Moment based estimation for the multivariate COGARCH(1,1) process

Thiago do Rêgo Sousa *　　Robert Stelzer †

Abstract

For the multivariate COGARCH process, we obtain explicit expressions for the second-order structure of the "squared returns" process observed on an equidistant grid. Based on this, we present a generalized method of moments estimator for its parameters. Under appropriate moment and strong mixing conditions, we show that the resulting estimator is consistent and asymptotically normal. Sufficient conditions for strong mixing, stationarity and identifiability of the model parameters are discussed in detail. We investigate the finite sample behavior of the estimator in a simulation study.

AMS 2010 Subject Classifications: primary: 62M05, 62M10, secondary: 60G51, 91B84.
Keywords: multivariate continuous time GARCH, second-order moment structure, estimation, generalized method of moments, model identification, Lévy process

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 introduced in Engle (1982); Bollerslev (1986). They are able to capture most of these stylized facts of financial data (see Cont (2001); Guillaume et al. (1997)). A special feature of GARCH like processes is that they usually exhibit heavy tails even if the driving noise is light tailed, a feature most other stochastic volatility models do not have (Fasen et al. (2006)).

In many financial applications, it is most natural to model the price evolution in continuous time, especially when dealing with high-frequency 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 (2014); Bibbona and Negri (2015); Haug et al. (2007); Klüppelberg et al. (2011); Maller et al. (2008); Müller (2010)).

*Center for Mathematical Sciences, Technical University of Munich, 85748 Garching, Boltzmannstr. 3, Germany, email: thiago.sousa@tum.de
†Institute of Mathematical Finance, Ulm University, 89075 Ulm, Helmholtzstr. 18, Germany, email: robert.stelzer@uni-ulm.de

In many cases one needs to model the joint price of several financial assets which exhibit a non-trivial dependence structure and therefore, multivariate models are needed. The MUCOGARCH process introduced in Stelzer (2010) 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 (1995). It is a $d-$dimensional stochastic process and it is defined as

$$G_t = \int_0^t V_{s-}^{1/2} \mathrm{d}L_s, \quad t \geq 0, \tag{1.1}$$

where L is an $\mathbb{R}^d$-valued Lévy process with non-zero Lévy measure and càdlàg sample paths. The matrix-valued volatility process $(V_s)_{s\in\mathbb{R}^+}$ depends on a parameter $\theta \in \Theta \subset \mathbb{R}^q$, it is predictable and its randomness depends only on L. We assume that we have a sample of size n of the log-price process (1.1) with true parameter $\theta_0 \in \Theta$ observed on a fixed grid of size $\Delta > 0$, and compute the log returns

$$\boldsymbol{G}_i = \int_{(i-1)\Delta}^{i\Delta} V_{s-}^{1/2} \mathrm{d}L_s, \quad i = 1, \ldots, n. \tag{1.2}$$

Therefore, an important question is how to estimate the true parameter θ_0 based on observations $(\boldsymbol{G}_i)_{i=1}^n$. In the univariate case, several methods have been proposed to estimate the parameters of the COGARCH process (Bayracı and Ünal (2014); Bibbona and Negri (2015); do Rêgo Sousa et al. (2019); Haug et al. (2007); Maller et al. (2008)). All these methods rely on the fact that the COGARCH process is, under certain regularity conditions, ergodic and strongly mixing.

In the univariate case, Fasen (2010) proved geometric ergodicity results for the COGARCH process (in fact, their results apply to a wider class of Lévy driven models). Recently, Stelzer and Vestweber (2019) derived sufficient conditions for the existence of a unique stationary distribution, for the geometric ergodicity, and for the finiteness 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 use their results to apply the generalized method of moments (GMM) for estimating the parameters of the MUCOGARCH process. To this end we compute the second-order structure of the squared returns in closed form, under appropriate assumptions.

Consistency and asymptotic normality of the GMM estimator is obtained under standard assumptions of strong mixing, existence of moments of the MUCOGARCH volatility process and model identifiability. Thus we discuss sufficient conditions, easily checkable for given parameter spaces ensuring strong mixing and existence of relevant moments.

The identifiability question is rather delicate, since the formulae 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 used in the one-dimensional COGARCH process cannot be generalised. In the end we can establish identifiability conditions that are not overly restrictive and easy to use.

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, strong mixing and identifiability of the model. In Section 6, we study the finite sample behavior of the estimators in a simulation study. Finally, Section A presents the proofs for the results of Sections 3 and 4.

2 Preliminaries

2.1 Notation

Denote the set of non-negative real numbers by $\mathbb{R}^+$. For $z \in \mathbb{C}, \Re(z)$ and $\Im(z)$ denote the real and imaginary part, respectively. 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 $GL_d(\mathbb{R})$, 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 Bernstein (2009); Horn and Johnson (1991). 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 in the ith line and jth column of A. Arbitrary norms of vectors or matrices are denoted by $\|\cdot\|$ in which case it is irrelevant which particular norm is used. The norm $\|\cdot\|_2$ denotes the operator norm on $M_{d^2}(\mathbb{R})$ associated with the usual Euclidean norm. 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 (2005)) 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 $t \in \mathbb{R}^+$, $i = 1, \dots, m$ and $j = 1, \dots, d$. We also refer to Lemma 2.2 in Behme (2012) 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

$$(\boldsymbol{Q}X)_{(k-1)d+l,(p-1)d+q} = X_{(k-1)d+p,(l-1)d+q} \quad \text{for all } k, l, p, q = 1, \dots, d,$$

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, 2009b, Theorem 4.3)). Let K_d be the commutation matrix characterized by $K_d \,\mathrm{vec}(A) = \mathrm{vec}(A^*)$ for all $A \in M_d(\mathbb{R})$ (see Magnus and Neudecker (1979) for more details). 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 $x \in \mathbb{R}^{d^4}$.

2.2 Lévy processes

A Lévy process $L = (L_t)_{t\in\mathbb{R}^+}$ in $\mathbb{R}^d$ is characterized by its characteristic function in Lévy-Khintchine form $\mathbb{E}e^{i\langle u,L_t\rangle} = \exp\{t\psi_L(u)\}$ for $t \in \mathbb{R}^+$ with

$$\psi_L(u) = i\langle\gamma_L, u\rangle - \frac{1}{2}\langle u, \Gamma_L u\rangle + \int_{\mathbb{R}^d} \left(e^{i\langle u,x\rangle} - 1 - i\langle u, x\rangle I_{[0,1]}(\|x\|)\right)\nu_L(dx), \quad u \in \mathbb{R}^d,$$

where $\gamma_L \in \mathbb{R}^d, \Gamma_L \in \mathbb{S}_d^+$ and the Lévy measure ν_L is a non-zero measure on $\mathbb{R}^d$ satisfying $\nu_L(\{0\}) = 0$ and $\int_{\mathbb{R}^d} \left(\|x\|^2 \wedge 1\right) \nu_L(dx) < \infty$. We assume w.l.o.g. 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 (2009); Sato (1999).

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. We can now recall the definition of the MUCOGARCH process.

Definition 3.1 (MUCOGARCH(1,1) - (Stelzer, 2010, Definition 3.1))**.** *Let L be an $\mathbb{R}^d$-valued Lévy process, $A, B \in M_d(\mathbb{R})$ and $C \in \mathbb{S}_d^{++}$. The process $G = (G_t)_{t\in\mathbb{R}^+}$ solving*

$$\begin{aligned} \mathrm{d}G_t &= V_{t-}^{1/2}\mathrm{d}L_t && (3.1)\\ V_t &= C + Y_t && (3.2)\\ \mathrm{d}Y_t &= (BY_{t-} + Y_{t-}B^*)\mathrm{d}t + AV_{t-}^{1/2}\mathrm{d}[L, L]_t^{\mathfrak{d}} V_{t-}^{1/2}A^* && (3.3) \end{aligned}$$

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 $Y = (Y_t)_{t\in\mathbb{R}^+}$ 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 $\sigma(B) \in \{z \in \mathbb{C} : \Re(z) < 0\}$, the process V, as long as no jump occurs, "mean reverts" to the level C at matrix exponential rate given by B. Since all jumps are positive semidefinite, C is not a mean level but, instead, a lower bound for V.

By (Stelzer, 2010, Theorems 3.2 and 4.4), the MUCOGARCH process is well-defined, 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.

Since the price process $(G_t)_{t\in\mathbb{R}^+}$ in (3.1) is defined in terms of the Lévy process L and $(Y_t)_{t\in\mathbb{R}_+}$, the existence of its moments is closely related to the existence of moments of L and the stationary distribution of $(Y_t)_{t\in\mathbb{R}_+}$.

Lemma 3.2. *Suppose that* $\mathbb{E}\|Y_0\|^p < \infty$ *and* $\mathbb{E}\|L_1\|^{2p} < \infty$ *for some* $p \geq 1$*. Then:*

(a) $\mathbb{E}\|Y_t\|^p < \infty$ *for all* $t \in \mathbb{R}^+$ *and* $t \mapsto \mathbb{E}\|Y_t\|^p$ *is locally bounded.*

(b) $\mathbb{E}\|G_t\|^{2p} < \infty$ *for all* $t \in \mathbb{R}^+$ *and* $t \mapsto \mathbb{E}\|G_t\|^{2p}$ *is locally bounded.*

4 Second-order structure of "squared returns"

In this section, we derive the second-order structure of the MUCOGARCH "squared returns" process $(\boldsymbol{G}_i \boldsymbol{G}_i^*)_{i \in \mathbb{N}}$ defined in terms of (1.2), which will be used in Section 5 to estimate the parameters A, B and C of the MUCOGARCH process. The proofs are postponed to Section A. We group the needed assumptions as follows.

Assumptions a (Lévy process)**.**

$(a.1)$ $\mathbb{E}L_1 = 0$.

$(a.2)$ $\mathrm{var}(L_1) = (\sigma_W + \sigma_L) I_d$, *with* $\sigma_W \geq 0$ *and* $\sigma_L > 0$.

$(a.3)$

$$\int_{\mathbb{R}^d} x_i x_j x_k \, \nu_L(dx) = 0, \quad \textit{for all } i, j, k \in \{1, \ldots, d\}.$$

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

$(a.5)$ *There exists a constant* $\rho_L > 0$ *such that*

$$\mathbb{E}[\mathrm{vec}([L, L^*]^{\mathfrak{d}}), \mathrm{vec}([L, L^*]^{\mathfrak{d}})^*]_1^{\mathfrak{d}} = \rho_L (I_{d^2} + K_d + \mathrm{vec}(I_d)\,\mathrm{vec}(I_d)^*).$$

$(a.6)$ $\mathbb{E}\|L_1\|^8 < \infty$.

Assumptions b (Parameters)**.**

$(b.1)$ $A \in GL_d(\mathbb{R})$.

$(b.2)$ *The matrices* $\mathcal{B}$ *and* $\mathcal{C}$ *defined below satisfy* $\sigma(\mathcal{B}), \sigma(\mathcal{C}) \in \{z \in \mathbb{C} : \Re(z) < 0\}$.

$$\begin{aligned} \mathcal{B} &:= B \otimes I + I \otimes B + \sigma_L (A \otimes A) \\ \mathcal{C} &:= \mathcal{B} \otimes I_{d^2} + I_{d^2} \otimes \mathcal{B} + \mathcal{A}\mathcal{R}, \end{aligned} \tag{4.1}$$

where $\mathcal{A} = (A \otimes A) \otimes (A \otimes A)$, $\mathcal{R} = \rho_L (\mathcal{Q} + \mathcal{K}_d \mathcal{Q} + I_{d^4})$, *and* $\mathcal{K}_d$ *and* $\mathcal{Q}$ *as in Section 2.1.*

Assumption c (MUCOGARCH volatility)**.**

$(c.1)$ $(Y_t)_{t \in \mathbb{R}_+}$ *is a second-order stationary MUCOGARCH volatility process.*

$(c.2)$ $(Y_t)_{t \in \mathbb{R}_+}$ *is a stationary MUCOGARCH volatility process and its stationary distribution satisfies* $\mathbb{E}\|Y_0\|^4 < \infty$.

Sufficient conditions for Assumption **c** are given in (Stelzer, 2010, Theorem 4.5). Note that $(c.2)$ implies $(c.1)$. We recall now the expressions for the second-order structure of the process Y and of the log-price returns process $(\boldsymbol{G}_i)_{i\in\mathbb{N}}$. First, for a second-order stationary $\mathbb{R}^d$-valued process, its autocovariance function $\mathrm{acov}_X : \mathbb{R} \mapsto M_d(\mathbb{R})$ is denoted by $\mathrm{acov}_X(h) = \mathrm{cov}\,(X_h, X_0) = \mathbb{E}\,(X_h X_0^*) - \mathbb{E}\,(X_0)\,E\,(X_0)^*$ 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 $\mathrm{acov}_Z = \mathrm{acov}_{\mathrm{vec}(Z)}$.

Proposition 4.1 ((Stelzer, 2010, Theorems 4.8, 4.11, Corollary 4.19 and Proposition 5.2))**.** *If Assumptions $(a.1)$-$(a.5)$, $(b.2)$ and $(c.1)$ hold, then*

$$\begin{aligned}
\mathbb{E}(\mathrm{vec}(Y_0)) &= -\sigma_L \mathcal{B}^{-1}(A\otimes A)\,\mathrm{vec}(C) && (4.2)\\
\mathrm{var}(\mathrm{vec}(Y_0)) &= \mathrm{var}(\mathrm{vec}(V_0)) = -\mathcal{C}^{-1}\big[\big(\sigma_L^2\mathcal{C}(\mathcal{B}^{-1}\otimes\mathcal{B}^{-1})\mathcal{A} + \mathcal{A}\mathcal{R}\big)(\mathrm{vec}(C)\otimes\mathrm{vec}(C))\\
&\quad + \big(\sigma_L(A\otimes A)\otimes I_{d^2} + \mathcal{A}\mathcal{R}\big)\,\mathrm{vec}(C)\otimes\mathbb{E}(\mathrm{vec}(Y_0))\\
&\quad + \big(\sigma_L I_{d^2}\otimes(A\otimes A) + \mathcal{A}\mathcal{R}\big)\mathbb{E}(\mathrm{vec}(Y_0))\otimes\mathrm{vec}(C)\big]\\
\mathrm{acov}_Y(h) &= \mathrm{acov}_V(h) = e^{\mathcal{B}h}\mathrm{var}(\mathrm{vec}(Y_0))\\
\mathbb{E}(\boldsymbol{G}_1) &= 0\\
\mathrm{var}(\boldsymbol{G}_1) &= (\sigma_L+\sigma_W)\Delta\mathbb{E}(C+Y_0) && (4.3)\\
\mathrm{acov}_{\boldsymbol{G}}(h) &= 0 \quad \textit{for all } h\in\mathbb{Z}\backslash\{0\}.
\end{aligned}$$

Based on Lemma 3.2 and Proposition 4.1, we obtain now the second-order properties of the MUCOGARCH process.

Lemma 4.2. *If Assumptions* $\boldsymbol{a}$,$\boldsymbol{b}$ *and* **c** *hold, then*

$$\begin{aligned}
\mathrm{acov}_{\boldsymbol{G}\boldsymbol{G}^*}(h) = e^{\mathcal{B}\Delta h}\mathcal{B}^{-1}(I_{d^2} - e^{-\mathcal{B}\Delta})(\sigma_L+\sigma_W)\mathrm{var}(\mathrm{vec}(V_0))&\\
\times\,(e^{\mathcal{B}^*\Delta} - I_{d^2})[(\sigma_W+\sigma_L)(\mathcal{B}^*)^{-1} - 2((A\otimes A)^*)^{-1}],\quad h\in\mathbb{N},&
\end{aligned} \tag{4.4}$$

$$\begin{aligned}
&\mathbb{E}\,\mathrm{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\,\mathrm{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)^* && (4.5)\\
&= \Delta\rho_L\big((\boldsymbol{Q} + K_d\boldsymbol{Q} + I_{d^2})(\mathbb{E}\,\mathrm{vec}(V_0)\,\mathrm{vec}(V_0)^*)\big)\times\\
&\quad (I_{d^2}+K_d)\boldsymbol{Q}(D^*)(I_{d^2}+K_d) + D + D^*,
\end{aligned}$$

with,

$$D := (\sigma_L+\sigma_W)\Big(\frac{1}{2}(\sigma_L+\sigma_W)\Delta^2\mathbb{E}\,\mathrm{vec}(V_0)\mathbb{E}\,\mathrm{vec}(V_0)^* + \mathrm{var}(\mathrm{vec}(V_0))\tilde{\mathcal{B}}\Big) \tag{4.6}$$

$$\tilde{\mathcal{B}} := \big[(\mathcal{B}^*)^{-1}(e^{\mathcal{B}^*\Delta} - I_{d^2}) - I_{d^2}\Delta\big]\big[(\sigma_W+\sigma_L)(\mathcal{B}^*)^{-1} - 2((A\otimes A)^*)^{-1}\big] \tag{4.7}$$

Remark 4.3. *If the Lévy process L has paths of finite variation, then Lemma 4.2 holds without the moment assumptions $(a.6)$ and $(c.2)$. This is because expectations involving stochastic integrals with finite variation Lévy integrators can be computed by using the compensation formula (see Remark A.2). In the following, we will define the moment based estimator for MUCOGARCH processes driven by general Lévy processes (without path restrictions). Only in Section 5.5 we will give a consistency result that distinguishes between Lévy process with paths of finite and infinite variation.*

Next, we define an estimator for the parameters A, B and C, which basically consists of comparing the sample moments to the model moments.

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 3.1 as depending on a parameter $\theta \in \Theta \subset \mathbb{R}^q$ for $q \in \mathbb{N}$.

The data used for estimation is an equidistant sample of d-dimensional log-prices $(\boldsymbol{G}_i)_{i=1}^n$ as defined in (1.2) with true parameter $\theta_0 \in \Theta$. We assume that the true σ_L, σ_W and ρ_L as used in Assumptions $(a.2)$ and $(a.5)$ are known. These assumptions are not very restrictive and are comparable to assuming iid standard normal noise in the discrete time multivariate GARCH process, which is very common (Francq and Zakoïan, 2019, eq. (10.6)).

5.1 Generalized Method of Moments (GMM) estimator

In order to estimate the parameter $\theta_0 \in \Theta$, we compare the sample moments (based on a sample of log-prices) to the model moments (based on the expressions (4.3), (4.4) and (4.5), provided they are well defined). More specifically, based on the observations $(\boldsymbol{G}_i)_{i=1}^n$ and a fixed $r < n$, the sample moments are defined as

$$\hat{k}_{n,r} = \frac{1}{n}\sum_{i=1}^{n-r} D_i = \frac{1}{n}\sum_{i=1}^{n-r} \begin{pmatrix} \operatorname{vec}(\boldsymbol{G}_i\boldsymbol{G}_i^*) \\ \operatorname{vec}(\operatorname{vec}(\boldsymbol{G}_i\boldsymbol{G}_i^*)\operatorname{vec}(\boldsymbol{G}_i\boldsymbol{G}_i^*)^*) \\ \vdots \\ \operatorname{vec}(\operatorname{vec}(\boldsymbol{G}_i\boldsymbol{G}_i^*)\operatorname{vec}(\boldsymbol{G}_{i+r}\boldsymbol{G}_{i+r}^*)^*) \end{pmatrix}. \tag{5.1}$$

The used number of lags of the true autocovariance function r 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 $\theta \in \Theta$, let

$$k_{\theta,r} = \begin{pmatrix} \mathbb{E}_\theta \operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*) \\ \mathbb{E}_\theta \operatorname{vec}(\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)^*) \\ \vdots \\ \mathbb{E}_\theta \operatorname{vec}(\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\operatorname{vec}(\boldsymbol{G}_{1+r}\boldsymbol{G}_{1+r}^*)^*) \end{pmatrix}, \tag{5.2}$$

where the expectations are explicitly given by (4.3), (4.4) and (4.5) 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

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\arg\min}\left\{(\hat{k}_{n,r} - k_{\theta,r})^T\Omega(\hat{k}_{n,r} - k_{\theta,r})\right\}, \tag{5.3}$$

where Ω is a positive definite weight matrix. The matrix Ω may be depend on the data but should converge in probability to a positive definite matrix of constants.

5.2 Asymptotic properties: general case

Additionally to Assumptions **a**, **b** and **c** we need assumptions for proving consistency and asymptotic normality of $\hat{\theta}_n$. These are mainly related to identifiability of the model parameters, stationarity, strong mixing and existence of certain moments of $(\boldsymbol{G}_i)_{i\in\mathbb{N}}$.

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) *[Identifiability]. Let* $r > 1$ *be fixed. For any* $\theta \neq \tilde{\theta} \in \Theta$ *we have* $k_{\theta,r} \neq k_{\tilde{\theta},r}$*.*

(d.4) *The map* $\theta \mapsto (A_\theta, B_\theta, C_\theta)$ *is continuously differentiable.*

(d.5) *The sequence* $(\boldsymbol{G}_i)_{i\in\mathbb{N}}$ *is strictly stationary and exponentially* α*-mixing.*

Assumption e (Moments)**.**

(e.1) *There exists a positive constant* $\delta > 0$ *such that* $\mathbb{E}\|\boldsymbol{G}_1\|^{8+\delta} < \infty$*.*

Assumption **e** can be written in terms of moments of L and Y_0 (see Lemma 3.2). We are now ready to state the strong consistency of the empirical moments in (5.1).

Lemma 5.1. *If Assumptions* ***a****,* ***b****,* ***c*** *and* (d.5) *hold, then* $\hat{k}_{n,r} \stackrel{\text{a.s.}}{\to} k_{\theta_0}$ *as* $n \to \infty$*.*

Proof. It follows from (d.5) that the log-price process $(\boldsymbol{G}_i)_{i\in\mathbb{N}}$ is ergodic and since 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 (Lemma 3.2 with $p = 2$ under (a.4) and (c.1)), we can apply Birkhoff's ergodic theorem ((Krengel, 1985, Theorem 4.4)) to conclude the result. $\square$

Next, we state the weak consistency property of the GMM estimator.

Theorem 5.2. *If Assumptions* ***a****,****b****,* ***c****,* (d.1)*,* (d.3)*-*(d.5) *hold, then the GMM estimator defined in* (5.3) *is weakly consistent.*

Proof. We check Assumptions 1.1-1.3 in Mátyás (1999) that ensure weak consistency of the GMM estimator in (5.3). Assumption 1.1 is satisfied due to our identifiability condition (d.3). It follows from (5.3) combined with Lemma 5.1 that

$$\sup_{\theta\in\Theta}\|\hat{k}_{n,r} - k_{\theta,r} - (k_{\theta_0,r} - k_{\theta,r})\| = \|\hat{k}_{n,r} - k_{\theta_0,r}\| \stackrel{\text{a.s.}}{\to} 0, \quad n \to \infty,$$

which is Assumption 1.2 of Mátyás (1999). Since the weight matrix Ω in (5.3) is non-random, their Assumption 1.3 is automatically satisfied, completing the proof. $\square$

In order to prove asymptotic normality of the GMM estimator, we need some auxiliary results.

Lemma 5.3. *If Assumptions* ***a****,* ***b****,* ***c****,* (d.1) *and* (d.4) *hold, then the map* $\Theta \mapsto k_{\theta,r}$ *in* (5.2) *is continuously differentiable.*

Proof. The the map $\Theta \mapsto k_{\theta,r}$ depends on the moments given in (4.3), (4.4) and (4.5). These moments are given in terms of products and Kronecker products involving the quantities A_θ, A_θ^{-1}, $\mathcal{B}_\theta$, $\mathcal{B}_\theta^{-1}$, $e^{-\alpha\mathcal{B}_\theta}$, $\alpha > 0$, C_θ, $\mathcal{C}_\theta$ and $\mathcal{C}_\theta^{-1}$. From (d.4) we obtain the continuous differentiability of $\mathcal{B}_\theta$,$\mathcal{B}_\theta^{-1}$, $\mathcal{C}_\theta$,$\mathcal{C}_\theta^{-1}$ and A_θ^{-1} on Θ. Let $i \in \{1,\dots,q\}$ be fixed. According to (2.1) in Wilcox (1967), the matrix exponential is differentiable and

$$\frac{\partial}{\partial\theta_i} e^{-\alpha\mathcal{B}_\theta} = -\int_0^\alpha e^{-(\alpha-u)\mathcal{B}_\theta}\left(\frac{\partial}{\partial\theta_i}\mathcal{B}_\theta\right)e^{-u\mathcal{B}_\theta}\mathrm{d}u. \tag{5.4}$$

Using the definition of $\mathcal{B}_\theta$ in (4.1) combined with $(d.1)$ and $(d.4)$ gives

$$\sup_{\theta\in\Theta}\|\mathcal{B}_\theta\| \leq 2\Big(\sup_{\theta\in\Theta}\|B_\theta\|\Big)\|I_d\| + \sigma_L\Big(\sup_{\theta\in\Theta}\|A_\theta\|^2\Big) < \infty.$$

Additionally, an application of the chain rule to $\frac{\partial}{\partial\theta_i}\mathcal{B}_\theta$ combined with $(d.1)$ and $(d.4)$ gives $\sup_{\theta\in\Theta}\|\frac{\partial}{\partial\theta_i}\mathcal{B}_\theta\| < \infty$ and, therefore,

$$\sup_{\theta\in\Theta}\left\|e^{-(\alpha-u)\mathcal{B}_\theta}\left(\frac{\partial}{\partial\theta_i}\mathcal{B}_\theta\right)e^{-u\mathcal{B}_\theta}\right\| \leq \sup_{\theta\in\Theta} e^{(|\alpha-u|+|u|)\|\mathcal{B}_\theta\|}\left(\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta_i}\mathcal{B}_\theta\right\|\right), \quad u\in[0,\alpha]. \tag{5.5}$$

Thus, the continuous differentiability of the map in (5.4) follows by dominated convergence with dominating function as in (5.5). Another application of the chain rule shows that the map $\theta \mapsto k_{\theta,r}$ is continuously differentiable on Θ. □

Lemma 5.4. *Assume that Assumptions* ***a***,***b***, ***c***, $(d.5)$ *and* $(e.1)$ *hold and let*

$$\Sigma_{\theta_0} = \mathbb{E}(F_1F_1^*) + \sum_{i=1}^\infty \mathbb{E}\big\{(F_1F_{1+i}^*) + \mathbb{E}(F_1F_{1+i}^*)^*\big\} \tag{5.6}$$

with $F_i = D_i - k_{\theta_0,r}$ *and* D_i *as defined in* (5.1). *Then for* $r\in\mathbb{N}_0$

$$\sqrt{n}(\hat{k}_{n,r} - k_{\theta_0,r}) \xrightarrow{d} \mathcal{N}(0,\Sigma_{\theta_0}), \quad n\to\infty.$$

Proof. For the asymptotic normality of (5.1) we use the Cramér-Wold device and show that

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n-r}\lambda^*F_i\Big) \xrightarrow{d} \mathcal{N}(0,\lambda^*\Sigma_{\theta_0}\lambda), \quad n\to\infty,$$

for all vectors $\lambda\in\mathbb{R}^{d^2+(r+1)d^4}$. Denote by $\alpha_{\boldsymbol{G}}$ the mixing coefficients of $(\boldsymbol{G}_i)_{i\in\mathbb{N}}$. Since each F_i is a measurable function of $\boldsymbol{G}_i,\dots,\boldsymbol{G}_{i+r}$ it follows from $(d.5)$ and Remark 1.8 of Bradley (2007) that $(\lambda^*F_i)_{i\in\mathbb{N}}$ is α-mixing with mixing coefficients satisfying $\alpha_F(n) \leq \alpha_{\boldsymbol{G}}(n-(r+1))$ for all $n\geq r+2$. Therefore, $\sum_{n=0}^\infty(\alpha_F(n))^{\frac{\epsilon}{2+\epsilon}} < \infty$ for all $\epsilon>0$. From $(e.1)$ we obtain $\mathbb{E}\|\lambda^*F_1\|^{2+\epsilon/4} < \infty$ for some $\epsilon>0$. Thus, the CLT for α-mixing sequences applies, see e.g. (Ibragimov and Linnik, 1971, Theorem 18.5.3), so that

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n-r}\lambda^*F_i\Big) \xrightarrow{d} \mathcal{N}(0,\zeta), \quad n\to\infty,$$

where

$$\zeta = \mathbb{E}\lambda^*F_1F_1^*\lambda + 2\sum_{i=1}^\infty \mathbb{E}\lambda^*F_1F_{1+i}^*\lambda.$$

Since $\lambda^*F_1F_{1+i}^*\lambda = \lambda^*(F_1F_{1+i}^*)^*\lambda$ we get (5.6) after rearranging the above equation. □

Theorem 5.5. *Assume that Assumptions **a**,**b**, **c**, **d** and* (e.1) *hold and that the matrix* Σ *in* (5.6) *is positive definite. Then the GMM estimator defined in* (5.3) *is asymptotically normal with covariance matrix*

$$(\mathcal{J}_{\theta_0})^{-1}\mathcal{I}_{\theta_0}(\mathcal{J}_{\theta_0})^{-1}, \tag{5.7}$$

where $\mathcal{J}_{\theta_0} = (\nabla_\theta k_{\theta_0,r})^\top \Omega(\nabla_\theta k_{\theta_0,r})$ *and* $\mathcal{I}_{\theta_0} = (\nabla_\theta k_{\theta_0,r})^\top \Omega\Sigma_{\theta_0}\Omega(\nabla_\theta k_{\theta_0,r})$.

Proof. We check Assumptions 1.7-1.9 of Theorem 1.2 in Mátyás (1999). Since by Lemma 5.3 the map $\theta \mapsto k_{\theta,r}$ is continuously differentiable, their 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 $\Theta \mapsto \frac{\partial}{\partial\theta}k_{\theta,r}$ in Lemma 5.3 that $\frac{\partial}{\partial\theta}(\hat{k}_{n,r} - k_{\theta_n}) \xrightarrow{P} (k_{\theta_0} - \frac{\partial}{\partial\theta}k_{\theta_0})$ as $n \to \infty$. Therefore, Assumption 1.8 in Mátyás (1999) is also satisfied. Since Lemma 5.4 implies Assumption 1.9, we conclude the result. $\square$

Remark 5.6. *In order to apply the results of Section 5.2 we need to check Assumption **c**, model identifiability* (d.3)*, strong mixing of the log-price returns sequence* (d.5) *and existence of certain moments of its stationary distribution (Assumption **e**) . In Sections 5.3 and 5.4 we give sufficient conditions for identifiability of the model parameters, strict stationarity and strong mixing. Then we use these results to derive in Section 5.5 more palpable conditions under which Theorems 5.2 and 5.5 can be applied.*

5.3 Sufficient conditions for strict stationarity and strong mixing

Sufficient conditions for the existence of a unique stationary distribution of $(Y_t)_{t\in\mathbb{R}^+}$, geometric ergodicity and for the finiteness of moments of order p of the stationary distribution have recently been given in Stelzer and Vestweber (2019). We state these conditions in the next theorem, which are conditions (i), (iv) and (v) of Theorem 4.3 in Stelzer and Vestweber (2019).

Theorem 5.7 (Geometric Ergodicity - (Stelzer and Vestweber, 2019, Theorem 4.3))**.** *Let* Y *be a MUCOGARCH volatility process which is* μ*-irreducible with the support of* μ *having non-empty interior and aperiodic. Assume that one of the following conditions is satisfied:*

(i) setting $p = 1$ *there exists* $\Xi \in \mathbb{S}_d^{++}$ *such that*

$$\Xi B + B^\top \Xi + \sigma_L A^\top \Xi A \in -\mathbb{S}_d^{++}, \tag{5.8}$$

(ii) there exist $p \in [1,\infty)$ *and* $\Xi \in \mathbb{S}_d^{++}$ *such that*

$$\int_{\mathbb{R}^d} \left(2^{p-1}\left(1 + K_{\Xi,A}\|y\|_2^2\right)^p - 1\right)\nu_L(dy) + pK_{\Xi,B} < 0, \tag{5.9}$$

where

$$K_{\Xi,B} = \max_{X\in\mathbb{S}_d^+,\mathrm{tr}(X)=1} \frac{\mathrm{tr}\left(\left(\Xi B + B^\top\Xi\right)X\right)}{\mathrm{tr}(\Xi X)} \text{ and } K_{\Xi,A} = \max_{X\in\mathbb{S}_d^+,\mathrm{tr}(X)=1} \frac{\mathrm{tr}\left(A^\top \Xi A X\right)}{\mathrm{tr}(\Xi X)},$$

(iii) there exist $p \in [1,\infty)$ *and* $\Xi \in \mathbb{S}_d^{++}$ *such that*

$$\max\left\{2^{p-2}, 1\right\} K_{\Xi,A} \int_{\mathbb{R}^d} \|y\|_2^2 \left(1 + \|y\|_2^2 K_{\Xi,A}\right)^{p-1} \nu_L(dy) + K_{\Xi,B} < 0 \tag{5.10}$$

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 its stationary distribution has a finite p-th moment.

A consequence of Theorem 5.7 is that the process Y is exponentially β-mixing. This implies α-mixing of the log-price process as we state next. For more details on mixing conditions we refer to Bradley (2007).

Corollary 5.8. *If Y is strictly stationary and exponentially β-mixing, then the log-price process $(\boldsymbol{G}_i)_{i\in\mathbb{N}}$ is stationary, exponentially α-mixing, and as a consequence also ergodic.*

Proof. Since Y is an exponentially β-mixing, homogeneous strong Markov process (Stelzer, 2010, Theorem 4.4), and driven only by the discrete part of the quadratic variation of L, the proof follows by the same arguments as for Theorem 3.4 in Haug et al. (2007). □

Next, we state a result which gives sufficient conditions for the irreducibility of the MUCOGARCH volatility process Y process, which is one of the sufficient conditions for the geometric ergodicity result in Theorem 5.7.

Theorem 5.9 (Irreducibility and Aperiodicity - (Stelzer, 2010, Theorem 5.1 and Corollary 5.2))**.** *Let Y be a MUCOGARCH volatility process driven by a Lévy process whose discrete part is 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 is irreducible w.r.t. the Lebesgue measure restricted to an open neighborhood of zero in $\mathbb{S}_d^+$ and aperiodic.*

5.4 Sufficient conditions for identifiability

In this Section we investigate the identifiability of the model parameters from the model moments, i.e., we investigate the injectivity of the map $\theta \mapsto k_{\theta,r}$ on an appropriate compact set Θ. Recall that we can divide the this map into the composition of $\theta \mapsto (A_\theta, B_\theta, C_\theta) \mapsto k_{\theta,r}$. Injectivity of $\theta \mapsto (A_\theta, B_\theta, C_\theta)$ holds if e.g. it simply maps the entries of θ to the entries of the matrices $(A_\theta, B_\theta, C_\theta)$. Thus, we only need to investigate the injectivity of the map $(A_\theta, B_\theta, C_\theta) \mapsto k_{\theta,r}$. As we will see, there will appear some restrictions on the matrices A_θ, B_θ, which are related to the fact that we need to take the logarithm of a matrix exponential, and we need to ensure this is well defined. We will omit θ from the notation, except when explicitly needed. We start with the identifiability of the matrix C.

Lemma 5.10. *Assume that Assumptions (a.1)-(a.5), (b.2) and (c.1) hold and that $\sigma(B) \subset \{z \in \mathbb{C} : \Re(z) < 0\}$. If the matrices A and B are known, then $\mathbb{E}(\boldsymbol{G}_1\boldsymbol{G}_1^*)$ uniquely determines C.*

Proof. Since $\sigma(B \otimes I + I \otimes B) = \sigma(B) + \sigma(B) \subset \{z \in \mathbb{C} : \Re(z) < 0\}$, the matrix $B \otimes I + I \otimes B$ is invertible. The rest of the proof follows by noting that from (4.2) and (4.3) it follows that

$$\operatorname{vec}(C) = (\sigma_L + \sigma_W)^{-1}\Delta^{-1}(B \otimes I + I \otimes B)^{-1}\mathcal{B}\operatorname{vec}(\mathbb{E}(\boldsymbol{G}_1\boldsymbol{G}_1^*)).$$

□

For the identification of the matrices A and B we need to use the second-order structure of the squared returns process in Lemma 4.2. We first state three auxiliary results, which provide conditions such that we can identify the components of the autocovariance function in (4.4).

Lemma 5.11. *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*

$$\left|\frac{\sigma_L - \sigma_W}{2\sigma_L}\right| \|A \otimes A\|_S < -2\max\{\Re(\sigma(B))\}, \tag{5.11}$$

with

$$\|X\|_S = \|(S^{-1} \otimes S^{-1})X(S \otimes S))\|_2, \quad X \in M_{d^2}(\mathbb{R}), S \in GL_d(\mathbb{C}), \tag{5.12}$$

then the matrix

$$(\sigma_W + \sigma_L)(\mathcal{B}^*)^{-1} - 2((A \otimes A)^*)^{-1} \tag{5.13}$$

is invertible.

Proof. From (Bernstein, 2009, 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 (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 diagonalizable, we can use (Bernstein, 2009, 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 diagonalizable. Now we rewrite the first equation on p. 106 in Stelzer (2010) 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-Fike Theorem (Horn and Johnson, 1991, Theorem 6.3.2) to see that (5.11) implies that all eigenvalues of $(X+Y)(\sigma_L+\sigma_W)$ are in $\{z \in \mathbb{C} : \Re(z) < 0\}$ and, therefore, $X+Y$ is invertible. ☐

Lemma 5.12. *If $A \in M_d(\mathbb{R})$ is such that $A_{(1,1)}, \ldots, A_{(1,j-1)} = 0$ and $A_{(1,j)} > 0$ for some $j \in \{1, \ldots, d\}$, 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 $i \in \{1, \ldots, d\}$, 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

$$(A^2_{(1,1)}, A_{(1,1)}A_{(2,1)}, \ldots, A_{(1,1)}A_{(d,1)}). \tag{5.14}$$

Since $A_{(1,1)} > 0$, (5.14) allows us to identify first $A_{(1,1)}$ and then $A_{(2,1)}, \ldots, A_{(d,1)}$. Now, for each $k \in \{2, \ldots, d\}$, note 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

$$(2A_{(1,1)}A_{(1,k)}, A_{(1,1)}A_{(2,k)} + A_{(1,k)}A_{(2,1)}, \ldots, A_{(1,1)}A_{(d,k)} + A_{(1,k)}A_{(d,1)}). \tag{5.15}$$

Since $A_{(1,1)} > 0$, we identify $A_{(1,k)}$ from the first entry of (5.15). Now, since also $A_{(2,1)}, \ldots, A_{(d,1)}$ are already known, we can identify $A_{(2,k)}, \ldots, A_{(d,k)}$. Thus, all entries of A can be identified. The cases $A_{(1,j)} > 0$ for some $j > 1$ follow similarly. ☐

Lemma 5.13. *Assume that the Assumptions* $\boldsymbol{a}$*,*$\boldsymbol{b}$ *and* $\boldsymbol{c}$ *and the conditions of Lemma 5.12 hold, that the matrix in* (5.13) *is invertible, that* $\sigma(\mathcal{B}) \subset \{z \in \mathbb{C} : -\pi < \Im(z)\Delta < \pi, \Re(z) < 0\}$ *and that* $\operatorname{var}(\operatorname{vech}(V_0))$ *is invertible. Define* $M = (e^{\mathcal{B}\Delta})^{-1}\operatorname{acov}_{\theta,\boldsymbol{GG}^*}(1)$*. Then,* $\operatorname{acov}_{\theta,\boldsymbol{GG}^*}(1)$ *and* $\operatorname{acov}_{\theta,\boldsymbol{GG}^*}(2)$ *uniquely identify* $\mathcal{B}$ *and* M*.*

Proof. Since M is given in terms of $\mathcal{B}$ and $\operatorname{acov}_{\theta,\boldsymbol{GG}^*}(1)$, we only need to identify $\mathcal{B}$. 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 take 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 $\operatorname{var}(\operatorname{vech}(V_0))$ 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

$$e^{\mathcal{B}\Delta h}\mathcal{B}^{-1}(I_{d^2} - e^{-\mathcal{B}\Delta})(\sigma_L + \sigma_W)\operatorname{var}(\operatorname{vec}(V_0))(e^{\mathcal{B}^*\Delta} - I_{d^2})[(\sigma_W + \sigma_L)(\mathcal{B}^*)^{-1} - 2((A \otimes A)^*)^{-1}]$$

are representing linear operators from $\mathbb{S}_d$ to itself. Under the assumptions made, the above product involves only invertible linear operators. Hence $\operatorname{acov}_{\theta,\boldsymbol{GG}^*}(h)$ is invertible (over $\mathbb{S}_d$) for every $h > 0$. Thus,

$$e^{\mathcal{B}\Delta} = \operatorname{acov}_{\theta,\boldsymbol{GG}^*}(2)[\operatorname{acov}_{\theta,\boldsymbol{GG}^*}(1)]^{-1}.$$

By the assumptions on the eigenvalues of $\mathcal{B}$ there is a unique logarithm for $e^{\mathcal{B}\Delta}$ (see (Horn and Johnson, 1991, Section 6.4) or (Schlemm and Stelzer, 2012, Lemma 3.11)), so $\mathcal{B}\Delta$ and thus $\mathcal{B}$ 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, 2009a, Proposition 3.1) and Lemma 5.12. □

Lemma 5.14 (Identifiability of A, B and C)**.** *For all* $\theta \in \Theta$*, assume the conditions of Lemma 5.13,* $\sigma(B_\theta) \subset \{z \in \mathbb{C} : \Re(z) < 0\}$ *and that the entries of the matrices* A_θ *and* B_θ *satisfy: for some* $k \neq l \in \{1, \ldots, d\}$*,* $A_{(k,l),\theta} > 0$*,* $A_{(k,l),\theta} \neq A_{(l,k),\theta}$ *and* $B_{(k,l),\theta} = B_{(l,k),\theta}$*. Then* $k_{\theta,2}$ *uniquely identifies* A_θ*,* B_θ *and* C_θ*.*

Proof. Recall that we omit θ in the notation. Assume w.l.o.g that $\sigma_L = 1$. Because of Lemma 5.10, we only need to show the identification of A and B.
Assume first that $d = 2$. Then the matrix $\mathcal{B}$ from (4.1) is

$$\begin{pmatrix} 2B_{(1,1)} + A_{(1,1)}^2 & B_{(1,2)} + A_{(1,1)}A_{(1,2)} & B_{(1,2)} + A_{(1,1)}A_{(1,2)} & A_{(1,2)}^2 \\ B_{(2,1)} + A_{(1,1)}A_{(2,1)} & B_{(1,1)} + B_{(2,2)} + A_{(1,1)}A_{(2,2)} & A_{(1,2)}A_{(2,1)} & B_{(1,2)} + A_{(1,2)}A_{(2,2)} \\ B_{(2,1)} + A_{(1,1)}A_{(2,1)} & A_{(1,2)}A_{(2,1)} & B_{(1,1)} + B_{(2,2)} + A_{(1,1)}A_{(2,2)} & B_{(1,2)} + A_{(1,2)}A_{(2,2)} \\ A_{(2,1)}^2 & B_{(2,1)} + A_{(2,1)}A_{(2,2)} & B_{(2,1)} + A_{(2,1)}A_{(2,2)} & 2B_{(2,2)} + A_{(2,2)}^2 \end{pmatrix}. \tag{5.16}$$

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 $d > 2$. We assume w.l.o.g. that $A_{(1,2)} > 0$, $A_{(1,2)} \neq A_{(2,1)}$ and $B_{(1,2)} = A_{(2,1)}$. Write the matrix $\mathcal{B}$ from (4.1) in the following block form:

$$\mathcal{B} = B \otimes I + I \otimes B + A \otimes A = \begin{pmatrix} \mathcal{B}^{(1,1)} & \dots & \mathcal{B}^{(1,d)} \\ \vdots & \ddots & \vdots \\ \mathcal{B}^{(d,1)} & \dots & \mathcal{B}^{(d,d)} \end{pmatrix}, \tag{5.17}$$

where $\mathcal{B}^{(i,j)} \in M_d(\mathbb{R})$ for all $i, j = 1, \dots, d$. First, we have that

$$\mathcal{B}^{(1,2)} = \begin{pmatrix} B_{(1,2)} + A_{(1,2)}A_{(1,1)} & A_{(1,2)}A_{(1,2)} & A_{(1,2)}A_{(1,3)} & \cdots & A_{(1,2)}A_{(1,d)} \\ A_{(1,2)}A_{(2,1)} & B_{(1,2)} + A_{(1,2)}A_{(2,2)} & A_{(1,2)}A_{(2,3)} & \cdots & A_{(1,2)}A_{(2,d)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{(1,2)}A_{(d,1)} & A_{(1,2)}A_{(d,2)} & A_{(1,2)}A_{(d,3)} & \cdots & B_{(1,2)} + A_{(1,2)}A_{(d,d)} \end{pmatrix}, \tag{5.18}$$

Since $A_{(1,2)} > 0$ we can identify it from (5.18), because $\mathcal{B}^{(1,2)}_{(1,2)} = A^2_{(1,2)}$. Then we use the off-diagonal entries of the matrix $\mathcal{B}^{(1,2)}$ in (5.18) together with $A_{(1,2)}$ to identify all the off-diagonal entries of the matrix A. Next we identify the diagonal entries of A. It follows from (5.17) that

$$\begin{cases} \mathcal{B}^{(k,k)}_{(1,2)} = B_{(1,2)} + A_{(k,k)}A_{(1,2)} \\ \mathcal{B}^{(k,k)}_{(2,1)} = B_{(2,1)} + A_{(k,k)}A_{(2,1)} \end{cases}, \quad k = 1, \dots, d. \tag{5.19}$$

Since $A_{(1,2)} - A_{(2,1)} \neq 0$ and $B_{(1,2)} = B_{(2,1)}$, the system of equations (5.19) gives

$$A_{(k,k)} = (\mathcal{B}^{(k,k)}_{(1,2)} - \mathcal{B}^{(k,k)}_{(2,1)})/(A_{(1,2)} - A_{(2,1)}), \quad k = 1, \dots, d.$$

Finally, since the matrix A is now completely known, we can use (5.17) to identify all entries of B. $\square$

In Lemma 5.14 we identify the matrices A and B only from $\mathcal{B}$ and, therefore, some mild restrictions on the off-diagonal entries of B appear. In order to avoid those restrictions, we could to take the structure of $\mathbb{E}\operatorname{vec}(\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)^*)$ in (4.5) into account when proving identifiability and we expect that one can improve the identification results since more moment conditions are used. However, already in the 2-dimensional case the results on 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 (4.5) is not one-to-one in the space of matrices of the form $\mathbb{E}\operatorname{vec}(V_0)\operatorname{vec}(V_0)^*$. In the end, in order to use the moment conditions $\mathbb{E}\operatorname{vec}(\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\operatorname{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)^*)$, we need to assume that the matrices $B \otimes I + I \otimes B$ and $A \otimes A$ commute (see (Do Rêgo Sousa, 2019, Lemma 3.5.18)). Since commutativity is a quite strong condition, it seems highly preferable to work with the class of MUCOGARCH processes, which are identifiable by Lemma 5.14. The exponential decay of the autocovariance function of the model is still quite flexible, because of the interplay between the matrices A and B (see (5.16), for instance).

5.5 Asymptotic properties: general case revisited

Here, we combine the results of Sections 5.2-5.4 to give easily verifiable conditions under which the GMM estimator $\hat{\theta}_n$ will be consistent and asymptotically normal. We assume that the parameter θ contains the entries of the matrices $(A_\theta, B_\theta, C_\theta)$ so that the map $\theta \mapsto (A_\theta, B_\theta, C_\theta)$ is automatically injective and continuously differentiable on Θ.

First, we define

$$\|x\|_S = \|(S^{-1} \otimes S^{-1})x\|_2, \quad x \in \mathbb{R}^{d^2}, S \in GL_d(\mathbb{C}) \tag{5.20}$$

$$K_{2,S} = \max_{X \in \mathbb{S}_d^+, \|X\|_2 = 1} \left(\frac{\|X\|_2}{\|\operatorname{vec}(X)\|_S} \right), \quad S \in GL_d(\mathbb{C}). \tag{5.21}$$

Consider now the following group of assumptions:

Assumptions f (Parameter space)**.** *For all $\theta \in \Theta$ it holds:*

$(f.1)$ *The matrices B_θ satisfy $\sigma(B_\theta) \subset \{z \in \mathbb{C} : \Re(z) < 0\}$.*

$(f.2)$ *The matrix $\mathcal{B}_\theta$ satisfy $\sigma(\mathcal{B}_\theta) \subset \{z \in \mathbb{C} : -\pi < \Im(z)\Delta < \pi, \Re(z) < 0\}$.*

$(f.3)$ *The matrix $B_\theta \in M_d(\mathbb{R})$ is diagonalizable with $S_\theta \in GL_d(\mathbb{C})$ such that $S_\theta^{-1} B_\theta S_\theta$ is diagonal.*

$(f.4)$ *The entries of the matrices A_θ and B_θ satisfy: for some $k \neq l \in \{1, \ldots, d\}$, $A_{(k,l),\theta} > 0$, $A_{(k,l),\theta} \neq A_{(l,k),\theta}$ and $B_{(k,l),\theta} = B_{(l,k),\theta}$.*

$(f.5)$ *The matrix $\operatorname{var}_\theta(\operatorname{vech}(V_0))$ is invertible.*

$(f.6)$ $|\frac{\sigma_L - \sigma_W}{2\sigma_L}| \|A_\theta \otimes A_\theta\|_{S_\theta} < -2\max\{\Re(\sigma(B_\theta))\}$ *with S_θ as in $(f.3)$ and $\|A_\theta \otimes A_\theta\|_{S_\theta}$ as in* (5.12).

$(f.7)$ *There exists $\Xi_\theta \in \mathbb{S}_d^{++}$ such that, condition* (5.8) *holds with A, B replaced by A_θ, B_θ.*

$(f.8)$ $m(4, \theta) < 0$ *where*

$$m(p, \theta) := \int_{\mathbb{R}^d} ((1 + \alpha_\theta \|\operatorname{vec}(yy^*)\|_{S_\theta})^p - 1)\nu_L(\mathrm{d}y) + 2p\max\{\Re(\sigma(B_\theta))\}, \tag{5.22}$$

$\alpha_\theta = \|S_\theta\|_2^2 \|S_\theta^{-1}\|_2^2 K_{2,B_\theta} \|A_\theta \otimes A_\theta\|_{S_\theta}$ *with K_{2,B_θ} as in* (5.21), $\|\operatorname{vec}(yy^*)\|_{S_\theta}$ *as in* (5.20) *and S_θ as in $(f.3)$.*

Assumptions g (MUCOGARCH process at θ_0)**.**

$(g.1)$ *The MUCOGARCH volatility process Y is stationary, μ-irreducible with the support of μ having non-empty interior and aperiodic.*

$(g.2)$ $m(p, \theta_0) < 0$ *for some $p > 4$.*

Assumption $(f.1)$-$(f.6)$ collect the needed identifiability assumptions from Section 5.4. Assumption $(f.7)$ is a sufficient condition under which we have uniqueness of the stationary distribution of Y and geometric ergodicity (Section 5.3). For the asymptotic results of the GMM estimator in Theorems 5.2 and 5.5, we need to ensure $\mathbb{E}\|Y_0\|^p < \infty$ for appropriate $p > 1$, and this would require checking Assumptions 5.9 or 5.10 with $p > 1$. However, this imposes strong conditions on the Lévy process (Stelzer and Vestweber, 2019, Remark 4.4). Instead, we require diagonalizability of the matrix B_θ ((f.3)), which is not a very restrictive assumption, and check (5.22) to ensure $\mathbb{E}\|Y_0\|^p < \infty$ (Stelzer, 2010, Theorem 4.5), which is less restrictive.

In view of the above assumptions and the results of Sections 5.1-5.4 we have the following consistency result.

Corollary 5.15 (Consistency of the GMM estimator - L has paths of infinite variation)**.** *Suppose that assumptions* ***a****,* ***b****,* (d.1),(f.1)-(f.8) *and* (g.1) *hold. Then the GMM estimator defined in* (5.3) *is weakly consistent.*

If the paths of the driving Lévy process are of finite variation, we can relax even more the conditions from Corollary 5.15. Before we state this result, 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

$$
\begin{aligned}
&\lim_{t\to\infty} \mathbb{E}\left(X_t\right) = \mu, \qquad \lim_{t\to\infty} \operatorname{var}\left(\operatorname{vec}\left(X_t\right)\right) = \Sigma \\
&\lim_{t\to\infty} \sup_{h\in\mathbb{R}^+} \left\{\left\|\operatorname{cov}\left(\operatorname{vec}\left(X_{t+h}\right), \operatorname{vec}\left(X_t\right)\right) - f(h)\right\|\right\} = 0.
\end{aligned}
$$

Corollary 5.16 (Consistency of the GMM estimator - L has paths of finite variation)**.** *Suppose that assumptions* ***a****,* ***b****,* (d.1)*,* (f.1)*,* (f.2)*,* (f.4)-(f.7)*,* (g.1) *hold and that L has paths of finite variation. Then, the GMM estimator defined in* (5.3) *is weakly consistent.*

Proof. Let $D \in \mathbb{S}_d^+$ be a constant matrix, and consider a MUCOGARCH process $(Y_t)_{t\in\mathbb{R}^+}$ solving (3.3) have starting value D. Then, a combination of Assumptions (a.2), (a.5) with the fact that the starting value D is non-random and the hypothesis imposed on the matrices $B_\theta, \mathcal{B}_\theta, \mathcal{C}_\theta$ allow us to apply Theorem 4.20(ii) in Stelzer (2010) to conclude that the process $(Y_t)_{t\in\mathbb{R}^+}$ is asymptotically second-order stationary. Additionally, Theorem 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 $Y_t \in \mathbb{S}_d^+$, and $\operatorname{tr}(Y^*Y)$ (with tr denoting the usual trace functional) defines a scalar product on $\mathbb{S}_d$ via $\operatorname{tr}(Y_t^*Y_t) = \operatorname{vec}(Y_t^*)\operatorname{vec}(Y_t)$ it follows that

$$
\begin{aligned}
\mathbb{E}\|Y_t\|_2^2 = &\operatorname{tr}(Y_t^*Y_t) = \operatorname{vec}(Y_t^*)\operatorname{vec}(Y_t) = \sum_{i,j}^{d} \mathbb{E}Y_{t,ij}^2 = \sum_{i,j}^{d} \operatorname{var}(Y_{t,ij}) + \sum_{i,j}^{d} (\mathbb{E}Y_{t,ij})^2 \\
&\operatorname{tr}(\operatorname{var}(Y_t)) + \|\mathbb{E}(Y_t)\|_2^2, \quad t > 0.
\end{aligned}
\tag{5.23}
$$

Since both maps $t \mapsto \mathbb{E}\|Y_t\|$ and $t \mapsto \operatorname{var}(Y_t)$ are continuous ((Stelzer, 2010, eqs. (4.7) and (4.16))), it follows from (5.23) that $\limsup_{t\geq 0} \mathbb{E}\|Y_t\|^2 < \infty$. Since Theorem 5.7(i) implies convergence of the transition probabilities in total variation, which in turns implies weak convergence (e.g. (Klenke, 2013, Exercise 13.2.2)), we have that $Y_t \xrightarrow{d} Y_0$ as $t \to \infty$, with Y_0 being the stationary version of Y. Hence, we can use the continuous mapping theorem and (Billingsley, 2008, Theorem 25.11) to conclude that $\mathbb{E}\|Y_0\|^2 < \infty$. Finally, the result follows by an application of Lemma 3.2, Theorem 5.2, Corollary 5.8 and Remark A.2. □

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 (cf. (e.1)). This is summarized in the next corollary.

Corollary 5.17 (Asymptotic normality of the GMM estimator)**.** *If Assumptions **a**, **b**, (d.1), (d.2), (d.4), **f** and **g** hold, then the GMM estimator defined in* (5.3) *is asymptotically normal with covariance matrix as in* (5.7)*.*

Proof. By the same arguments of (Stelzer, 2010, Theorem 4.5) combined with (Lindner and Maller, 2005, Proposition 4.1) and (g.2), it follows that $\mathbb{E}\|Y_0\|^p < \infty$ for some $p > 4$. The rest of the proof is just an application of Theorem 5.5. □

Remark 5.18. *The advantage of Corollaries 5.15-5.17 is that Assumption **f** can be checked numerically and Assumption **g** holds true if e.g., the Lévy process L is a compound Poisson process with jump distribution having a density which is strictly positive around zero (see Theorem 5.9).*

If (g.1) holds, the stationary distribution of Y is automatically a maximum irreducibility measure. All maximal irreducibility measures are equivalent and thus the support of the stationary distribution has a support which has non-empty interior. The latter in turn implies that the variance has to be an invertible operator (non-invertibility is equivalent to the distribution being concentrated on a proper linear subspace) which is (f.5) for θ_0.

In the next section, we investigate the finite sample performance of the estimators in a simulation study.

6 Simulation study

To assess the performance of the GMM estimator, we will focus on the MUCOGARCH model in dimension $d = 2$. We fix $L_t = L_t^{\mathfrak{d}} + \sqrt{\sigma_W} W_t$ for $t \in \mathbb{R}^+$ where $L^{\mathfrak{d}}$ is a bivariate compound Poisson process (CPP), W is a standard bivariate Brownian motion, independent of $L^{\mathfrak{d}}$ and $\sigma_W \geq 0$ is fixed. 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 (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 (2007)). In applications, a CPP has also been used in combination with the univariate COGARCH(1,1) process for modeling high frequency data (see Müller (2010)). The jump distribution of $L^{\mathfrak{d}}$ is chosen as $N(0, 1/4I_2)$ and the jump 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 Assumptions **a** from Section 4 (with $\sigma_L = 1$ and $\sigma_W > 0$). Based on the identification Lemma 5.14, 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:

$$A_\theta = \begin{pmatrix} \theta^{(1)} & \theta^{(2)} \\ \theta^{(3)} & \theta^{(4)} \end{pmatrix}, \quad B_\theta = \begin{pmatrix} \theta^{(5)} & \theta^{(6)} \\ \theta^{(6)} & \theta^{(7)} \end{pmatrix} \quad \text{and} \quad C_\theta = \begin{pmatrix} \theta^{(8)} & \theta^{(9)} \\ \theta^{(9)} & \theta^{(10)} \end{pmatrix},$$

with $\theta^{(2)} > 0$ and $\theta^{(2)} \neq \theta^{(3)}$. Thus, Assumption $(d.4)$ and $(f.4)$ are automatically satisfied. The data used for estimation is a sample of the log-price process $\boldsymbol{G} = (\boldsymbol{G}_i)_{i=1}^n$ as defined in (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 6.1. *We fix* $\sigma_W = 1$*,*

$$A_{\theta_0} = \begin{pmatrix} 0.85 & 0.10 \\ -0.10 & 0.75 \end{pmatrix}, \quad B_{\theta_0} = \begin{pmatrix} -2.43 & 0.05 \\ 0.05 & -2.42 \end{pmatrix} \quad \text{and} \quad C_{\theta_0} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}. \tag{6.1}$$

Example 6.2. *We fix* $\sigma_W = 0$*,* A_{θ_0} *and* C_{θ_0} *are as in Example 6.1 and*

$$B_{\theta_0} = \frac{1}{4} \begin{pmatrix} -2.43 & 0.05 \\ 0.05 & -2.42 \end{pmatrix}. \tag{6.2}$$

For the chosen Lévy process here, Assumption $(g.1)$ is satisfied. In Example 6.1, θ_0 is chosen in such a way that the asymptotic normality of $\hat{\theta}_n$ can be verified. Then, in Example 6.2 we rescale B_{θ_0} from Example 6.1 in such a way that our sufficient conditions for weak consistency are satisfied, but our sufficient conditions for asymptotic normality in Corollary 5.17 are not satisfied.

Due to the identifiability Lemma 5.14 we need to choose $r \geq 2$. For comparison purposes, we perform the estimation for maximum lags $r \in \{2, 5, 10\}$ and sample sizes $n \in \{1\,000, 10\,000, 100\,000\}$. The computations are performed with the `optim` routine in combination with the `Nelder-Mead` algorithm in R (R Core Team (2017)). Initial values for the estimation were found by the `DEoptim` routine on a neighborhood around the true parameter θ_0. We only consider estimators based on the identity matrix for the weight matrix Ω in (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 6.1 and 6.2.

6.1 Simulation results for Example 6.1

We can check numerically that the matrices $A_{\theta_0}, B_{\theta_0}$ and A_{θ_0} are such that Assumptions **b** and $(f.2)$-$(f.6)$ hold. Additionally, the eigenvalues of the matrix $B_{\theta_0} + B_{\theta_0}^* + \sigma_L A_{\theta_0}^* A_{\theta_0}$ are -4.067 and -4.328, so it is negative definite and Assumptions $(f.7)$ holds. For our choice of θ_0 we have that B_{θ_0} is diagonalizable with $B_{\theta_0} = S_{\theta_0} D_{\theta_0} S_{\theta_0}^{-1}$, where

$$S_{\theta_0} = \begin{pmatrix} -0.671 & -0.741 \\ -0.741 & 0.671 \end{pmatrix} \quad \text{and} \quad D_{\theta_0} = \begin{pmatrix} -2.375 & 0 \\ 0 & -2.475 \end{pmatrix}.$$

In addition, for $p = 4.001$,

$$\int_{\mathbb{R}^2} ((1 + \alpha_{\theta_0} \|\operatorname{vec}(yy^*)\|_{S_{\theta_0}})^p - 1)\nu_L(\mathrm{d}y) + 2p \max\{\Re(\sigma(B_{\theta_0}))\} = -0.024 < 0, \tag{6.3}$$

so $(g.2)$ is also valid. Therefore, all assumptions for applying Corollary 5.17 can be verified, which imply assumption **e**, and ensure asymptotic normality. We also note that the chosen parameters are very close to not satisfying Assumption (6.3).

We investigate the behavior of the bias and standard deviation in Figure 1, where we excluded those paths for which the algorithm did not converge successfully (around 10 percent of the paths of length $n = 1\,000$ and less than 3 percent for larger n). Figure 1 show the estimated absolute values of the bias and standard deviation for different lags r and varying n. As expected, they decay when n increases. Additionally, the results favor the choice of maximum lag $r = 10$, which is already expected since using more lags of the autocovariance function usually helps to give a better fit. It is also worth noting that the estimation of the parameters in the matrix B_{θ_0} is more difficult than the other parameters, specially for $n \in \{1\,000, 10\,000\}$.

Figures 2 and 3 assess asymptotic normality though normal QQ-plots. Based on the previous findings we fix $r = 10$, since it gave the best results. This might have to do with the fact that using just a few lags for the autocovariance function ($r = 2$ or $r = 5$) are not sufficient for a good fit. 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 2 and 3). These plots are clearly in line with the 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 $n \in \{1\,000, 10\,000\}$, but they get closer to a normal distribution for $n = 100\,000$. The tails of the plots for $A_{(2,1),\hat{\theta}_n}$ in Figure 2 is not close to a normal (although the plots show its convergence). This is maybe due to identifiability condition in Lemma 5.14 which requires $A_{(2,1),\theta} > 0$ but $A_{(2,1),\theta_0} = 0.1$ is very close to the boundary. For $n = 1\,000$, there are very large negative outliers for the estimates of B_{θ_0}, which affects the bias substantially.

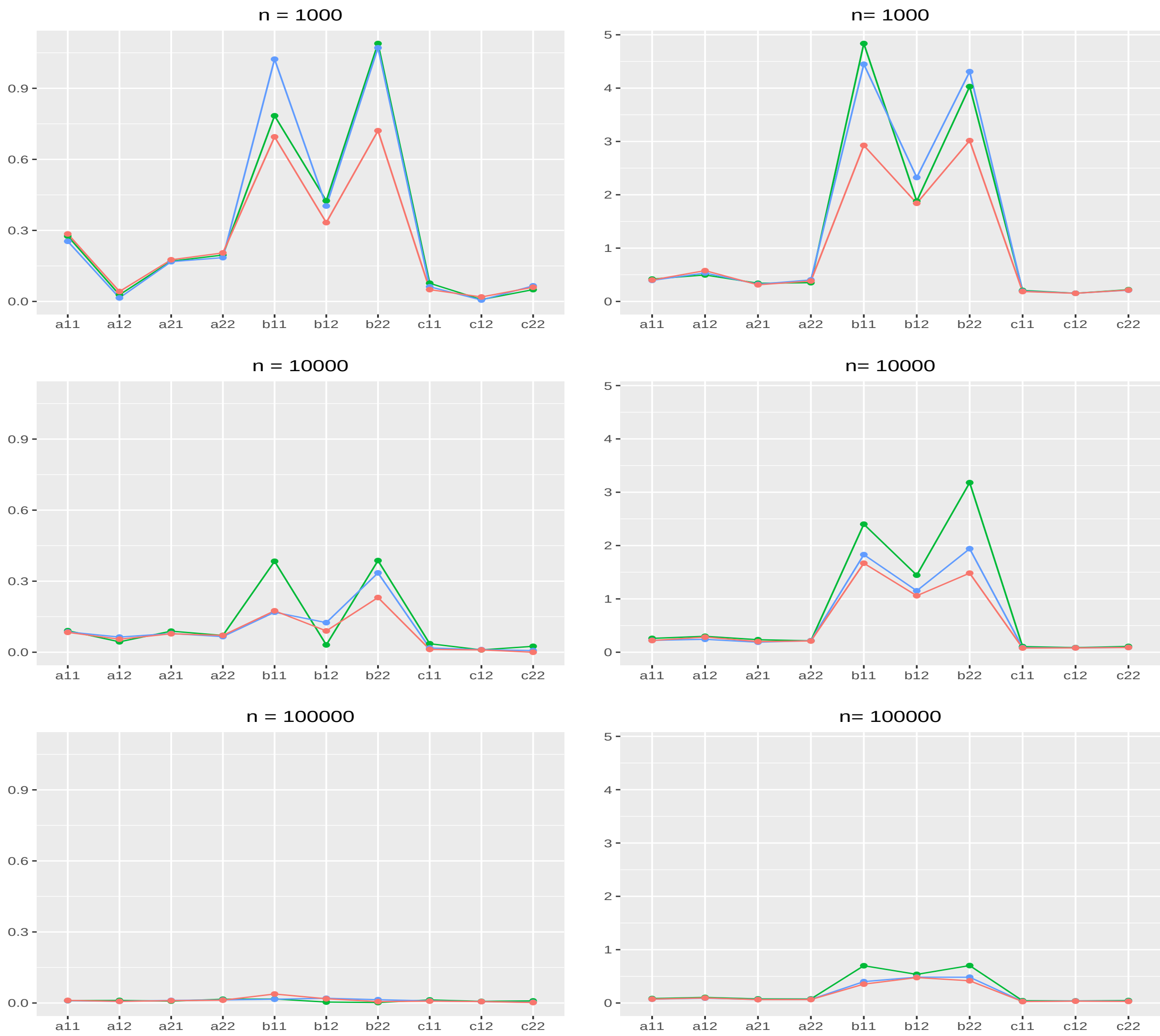

Figure 1: Example 1: Estimated absolute bias (lhs) and standard deviation (rhs) of $\hat{\theta}_{n,r}$. The colors green, blue and red correspond to $r = 2, 5$ and 10, respectively.

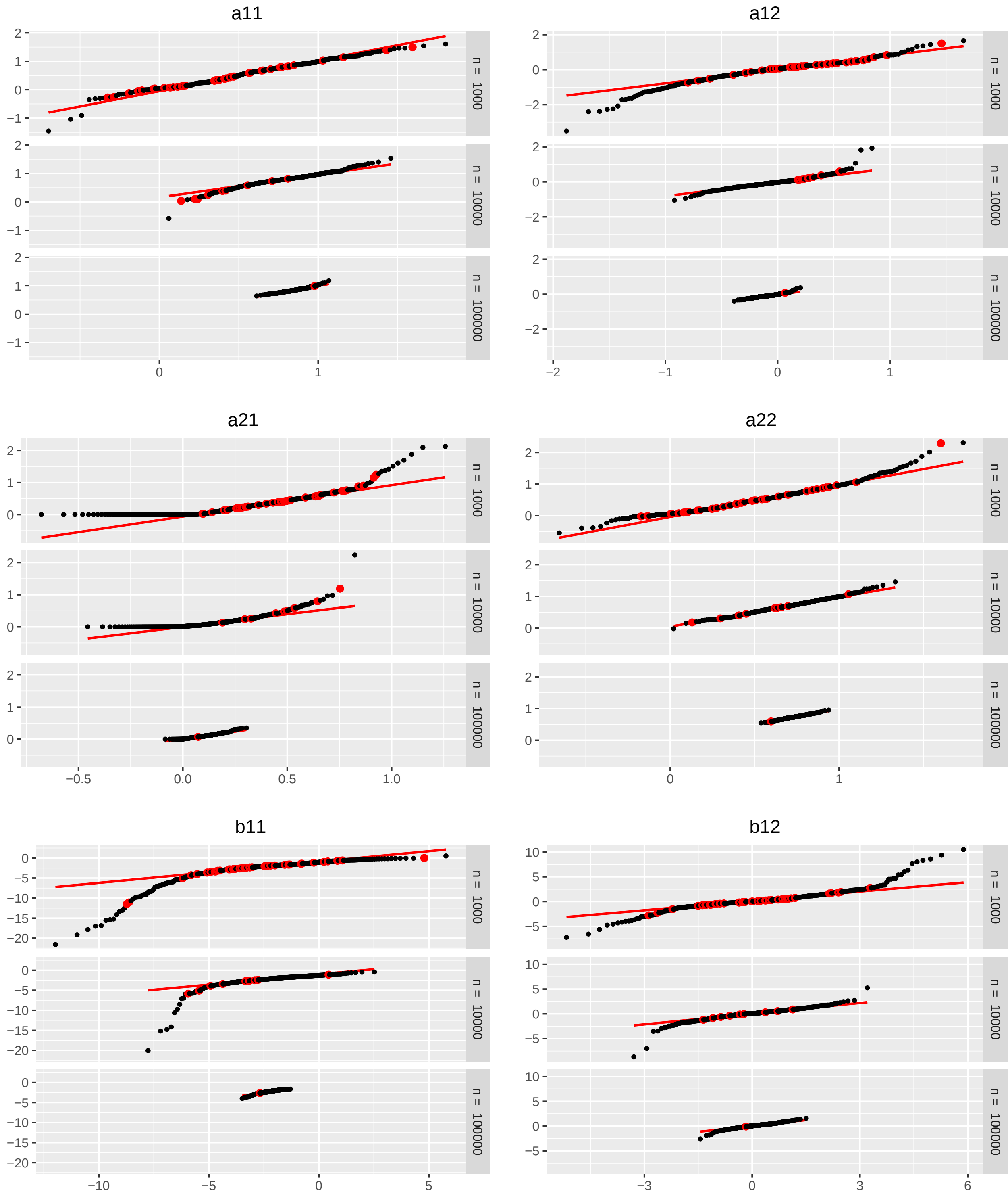

Figure 2: Example 1: Normal QQ-plots of $\hat{\theta}_{n,10}$ for θ_0 as in (6.2). The red dots are values for which the algorithm did not converge.

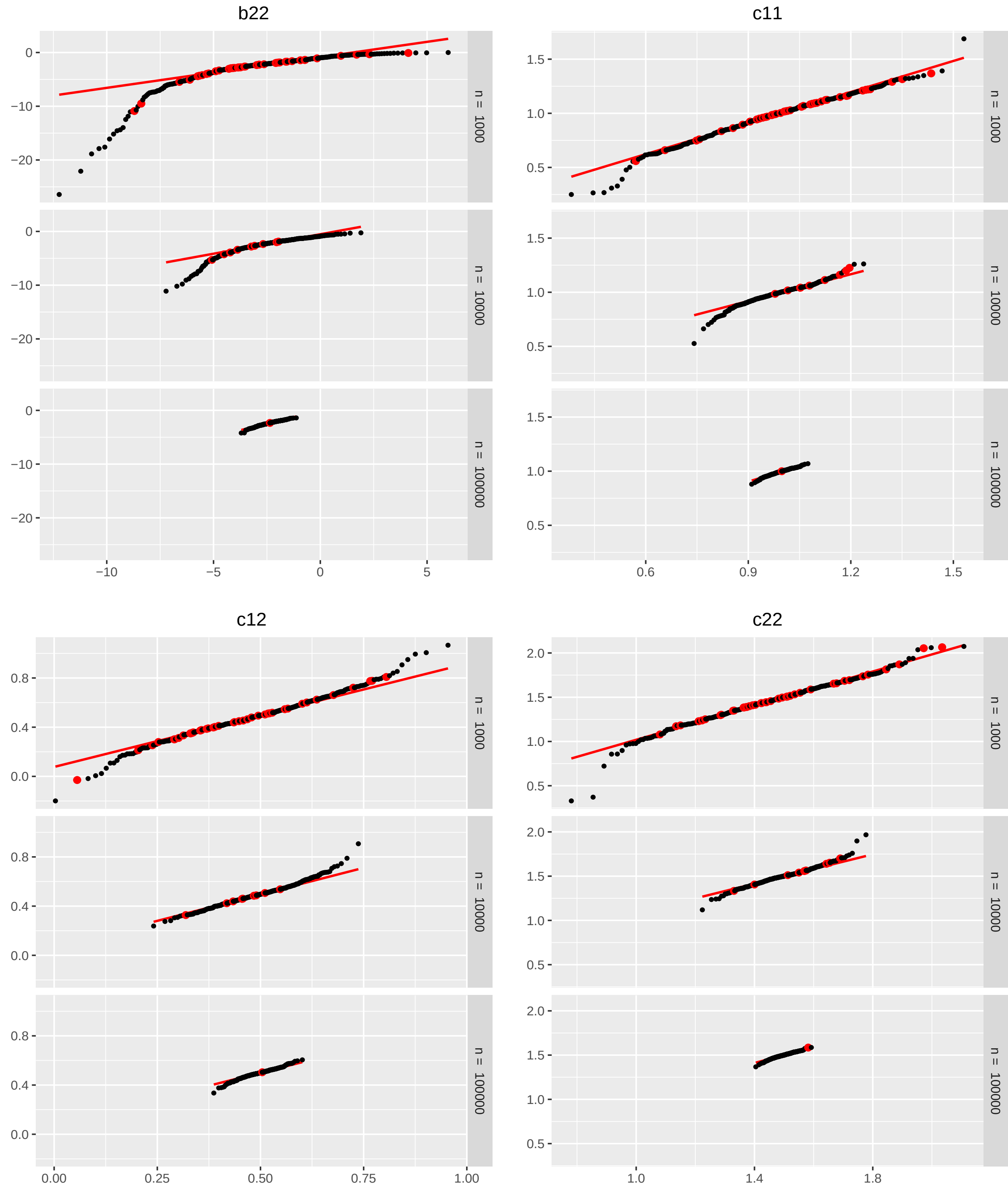

Figure 3: Example 1: Normal QQ-plots of $\hat{\theta}_{n,10}$ for θ_0 as in (6.2). The red dots are values for which the algorithm did not converge.

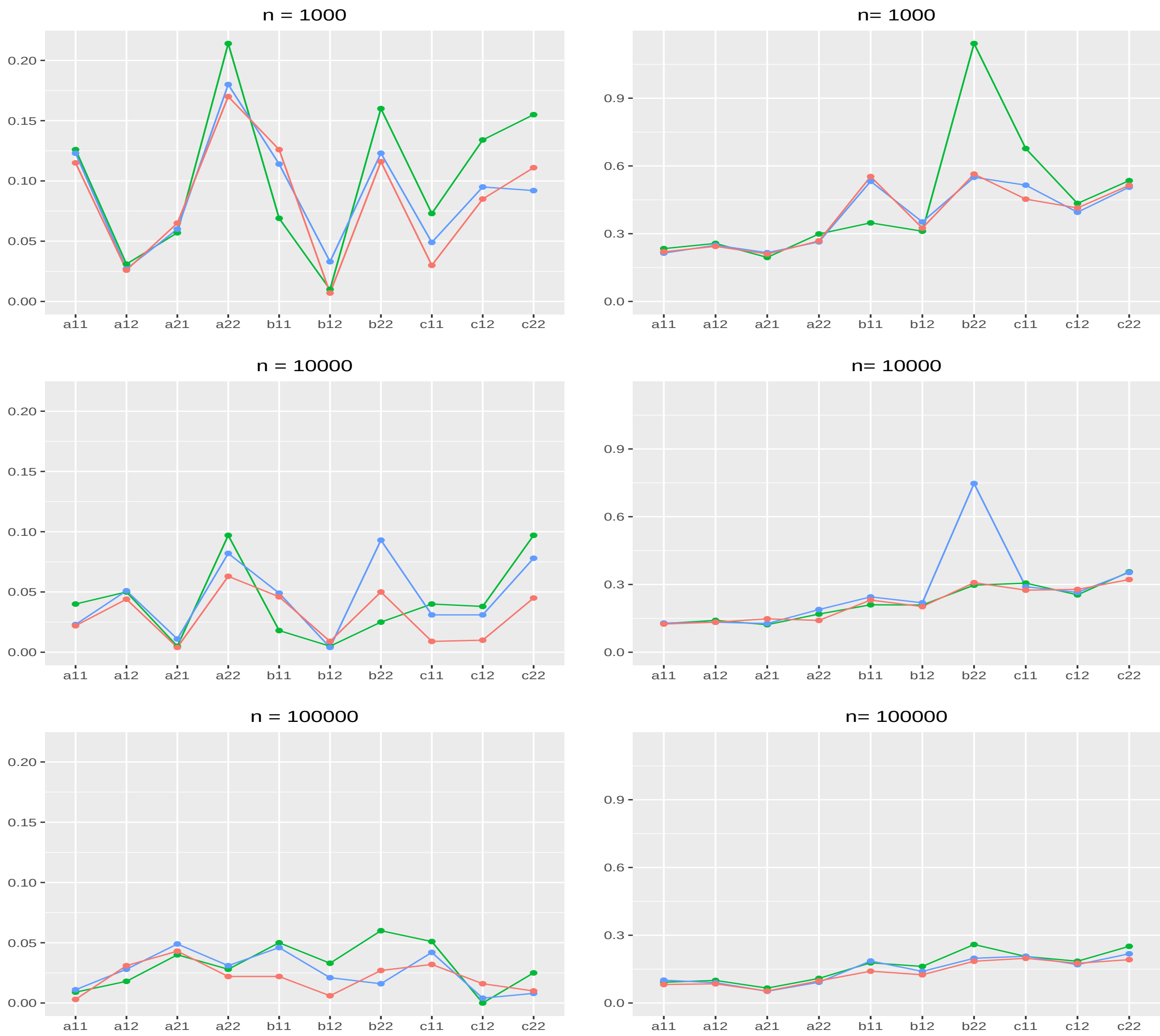

Figure 4: Example 2: Estimated bias and std of $\hat{\theta}_{n,r}$. The colors green, blue and red correspond to $r = 2, 5$ and 10, respectively.

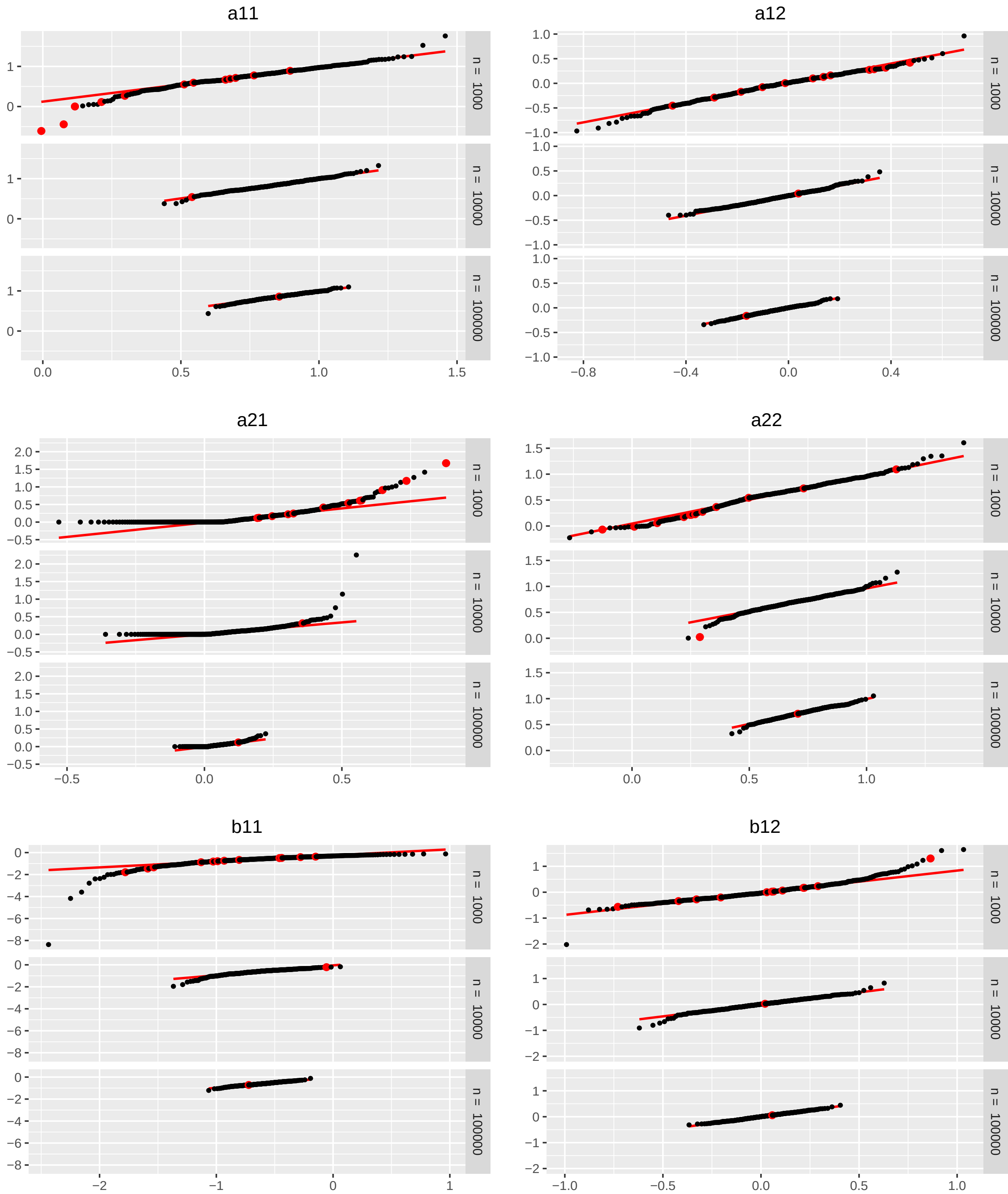

Figure 5: Example 2: Normal QQ-plots of $\hat{\theta}_{n,10}$ for θ_0 as in (6.2). The red dots are values for which the algorithm did not converge.

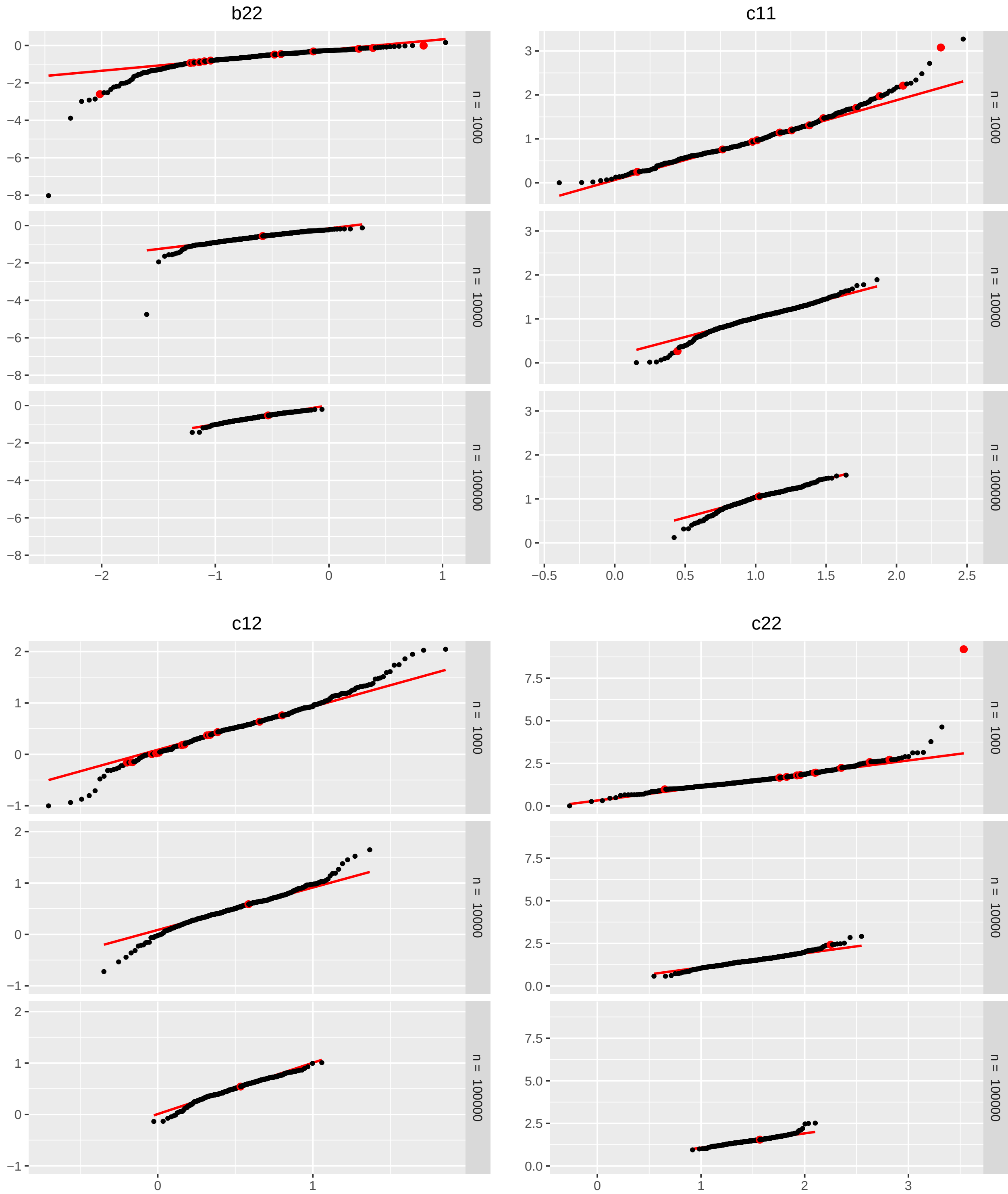

Figure 6: Example 2: Normal QQ-plots of $\hat{\theta}_{n,10}$ for θ_0 as in (6.2). The red dots are values for which the algorithm did not converged.

6.2 Simulation results for Example 6.2

In this section we analyze the behavior of the GMM estimator when the consistency conditions are valid, but we cannot check the conditions for asymptotic normality. Here, we have $\sigma(B_{\theta_0}/4+$

$B^*_{\theta_0}/4 + \sigma_L A^*_{\theta_0} A_{\theta_0}) = \{-0.594, -0.619\} \in (-\infty, 0) + i\mathbb{R}$. Thus, Corollary 5.16 applies and gives weak consistency of the GMM estimator. On the other hand, for $p = 4.001$ the integral in (6.3) is $14.22 > 0$, and thus, we cannot apply Corollary 5.17 to ensure asymptotic normality.

The results for Example 6.2 are given in Figures 4-6. 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}, as observed in the previous example. Also, the estimated bias and std decreases in general as n grows, showing consistency of the estimators. Also, the convergence rate seems slow and, therefore, probably smaller than $n^{1/2}$ (the asymptotic normality rate from Theorem 5.5). The QQ-plots for the estimation of the parameters $A_{(2,1)}$, $C_{(1,1)}$ and $C_{(2,1)}$ also show some deviation from the normal distribution.

7 Real data analysis

In this Section, we fit the MUCOGARCH model to 5 minutes log-returns of stock prices corresponding to the SAP SE and Siemens AG companies (the data was obtained from the Refinitiv EIKON system). For both datasets we have excluded overnight returns. The resulting bivariate dataset has a total length of 12 135 (from 30-jun-2020 to 15-dez-2020) and is shown in Figure 7.

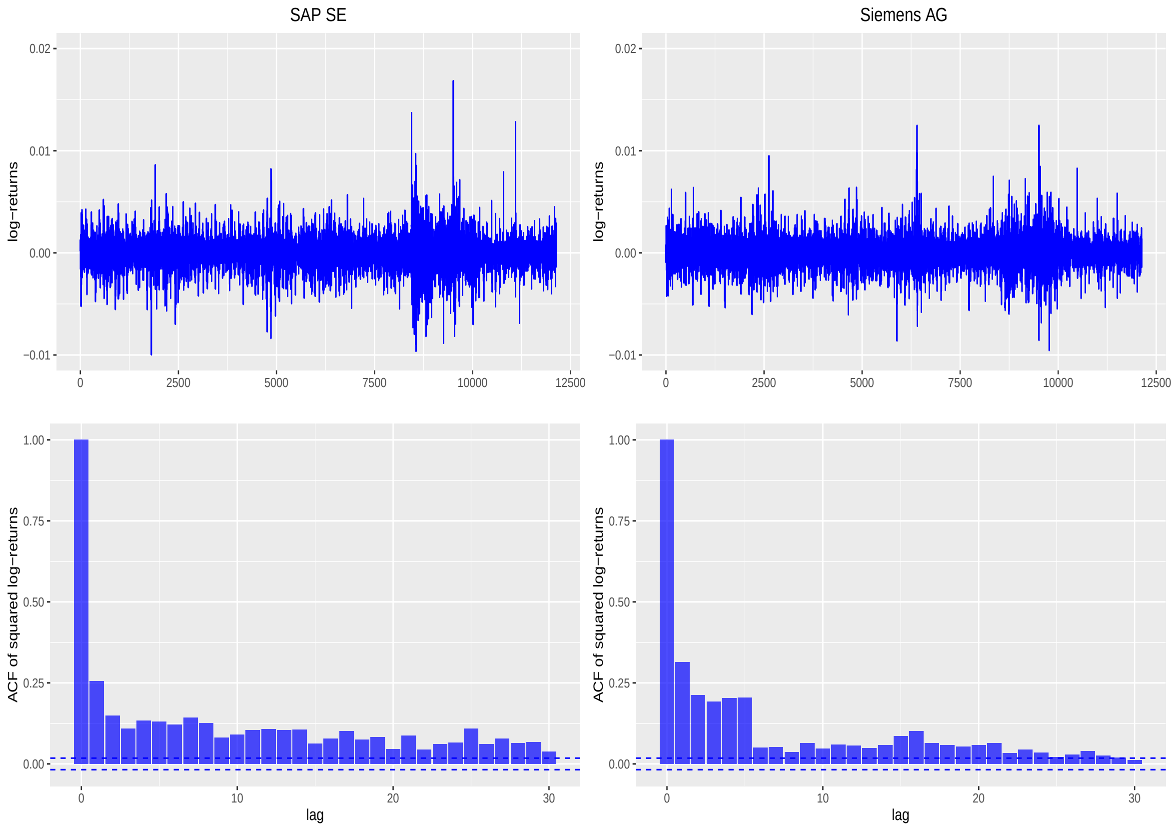

Figure 7: Log-returns (top) and sample acf of squared log-returns (bottom) for SAP SE (left) and Siemens AG (right) exchange rates.

Based on the sample autocorrelation function of the squared log-returns we decided to use 15 lags to estimate a bivariate MUCOGARCH model. Before moving to the estimation step, we re-scaled the data multiplying it by 1000 for numerical reasons. Since the GMM algorithm from (5.3) requires a starting value, we first fit a one-dimensional COGARCH to each dataset with the algorithm described in (Haug et al., 2007, Algorithm 1) and used them to construct the corresponding 2-dimensional MUCOGARCH describing two independent log-return price process. This will be a MUCOGARCH process whose parameters A, B, C are diagonal matrices and the driving Lévy process has independent components (Stelzer, 2010, Example 4.2). To construct a second step GMM estimator, we use the estimated parameters from the first step to estimate a weight matrix to replace Ω in (5.3). Indeed, as explained in (Mátyás, 1999, Section 1.3.3), and given a consistent estimator $\hat{\Sigma}_{\theta_0}$ of Σ_{θ_0}, an asymptotically more efficient estimator can be constructed as

$$\hat{\theta}_n^{(2)} = \underset{\theta\in\Theta}{\arg\min} \left\{ (\hat{k}_{n,r} - k_{\theta,r})^T \hat{\Sigma}_{\theta_0} (\hat{k}_{n,r} - k_{\theta,r}) \right\}. \tag{7.1}$$

Given a bivariate dataset of log-price returns of size n following a MUCOGARCH model, we can see from (5.6) that a natural estimator for Σ_{θ_0} is

$$\begin{aligned}\hat{\Sigma}_{n,M} = &\frac{1}{n}\sum_{t=1}^{n}(D_t - \hat{k}_{n,r})(D_t - \hat{k}_{n,r})^* + \\ &\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{M}\left\{(D_t - \hat{k}_{n,r})(D_{t+i} - \hat{k}_{n,r})^* + (D_{t+i} - \hat{k}_{n,r})^*(D_t - \hat{k}_{n,r})\right\}, \quad M \in \mathbb{N},\end{aligned} \tag{7.2}$$

where we truncated the infinity sum in (5.6). The above estimator is a symmetric matrix, regardless of the values chosen for M and n. As done in the proof of Lemma 5.1, sufficient conditions for it to be consistent are assumptions **a**, **b**, **c** and (d.5). On the other hand, we cannot guarantee that it is going to be positive semidefinite, which is a condition required to use it as a weight matrix in the estimator from (7.1). One way around it is to use the simpler estimator (as in (Stelzer et al., 2015, Section 4.1)) of Σ_{θ_0}, which ignores the second summation on the rhs of (7.2), namely,

$$\hat{\Sigma}_{n,M}^{(\text{basicCOV})} = \frac{1}{n}\sum_{t=1}^{n}(D_t - \hat{k}_{n,r})(D_t - \hat{k}_{n,r})^*, \quad M \in \mathbb{N}, \tag{7.3}$$

The resulting estimators when applying the 2 step GMM with (7.3) for the bivariate dataset of 5min log-returns are given by:

$$A = \begin{pmatrix} 0.442 & 0.259 \\ 0.054 & 0.054 \end{pmatrix}, \quad B = \begin{pmatrix} -0.146 & -0.014 \\ -0.014 & -0.080 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 0.257 & -0.134 \\ -0.134 & 0.070 \end{pmatrix}. \tag{7.4}$$

The parameters of the Levy process were chosen to be the same from example 1 as they allow us to check consistency and asymptotic normality conditions. According to Corollary 5.16, the estimated parameters given in (7.4) describe a MUCOGARCH model for which the GMM estimators is consistent. For the asymptotic normality, we need to verify assumption (f.8), which requires computing $m(p, \theta_0)$ with θ_0 replaced by a vector formed with the entries of the matrices in (7.4). This computation results in 2.98, and therefore (f.8) is violated, and we cannot ensure asymptotic normality.

To assess performance of the GMM estimator used to estimate the MUCOGARCH model for real data, we perform a simulation study using the values from (7.4) as θ_0. The results are

Estimated Bias - GMM

	True Value	1 step	2 step (BasicCOV)	2 step ($M = 10$)	2 step (diagCOV)
$\theta^{(1)}$	0.442	-0.024	-0.094	-0.055	-0.073
$\theta^{(2)}$	0.259	0.016	-0.026	0.040	-0.034
$\theta^{(3)}$	0.054	0.086	0.120	0.121	0.100
$\theta^{(4)}$	0.194	0.164	0.170	0.151	0.169
$\theta^{(5)}$	-0.146	0.010	0.014	-0.043	0.031
$\theta^{(6)}$	-0.014	-0.085	-0.076	-0.091	-0.072
$\theta^{(7)}$	-0.080	-0.081	-0.077	-0.162	-0.058
$\theta^{(8)}$	0.257	0.091	0.129	0.248	0.083
$\theta^{(9)}$	-0.134	0.147	0.148	0.176	0.148
$\theta^{(10)}$	0.070	0.136	0.157	0.315	0.137

Estimated Std - GMM

	True Value	1 step	2 step (BasicCOV)	2 step ($M = 10$)	2 step (diagCOV)
$\theta^{(1)}$	0.442	0.101	0.084	0.207	0.100
$\theta^{(2)}$	0.259	0.123	0.112	0.261	0.112
$\theta^{(3)}$	0.054	0.088	0.086	0.169	0.093
$\theta^{(4)}$	0.194	0.109	0.108	0.193	0.110
$\theta^{(5)}$	-0.146	0.080	0.091	0.402	0.067
$\theta^{(6)}$	-0.014	0.073	0.081	0.398	0.065
$\theta^{(7)}$	-0.080	0.072	0.085	0.650	0.062
$\theta^{(8)}$	0.257	0.112	0.113	0.909	0.168
$\theta^{(9)}$	-0.134	0.093	0.101	0.645	0.139
$\theta^{(10)}$	0.070	0.091	0.103	1.137	0.126

Table 7.1: Estimated bias and std of $\hat{\theta}_{n,r}$ (1 step GMM) and $\hat{\theta}^{(2)}_{n,r}$, the 2 step GMM with 3 different weighting matrices, namely: basicCOV from (7.3), fullCOV from (7.2) with $M = 10$ and diagCOV, a diagonal weighting matrix formed with the diagonal entries of (7.3).

reported in Table 7.1 and referred by basicCOV. We also compare it with the first step GMM and two other 2 step GMM estimators: one using $\hat{\Sigma}_{n,M}$ from (7.2) with $M = 10$ as a weight matrix and another which uses a weighing diagonal matrix formed with the diagonal of $\hat{\Sigma}_{n,M}$. These are referred in Table 7.1 as fullCOV and diagCOV, respectively. Using fullCOV did not improved the estimates when compared with basicCOV and diagCOV, and that might be to do with the fact that inverting the estimated covariance matrix gave several warnings during the estimation. The estimator based on diagCOV was somehow similar to the 1 step GMM, giving smaller bias and std for the parameters $\theta^{(6)} - \theta^{(8)}$, which correspond to the matrix B. The fact that the 2 step GMM did not improved the 1step GMM results, might be because the sample size n used to estimate the covariance matrix is too small.

Acknowledgement

Both authors take pleasure to thank Claudia Klüppelberg for motivating this work and for helpful comments. Thiago do Rêgo Sousa gratefully acknowledges support from the National Council for Scientific and Technological Development (CNPq - Brazil) and the TUM Graduate School.

He also thanks the Institute of Mathematical Finance at Ulm University for its hospitality and takes pleasure to thank Carlos Améndola for helpful discussions.

A Proofs

A.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 (2012). Furthermore, we need the following.

Fact A.1 ((Stelzer, 2010, Lemma 6.9) with drift)**.** *Assume that* $(X_t)_{t\in\mathbb{R}^+}$ *is an adapted cadlag* $M_d(\mathbb{R})$*-valued process satisfying* $\mathbb{E}(\|X_t\|) < \infty$ *for all* $t \in \mathbb{R}^+$*,* $t \mapsto \mathbb{E}(\|X_t\|)$ *is locally bounded and* $(L_t)_{t\in\mathbb{R}_+}$ *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 A.2. *Let* $(A_t)_{t\in\mathbb{R}^+}$ *in* $M_{d^2}(\mathbb{R})$*,* $(B_t)_{t\in\mathbb{R}^+}$ *in* $M_{1,d^2}(\mathbb{R})$ *be adapted caglad processes satisfying* $\mathbb{E}\|A_t\|\|B_t\| < \infty$ *for all* $t \in \mathbb{R}^+$*,* $t \mapsto \mathbb{E}\|A_t\|\|B_t\|$ *is locally bounded and* $(L_t)_{t\in\mathbb{R}_+}$ *be an* $\mathbb{R}^d$ *valued Lévy process satisfying Assumption 5.2 in Stelzer (2010). Then,*

$$\mathbb{E}\int_0^t A_s\mathrm{d}(\mathrm{vec}([L,L]_s))B_s = (\sigma_W + \sigma_L)\int_0^t \mathbb{E}[A_s\,\mathrm{vec}(I_d)B_s]\mathrm{d}s.$$

Proof. First notice that $\mathrm{vec}([L,L]_s)$ is an $\mathbb{R}^{d^2}$-valued Lévy process with finite variation. Then it follows from Fact A.1 that

$$\begin{aligned}
&\mathrm{vec}\left(\mathbb{E}\int_0^t A_s\mathrm{d}(\mathrm{vec}([L,L]_s))B_s\right) = \mathbb{E}\int_0^t (B_s\otimes A_s)\mathrm{d}(\mathrm{vec}([L,L]_s)\\
&= \int_0^t \mathbb{E}(B_s\otimes A_s)\mathbb{E}(\mathrm{vec}([L,L]_1))\mathrm{d}s = (\sigma_W+\sigma_L)\int_0^t \mathbb{E}(B_s\otimes A_s)\,\mathrm{vec}(I_d)\mathrm{d}s\\
&= (\sigma_W+\sigma_L)\,\mathrm{vec}\left(\int_0^t \mathbb{E}(A_sI_dB_s)\mathrm{d}s\right),
\end{aligned}$$

so the result follows by an application of vec^{-1}. □

Proof of Lemma 3.2: It follows from (Stelzer, 2010, Proposition 4.7) (with $k = p$) that $\mathbb{E}\|Y_t\|^p < \infty$ for all $t \in \mathbb{R}^+$ and $t \mapsto \mathbb{E}\|Y_t\|^p$ is locally bounded. Then an application of (Protter, 2005, Theorem 66 of Ch. 5) together with the fact that $\mathbb{E}\|L_1\|^{2p} < \infty$ and the definition of $(V_t)_{t\in\mathbb{R}_+}$ in (3.2) gives for all $t > 0$

$$\mathbb{E}\|G_t\|^{2p} = \mathbb{E}\left\|\int_0^t V_{s-}^{1/2}\mathrm{d}L_s\right\|^{2p} \le c\int_0^t \mathbb{E}\|V_{s-}^{1/2}\|^{2p}\mathrm{d}s \le c\int_0^t \mathbb{E}\|C+Y_{s-}\|^p\mathrm{d}s.$$

Lemma A.1. *Assume that Assumptions* (*a*.1)*-*(*a*.4)*,* $\boldsymbol{b}$ *and* (*c*.2) *hold. Then,*

$$\begin{aligned}
&\mathrm{cov}(\mathrm{vec}(Y_\Delta),\mathrm{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)) = \mathrm{cov}(\mathrm{vec}(Y_\Delta),\mathrm{vec}(G_\Delta G_\Delta^*))\\
&= \mathrm{var}(\mathrm{vec}(V_0))(e^{\mathcal{B}^*\Delta} - I_{d^2})[(\sigma_W+\sigma_L)(\mathcal{B}^*)^{-1} - 2((A\otimes A)^*)^{-1}], \quad \Delta \ge 0.
\end{aligned} \tag{A.1}$$

Proof. Since $(a.4)$, $(a.1)$ and $(c.1)$ hold, we can apply Lemma 3.2 with $p = 2$ 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 (A.1) is finite. Integration by parts formula (Stelzer, 2010, p. 111) gives

$$G_\Delta G_\Delta^* = \int_0^\Delta V_{s-}^{1/2} \mathrm{d}L_s G_{s-}^* + \int_0^\Delta G_{s-} \mathrm{d}L_s^* V_{s-}^{1/2} + \int_0^\Delta V_{s-}^{1/2} \mathrm{d}[L, L^*]_s V_{s-}^{1/2} := A_\Delta + A_\Delta^* + C_\Delta. \quad \text{(A.2)}$$

It follows from Lemma 3.2(a) and (b) with $p = 2$ together with the Cauchy-Schwarz inequality that

$$\int_0^t \mathbb{E}(\|V_{s-}^{1/2}\|_2 \|G_{s-}\|_2)^2 \mathrm{d}s \le \int_0^t \big(\mathbb{E}\|V_{s-}\|_2^2\big)^{1/2} \big(\mathbb{E}\|G_{s-}\|_2^4\big)^{1/2} \mathrm{d}s < \infty, \quad \text{(A.3)}$$

where the finiteness is due to the fact that the integrand is locally bounded, and thus, also bounded on $(0, t)$. Therefore $(A_t)_{t\in\mathbb{R}^+}$ is a martingale and $A_t \in L^2$ for all $t \ge 0$. 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 Behme (2012)) imply

$$\begin{aligned}
&\operatorname{cov}(\operatorname{vec}(Y_\Delta), \operatorname{vec}(A_\Delta)) \\
&= \mathbb{E}\big(\operatorname{vec}(Y_\Delta)(\operatorname{vec}(A_\Delta))^*\big) - \mathbb{E}(\operatorname{vec}(Y_\Delta))\mathbb{E}(\operatorname{vec}(A_\Delta))^* \\
&= \mathbb{E}\bigg(\int_0^\Delta \operatorname{vec}(Y_{s-})\mathrm{d}(\operatorname{vec}(A_s))^* + \int_0^\Delta \mathrm{d}\operatorname{vec}(Y_s)(\operatorname{vec}(A_{s-}))^* + [\operatorname{vec}(Y), (\operatorname{vec}(A))^*]_\Delta\bigg) - 0 \\
&= 0 + \mathbb{E}\int_0^\Delta \mathrm{d}\operatorname{vec}(Y_s)(\operatorname{vec}(A_{s-}))^* + \mathbb{E}([\operatorname{vec}(Y), (\operatorname{vec}(A))^*]_\Delta).
\end{aligned} \quad \text{(A.4)}$$

The first expectation in (A.4) vanishes since

$$\begin{aligned}
&\int_0^\Delta \mathbb{E}\|\operatorname{vec}(Y_{s-})\|^2 \|G_{s-}\|^2 \|V_{s-}^{1/2}\|^2 \mathrm{d}s \\
&\le \int_0^\Delta (\mathbb{E}\|\operatorname{vec}(Y_{s-})\|^4)^{1/2} (\mathbb{E}\|G_{s-}\|^8)^{1/4} (\mathbb{E}\|V_{s-}^{1/2}\|^8)^{1/4} \mathrm{d}s < \infty
\end{aligned}$$

by the generalized Hölder inequality with $(1/2 + 1/4 + 1/4 = 1)$ (see e.g. (Kufner et al., 1977, Theorem 2.1)), Lemma 3.2 and the fact that $(L_t)_{t\in\mathbb{R}_+}$ is an L^2-martingale. Let $\tilde{C} := (B \otimes I + I \otimes B)$ and recall from p. 84 in Stelzer (2010) that

$$\mathrm{d}\operatorname{vec}(Y_s) = \tilde{C}\operatorname{vec}(Y_{s-})\mathrm{d}s + (A \otimes A)(V_{s-}^{1/2} \otimes V_{s-}^{1/2})\mathrm{d}\operatorname{vec}([L, L]_s^{\mathfrak{d}}). \quad \text{(A.5)}$$

Using (A.5), the bilinearity of the quadratic covariation process, (Behme, 2012, eq. (2.1)), Lemma 3.2, Facts A.1, $(a.3)$, (A.3) and the Itô isometry we obtain

$$\begin{aligned}
&[\operatorname{vec}(Y), (\operatorname{vec}(A))^*]_\Delta \\
=& \left[\int_0^\cdot \tilde{C}\operatorname{vec}(Y_{s-})\mathrm{d}s + \int_0^\cdot (A \otimes A)(V_{s-}^{1/2} \otimes V_{s-}^{1/2})\mathrm{d}\operatorname{vec}([L, L]_s^{\mathfrak{d}}), \int_0^\cdot \mathrm{d}L_s^*(G_{s-}^* \otimes V_{s-}^{1/2})\right]_\Delta \\
=& \int_0^\Delta (A \otimes A)(V_{s-}^{1/2} \otimes V_{s-}^{1/2})\mathrm{d}[\operatorname{vec}([L, L]^{\mathfrak{d}}), L^*]_s (G_{s-}^* \otimes V_{s-}^{1/2}).
\end{aligned} \quad \text{(A.6)}$$

Recall that for arbitrary matrices $M \in M_{m,n}(\mathbb{R})$ and $N \in M_{k,l}(\mathbb{R})$ it holds $\|A \otimes B\|_2 = \|A\|_2 \|B\|_2$ (Bernstein, 2009, Fact 9.9.61). This together with the Hölder inequality with $(3/4 + 1/4 = 1)$

and Lemma 3.2 with $p = 4$ gives

$$\int_0^\Delta \mathbb{E}\|V_{s-}^{1/2} \otimes V_{s-}^{1/2}\|_2 \|(G_{s-}^* \otimes V_{s-}^{1/2})\|_2 \mathrm{d}s = \int_0^\Delta \mathbb{E}\|V_{s-}^{1/2}\|_2^3 \|G_{s-}^*\|_2 \mathrm{d}s$$
$$= \int_0^\Delta \mathbb{E}\|V_{s-}\|_2^{3/2} \|G_{s-}^*\|_2 \mathrm{d}s \le \int_0^\Delta (\mathbb{E}\|V_{s-}\|_2^2)^{3/4} (\mathbb{E}\|G_{s-}^*\|_2^4)^{1/4} \mathrm{d}s < \infty.$$

Thus, applying expectations at both sides of (A.6) gives

$$E[\mathrm{vec}(Y), (\mathrm{vec}(A))^*]_\Delta = 0. \tag{A.7}$$

Let $l_s := \mathbb{E}\,\mathrm{vec}(Y_s)(\mathrm{vec}(A_s))^*$ and notice that it follows from Lemma 3.2 and the Cauchy-Schwarz inequality that $\mathbb{E}\|l_s\| < \infty$ and $s \mapsto \mathbb{E}\|l_s\|$ is locally bounded. Use (A.5), (A.7), the compensation formula, (Bernstein, 2009, Proposition 7.1.9), $\mathbb{E}\,\mathrm{vec}(V_s)\,\mathrm{vec}(A_{s-}) = l_s$, the Itô isometry and $(a.2)$ to get

$$\begin{aligned}
l_\Delta &= \mathbb{E}\int_0^\Delta \mathrm{d}\,\mathrm{vec}(Y_s)(\mathrm{vec}(A_{s-}))^* \\
&= \mathbb{E}\int_0^\Delta \left[\tilde{C}\,\mathrm{vec}(Y_{s-})\mathrm{d}s + (A\otimes A)(V_{s-}^{1/2}\otimes V_{s-}^{1/2})\mathrm{d}(\mathrm{vec}([L,L]_s^{\mathfrak{d}})\right](\mathrm{vec}(A_{s-}))^* \\
&= \tilde{C}\int_0^\Delta \mathbb{E}\,\mathrm{vec}(Y_{s-})(\mathrm{vec}(A_{s-}))^*\mathrm{d}s \\
&\quad + \sigma_L \int_0^\Delta \mathbb{E}[(A\otimes A)(V_{s-}^{1/2}\otimes V_{s-}^{1/2})\,\mathrm{vec}(I_d)(\mathrm{vec}(A_{s-}))^*]\mathrm{d}s \\
&= (\tilde{C} + \sigma_L(A\otimes A))\int_0^\Delta l_s \mathrm{d}s.
\end{aligned} \tag{A.8}$$

Solving the matrix-valued integral equation in (A.8) and using that $A_0 = 0$ implies $l_0 = 0$, gives $l_s = 0$ for all $s \ge 0$ (see Haug et al. (2007)). Thus, it follows from (A.4)-(A.8) that

$$\mathrm{cov}(\mathrm{vec}(Y_\Delta), \mathrm{vec}(A_\Delta)) = 0, \tag{A.9}$$

and, as a consequence $\mathrm{cov}(\mathrm{vec}(Y_\Delta), \mathrm{vec}(A_\Delta^*)) = 0$. Let $\mathcal{V}_{s-} := V_{s-}^{1/2} \otimes V_{s-}^{1/2}$. Then,

$$\begin{aligned}
\mathrm{vec}(C_\Delta) &= \int_0^\Delta \mathcal{V}_{s-}\mathrm{d}\,\mathrm{vec}([L,L^*]_s) = \int_0^\Delta \mathcal{V}_{s-}\mathrm{d}\,\mathrm{vec}([L,L^*]_s^{\mathfrak{d}}) \\
&+\sigma_W \int_0^\Delta (V_{s-}^{1/2}\otimes V_{s-}^{1/2})\,\mathrm{vec}(I_d)\mathrm{d}s = \int_0^\Delta \mathcal{V}_{s-}\mathrm{d}\,\mathrm{vec}([L,L^*]_s^{\mathfrak{d}}) + \sigma_W \int_0^\Delta \mathrm{vec}(V_{s-})\mathrm{d}s.
\end{aligned} \tag{A.10}$$

Using the compensation formula, Fact A.1 and the stationarity of $(V_s)_{s\in\mathbb{R}_+}$ we get

$$\mathbb{E}\int_0^\Delta \mathcal{V}_{s-}\mathrm{d}\,\mathrm{vec}([L,L^*]_s) = (\sigma_W + \sigma_L)\int_0^\Delta \mathbb{E}\mathcal{V}_{s-}\,\mathrm{vec}(I_d)\mathrm{d}s = \Delta(\sigma_W+\sigma_L)\mathbb{E}\,\mathrm{vec}(V_0). \tag{A.11}$$

Additionally, it follows from Lemma 3.2 that $\mathbb{E}\|\,\mathrm{vec}(V_s)\,\mathrm{vec}(Y_\Delta)^*\| < \infty$ for all $s \ge 0$ and that $s \mapsto \mathbb{E}\|\,\mathrm{vec}(V_s)\,\mathrm{vec}(Y_\Delta)^*\|$ is locally bounded. Then,

$$\begin{aligned}
\mathbb{E}\Big(\int_0^\Delta \mathrm{vec}(V_{s-})\mathrm{d}s\,(\mathrm{vec}(Y_\Delta))^*\Big) &= \int_0^\Delta \mathbb{E}\,\mathrm{vec}(V_{s-})(\mathrm{vec}(Y_\Delta))^*\mathrm{d}s \\
&= \Delta\,\mathrm{vec}(C)(\mathbb{E}\,\mathrm{vec}(Y_0))^* + \int_0^\Delta \mathbb{E}\,\mathrm{vec}(Y_s)(\mathrm{vec}(Y_\Delta))^*\mathrm{d}s.
\end{aligned}$$

Now it follows from the invertibility of $(A \otimes A)$ and from the second equation following (3.5) in Stelzer (2010) that

$$\begin{aligned}&\int_0^\Delta \mathcal{V}_{s-} \mathrm{d}\,\mathrm{vec}([L, L^*]_s^{\mathfrak{d}})\\ &= (A \otimes A)^{-1}\Big(\mathrm{vec}(Y_\Delta) - \mathrm{vec}(Y_0) - \int_0^\Delta (B \otimes I + I \otimes B)\,\mathrm{vec}(Y_{s-})\mathrm{d}s\Big).\end{aligned} \tag{A.12}$$

The representation in (A.12) gives

$$\begin{aligned}&\mathbb{E}\Big[\Big(\int_0^\Delta \mathcal{V}_{s-} \mathrm{d}\,\mathrm{vec}([L, L^*]_s^{\mathfrak{d}})\Big)(\mathrm{vec}(Y_\Delta))^*\Big]\\ =\ &\mathbb{E}\Big[(A \otimes A)^{-1}\Big(\mathrm{vec}(Y_\Delta) - \mathrm{vec}(Y_0) - \int_0^\Delta (B \otimes I + I \otimes B)\,\mathrm{vec}(Y_{s-})\mathrm{d}s\Big)(\mathrm{vec}(Y_\Delta))^*\Big]\\ =\ &(A \otimes A)^{-1}\Big[\mathbb{E}\,\mathrm{vec}(Y_\Delta)(\mathrm{vec}(Y_\Delta))^* - \mathbb{E}\,\mathrm{vec}(Y_0)(\mathrm{vec}(Y_\Delta))^*\\ &\quad -(B \otimes I + I \otimes B)\int_0^\Delta \mathbb{E}\,\mathrm{vec}(Y_{s-})(\mathrm{vec}(Y_\Delta))^*\mathrm{d}s\Big].\end{aligned} \tag{A.13}$$

Using the definition of C_Δ in (A.2), together with (A.10), (A.11) and (A.13) gives

$$\begin{aligned}&\mathrm{cov}(\mathrm{vec}(C_\Delta), \mathrm{vec}(Y_\Delta)) = (A \otimes A)^{-1}\Big[\mathbb{E}\,\mathrm{vec}(Y_\Delta)(\mathrm{vec}(Y_\Delta))^* - \mathbb{E}\,\mathrm{vec}(Y_0)(\mathrm{vec}(Y_\Delta))^*\\ &-(B \otimes I + I \otimes B)\Big(\int_0^\Delta \mathbb{E}\,\mathrm{vec}(Y_{s-})(\mathrm{vec}(Y_\Delta))^*\mathrm{d}s\Big)\Big]\\ &+\Delta\sigma_W\,\mathrm{vec}(C)(\mathbb{E}\,\mathrm{vec}(Y_0))^* + \sigma_W\int_0^\Delta \mathbb{E}\,\mathrm{vec}(Y_{s-})(\mathrm{vec}(Y_\Delta))^*\mathrm{d}s\\ &-\Delta(\sigma_W + \sigma_L)\mathbb{E}\,\mathrm{vec}(V_0)\mathbb{E}(\mathrm{vec}(Y_\Delta))^*\\ &= [\sigma_W I_{d^2} - (A \otimes A)^{-1}(B \otimes I + I \otimes B)]\int_0^\Delta \mathbb{E}\,\mathrm{vec}(Y_s)(\mathrm{vec}(Y_\Delta))^*\mathrm{d}s\\ &+(A \otimes A)^{-1}\big[\mathrm{var}(\mathrm{vec}(Y_0)) - \mathrm{cov}(\mathrm{vec}(Y_0), \mathrm{vec}(Y_\Delta))\big] - \Delta\sigma_L\,\mathrm{vec}(C)\mathbb{E}(\mathrm{vec}(Y_0))^*\\ &-\Delta(\sigma_W + \sigma_L)\mathbb{E}\,\mathrm{vec}(Y_0)\mathbb{E}(\mathrm{vec}(Y_0))^*,\end{aligned}$$

where the last equality follows from $V_0 = C + Y_0$ and the stationarity of $(Y_s)_{s\in\mathbb{R}_+}$. Using (4.2) it follows first that

$$\begin{aligned}\int_0^\Delta \mathbb{E}\,\mathrm{vec}(Y_s)(\mathrm{vec}(Y_\Delta))^*\mathrm{d}s &= \int_0^\Delta e^{\mathcal{B}(\Delta-s)}\mathrm{var}(\mathrm{vec}(Y_0))\mathrm{d}s + \Delta\mathbb{E}\,\mathrm{vec}(Y_0)\mathbb{E}(\mathrm{vec}(Y_0))^*\\ &= \mathcal{B}^{-1}(e^{\mathcal{B}\Delta} - I_{d^2})\mathrm{var}(\mathrm{vec}(Y_0)) + \Delta\mathbb{E}\,\mathrm{vec}(Y_0)\mathbb{E}(\mathrm{vec}(Y_0))^*,\end{aligned} \tag{A.14}$$

and second that

$$\mathrm{var}(\mathrm{vec}(Y_0)) - \mathrm{cov}(\mathrm{vec}(Y_0), \mathrm{vec}(Y_\Delta)) = -(e^{\mathcal{B}\Delta} - I_{d^2})\mathrm{var}(\mathrm{vec}(Y_0)). \tag{A.15}$$

Substituting $B \otimes I + I \otimes B = \mathcal{B} - \sigma_L(A \otimes A)$, using (A.14), (A.15), (4.1) and the formula for

$\mathbb{E}\,\mathrm{vec}(Y_0)$ in (4.2) gives

$$
\begin{aligned}
&\mathrm{cov}(\mathrm{vec}(C_\Delta), \mathrm{vec}(Y_\Delta)) \\
&= \big[\sigma_W I_{d^2} - (A\otimes A)^{-1}(\mathcal{B} - \sigma_L(A\otimes A))\big]\big[\mathcal{B}^{-1}(e^{\mathcal{B}\Delta} - I_{d^2})\mathrm{var}(\mathrm{vec}(Y_0)) \\
&+ \Delta\mathbb{E}\,\mathrm{vec}(Y_0)\mathbb{E}(\mathrm{vec}(Y_0))^*\big] \\
&- (A\otimes A)^{-1}(e^{\mathcal{B}\Delta} - I_{d^2})\mathrm{var}(\mathrm{vec}(Y_0)) - \Delta\sigma_L\,\mathrm{vec}(C)\mathbb{E}(\mathrm{vec}(Y_0))^* \\
&- \Delta(\sigma_W + \sigma_L)\mathbb{E}\,\mathrm{vec}(Y_0)\mathbb{E}(\mathrm{vec}(Y_0))^* \\
&= \big[(\sigma_W+\sigma_L)\mathcal{B}^{-1} - 2(A\otimes A)^{-1}\big](e^{\mathcal{B}\Delta} - I_{d^2})\mathrm{var}(\mathrm{vec}(Y_0)) \\
&\quad - \big[(A\otimes A)^{-1}\mathcal{B}\mathbb{E}\,\mathrm{vec}(Y_0) + \sigma_L\,\mathrm{vec}(C)\big]\Delta\mathbb{E}(\mathrm{vec}(Y_0))^* \\
&= \big[(\sigma_W+\sigma_L)\mathcal{B}^{-1} - 2(A\otimes A)^{-1}\big](e^{\mathcal{B}\Delta} - I_{d^2})\mathrm{var}(\mathrm{vec}(Y_0)) \\
&\quad - \big[(A\otimes A)^{-1}\mathcal{B}(-\sigma_L\mathcal{B}^{-1}(A\otimes A)\,\mathrm{vec}(C)) + \sigma_L\,\mathrm{vec}(C)\big]\Delta\mathbb{E}(\mathrm{vec}(Y_0))^* \\
&= \big[(\sigma_W+\sigma_L)\mathcal{B}^{-1} - 2(A\otimes A)^{-1}\big](e^{\mathcal{B}\Delta} - I_{d^2})\mathrm{var}(\mathrm{vec}(Y_0)).
\end{aligned}
\tag{A.16}
$$

Finally, the result of the Lemma follows from (A.2), (A.9), (A.16) and the fact that

$$\mathrm{cov}(\mathrm{vec}(Y_\Delta), \mathrm{vec}(G_\Delta G_\Delta^*)) = (\mathrm{cov}(\mathrm{vec}(G_\Delta G_\Delta^*), \mathrm{vec}(Y_\Delta)))^* = (\mathrm{cov}(\mathrm{vec}(C_\Delta), \mathrm{vec}(Y_\Delta)))^*.$$

□

A.2 Proof of Lemma 4.2

(i) The proof of Lemma 4.2 (i) follows directly from Lemma A.1 combined with (5.7) in Stelzer (2010).
(ii) Denoting by $\|\cdot\|_F$ the Frobenius norm we have by Lemma 3.2(b) with $p = 2$

$$
\begin{aligned}
\mathbb{E}\|\,\mathrm{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\,\mathrm{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)^*\|_2 &= \mathbb{E}\|\,\mathrm{vec}(\boldsymbol{G}_1\boldsymbol{G}_1^*)\|_2^2 = \mathbb{E}\|\boldsymbol{G}_1\boldsymbol{G}_1^*\|_F^2 \\
&= tr(\boldsymbol{G}_1\boldsymbol{G}_1^*\boldsymbol{G}_1\boldsymbol{G}_1^*) = \mathbb{E}\|\boldsymbol{G}_1\|_2^4 < \infty.
\end{aligned}
$$

Let $a_s := \mathrm{vec}(G_sG_s^*)$, $s \in [0,\Delta]$ and use the integration by parts formula to write

$$
\begin{aligned}
a_\Delta a_\Delta^* &= \int_0^\Delta a_{s-}\mathrm{d}(a_s^*) + \int_0^\Delta \mathrm{d}a_s(a_{s-}^*) + [a,a^*]_\Delta \\
&= \Big(\int_0^\Delta \mathrm{d}a_s(a_{s-}^*)\Big)^* + \int_0^\Delta \mathrm{d}a_s(a_{s-}^*) + [a,a^*]_\Delta
\end{aligned}
\tag{A.17}
$$

hence we only need to prove that the random variables

$$\int_0^\Delta \mathrm{d}a_s(a_{s-}^*) \quad \text{and} \quad [a,a^*]_\Delta,$$

have finite expectations and compute them in closed form. From (A.2), Lemma 2.1(vi) in Behme (2012) and the symmetry of $(V_t)_{t\in\mathbb{R}_+}$ it follows that

$$
\begin{aligned}
\mathrm{d}a_t &= \mathrm{d}(\mathrm{vec}(G_tG_t^*)) \\
&= \mathrm{d}\Big(\mathrm{vec}\Big(\int_0^t V_{s-}^{1/2}\mathrm{d}L_sG_{s-}^* + \int_0^t G_{s-}\mathrm{d}L_s^*V_{s-}^{1/2} + \int_0^t V_{s-}^{1/2}\mathrm{d}[L,L^*]_sV_{s-}^{1/2}\Big)\Big) \\
&= \mathrm{d}\Big(\int_0^t (G_{s-}\otimes V_{s-}^{1/2})\mathrm{d}L_s + \int_0^t (V_{s-}^{1/2}\otimes G_{s-})\mathrm{d}L_s + \int_0^t (V_{s-}^{1/2}\otimes V_{s-}^{1/2})\mathrm{d}\,\mathrm{vec}([L,L^*]_s)\Big) \\
&= (G_{t-}\otimes V_{t-}^{1/2} + V_{t-}^{1/2}\otimes G_{t-})\mathrm{d}L_t + (V_{t-}^{1/2}\otimes V_{t-}^{1/2})\mathrm{d}\,\mathrm{vec}([L,L^*]_t), \quad t\geq 0.
\end{aligned}
\tag{A.18}
$$

By the sub-multiplicative property of $\|\cdot\|_2$, the generalized Hölder inequality with $(1/4+1/4+1/2=1)$ we have

$$\int_0^\Delta \mathbb{E}\|G_{s-}\otimes V_{s-}^{1/2}\|_2^2\|a_{s-}\|_2^2\mathrm{d}s = \int_0^\Delta \mathbb{E}\|G_{s-}\|_2^2\|V_{s-}^{1/2}\|_2^2\|\operatorname{vec}(G_{s-}G_{s-}^*)\|_2^2\mathrm{d}s \tag{A.19}$$
$$= \int_0^\Delta \mathbb{E}\|G_{s-}\|_2^2\|V_{s-}^{1/2}\|_2^2\|G_{s-}G_{s-}^*\|_2^2\mathrm{d}s \leq \int_0^\Delta (\mathbb{E}\|G_{s-}\|_2^8)^{1/4}(\mathbb{E}\|V_{s-}\|_2^4)^{1/4}(\mathbb{E}\|G_{s-}\|_2^8)^{1/2}\mathrm{d}s,$$

which is finite by Lemma 3.2 with $p=4$. Additionally, similar calculations and Lemma 3.2 with $p=2$ shows that $\mathbb{E}(\|V_{s-}^{1/2}\otimes V_{s-}^{1/2}\|_2)\|a_{s-}\|_2 \leq (\mathbb{E}\|V_{s-}\|_2^2)^{1/2}(\mathbb{E}\|G_{s-}\|_2^4)^{1/2} < \infty$ for all $s>0$ and the map $s\mapsto \mathbb{E}(\|V_{s-}^{1/2}\otimes V_{s-}^{1/2}\|_2\|a_{s-}\|_2)$ is locally bounded. Thus it follows from (A.18), the Itô isometry, the fact that $[L,L^*]_t = [L,L^*]_t^{\mathfrak{d}} + \sigma_w I_d t$ and fact A.2 that

$$\begin{aligned}
&\mathbb{E}\int_0^\Delta \mathrm{d}a_s(a_{s-}^*)\\
&= \mathbb{E}\Big(\int_0^\Delta (G_{s-}\otimes V_{s-}^{1/2} + V_{s-}^{1/2}\otimes G_{s-})\mathrm{d}L_s a_{s-}^* + \int_0^\Delta (V_{s-}^{1/2}\otimes V_{s-}^{1/2})\mathrm{d}(\operatorname{vec}([L,L^*]_s)a_{s-}^*\Big)\\
&= (\sigma_L+\sigma_W)\Big(\int_0^\Delta \mathbb{E}\big((V_{s-}^{1/2}\otimes V_{s-}^{1/2})\operatorname{vec}(I_d)a_{s-}^*\big)\mathrm{d}s\Big)\\
&= (\sigma_L+\sigma_W)\int_0^\Delta \mathbb{E}(\operatorname{vec}(V_{s-})a_{s-}^*)\mathrm{d}s.
\end{aligned} \tag{A.20}$$

It follows from (5.6) in Stelzer (2010) that

$$\int_0^\Delta \mathbb{E}a_{s-}^*\mathrm{d}s = \int_0^\Delta \big(\operatorname{vec}((\sigma_L+\sigma_W)s\mathbb{E}V_0)\big)^*\mathrm{d}s = \frac{1}{2}(\sigma_L+\sigma_W)\Delta^2\mathbb{E}\operatorname{vec}(V_0)^*. \tag{A.21}$$

Since we assumed here that all hypothesis for using Lemma A.1 are valid, we can use (A.1) with $\Delta = s$ to get

$$\begin{aligned}
&\int_0^\Delta \operatorname{cov}(\operatorname{vec}(Y_{s-}), a_{s-})\mathrm{d}s\\
&= \operatorname{var}(\operatorname{vec}(Y_0))\Big(\int_0^\Delta (e^{\mathcal{B}^*s} - I_{d^2})\mathrm{d}s\Big)\big[(\sigma_W+\sigma_L)(\mathcal{B}^*)^{-1} - 2((A\otimes A)^*)^{-1}\big]\\
&= \operatorname{var}(\operatorname{vec}(Y_0))\tilde{\mathcal{B}},
\end{aligned} \tag{A.22}$$

where $\tilde{\mathcal{B}}$ is defined in (4.7). Using (A.20), $(c.1)$ (A.21), (A.22) gives

$$\begin{aligned}
\int_0^\Delta \mathbb{E}\operatorname{vec}(V_{s-})a_{s-}^*\mathrm{d}s &= \int_0^\Delta \operatorname{cov}(\operatorname{vec}(V_{s-}), a_{s-})\mathrm{d}s + (\mathbb{E}\operatorname{vec}(V_s))\int_0^\Delta \mathbb{E}(a_{s-}^*)\mathrm{d}s\\
&= \int_0^\Delta \operatorname{cov}(\operatorname{vec}(Y_{s-}), a_{s-})\mathrm{d}s + (\mathbb{E}\operatorname{vec}(V_0))\int_0^\Delta \mathbb{E}(a_{s-}^*)\mathrm{d}s\\
&= \frac{1}{2}(\sigma_L+\sigma_W)\Delta^2\mathbb{E}\operatorname{vec}(V_0)E\operatorname{vec}(V_0)^* + \operatorname{var}(\operatorname{vec}(Y_0))\tilde{\mathcal{B}}\\
&= (\sigma_L+\sigma_W)^{-1}D,
\end{aligned} \tag{A.23}$$

where D is defined in (4.6). 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 (A.2), Lemma 2.1(vi) in Behme (2012) and the symmetry of $V_{s-}^{1/2}$ gives

$$
\begin{aligned}
&[a, a^*]_\Delta \\
&= \Bigg[\operatorname{vec}\Bigg(\int_0^\cdot V_{s-}^{1/2} \mathrm{d}L_s G_{s-}^* + \int_0^\cdot G_{s-} \mathrm{d}L_s^* V_{s-}^{1/2} + \int_0^\cdot V_{s-}^{1/2} \mathrm{d}[L, L^*]_s V_{s-}^{1/2} \Bigg), \\
&\Bigg(\operatorname{vec}\Bigg(\int_0^\cdot V_{s-}^{1/2} \mathrm{d}L_s G_{s-}^* + \int_0^\cdot G_{s-} \mathrm{d}L_s^* V_{s-}^{1/2} + \int_0^\cdot V_{s-}^{1/2} \mathrm{d}[L, L^*]_s V_{s-}^{1/2} \Bigg) \Bigg)^* \Bigg]_\Delta \\
&= \Bigg[\int_0^\cdot f_{s-} \mathrm{d}L_s + \int_0^\cdot \mathcal{V}_{s-} \mathrm{d} \operatorname{vec}([L, L^*]_s), \int_0^\cdot \mathrm{d}L_s^* f_{s-}^* + \int_0^\cdot \mathrm{d}(\operatorname{vec}([L, L^*]_s)^*) \mathcal{V}_{s-} \Bigg]_\Delta \\
&= \int_0^\Delta f_{s-} \mathrm{d}[L, L^*]_s f_{s-}^* + \int_0^\Delta f_{s-} \mathrm{d}[L, \operatorname{vec}([L, L^*])^*]_s \mathcal{V}_{s-} \\
&+ \int_0^\Delta \mathcal{V}_{s-} \mathrm{d}[\operatorname{vec}([L, L^*]), L^*]_s f_{s-}^* + \int_0^\Delta \mathcal{V}_{s-} \mathrm{d}[\operatorname{vec}([L, L^*]), \operatorname{vec}([L, L^*])^*] \mathcal{V}_{s-} \\
&:= I_1 + I_2 + I_3 + I_4.
\end{aligned}
\tag{A.24}
$$

By Lemma 3.2 with $p = 2$ and similar calculations as in (A.19) it follows that $\mathbb{E}\|\mathcal{V}_{s-}\|\|f_{s-}\| < \infty$ for all $s > 0$ and the map $s \mapsto \mathbb{E}\|\mathcal{V}_{s-}\|\|f_{s-}\|$ is locally bounded. Thus, it follows from $(a.3)$ that we have $\mathbb{E}I_2 = \mathbb{E}I_3 = 0$. Now, Lemma 3.2 gives $\mathbb{E}\|\mathcal{V}_{s-}\|^2 < \infty$ for all $s > 0$ and local boundedness of the map $s \mapsto \mathbb{E}\|\mathcal{V}_{s-}\|^2$. 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 Stelzer (2010)

$$
\begin{aligned}
\mathbb{E}I_4 &= \mathbb{E}\Bigg(\int_0^\Delta \mathcal{V}_{s-} \mathrm{d}[\operatorname{vec}([L, L^*]), \operatorname{vec}([L, L^*])^*] \mathcal{V}_{s-} \Bigg) \\
&= \mathbb{E}\Bigg(\int_0^\Delta \mathcal{V}_{s-} \mathrm{d}[\operatorname{vec}([L, L^*]^\mathfrak{d}), (\operatorname{vec}([L, L^*]^\mathfrak{d}))^*]^\mathfrak{d} \mathcal{V}_{s-} \Bigg) \\
&= \int_0^\Delta \mathbb{E}\big(\mathcal{V}_{s-} \rho_L [I_{d^2} + K_d + \operatorname{vec}(I_d) \operatorname{vec}(I_d)^*] \mathcal{V}_{s-} \big) \mathrm{d}s \\
&= \rho_L \int_0^\Delta (\boldsymbol{Q} + K_d \boldsymbol{Q} + I_{d^2}) \mathbb{E}(\operatorname{vec}(V_s) \operatorname{vec}(V_s)^*) \mathrm{d}s \\
&= \Delta \rho_L (\boldsymbol{Q} + K_d \boldsymbol{Q} + I_{d^2}) \mathbb{E} \operatorname{vec}(V_0) \operatorname{vec}(V_0)^*.
\end{aligned}
\tag{A.25}
$$

To compute $\mathbb{E}I_1$ we will need the following matrix identity, which is based on Fact 7.4.30 (xiv) in Bernstein (2009). Let $A \in M_{d,1}(\mathbb{R})$ and $B, B^2 \in M_{d,d}(\mathbb{R})$ be symmetric matrices. Then,

$$
\begin{aligned}
&(A \otimes B + B \otimes A)(A \otimes B + B \otimes A)^* = (A \otimes B + K_d(A \otimes B))(A \otimes B + K_d(A \otimes B))^* \\
&= (I + K_d)(A \otimes B)(A^* \otimes B)(I + K_d) = (I + K_d) \boldsymbol{Q} \operatorname{vec}(AA^*) \operatorname{vec}(B^2)(I + K_d).
\end{aligned}
\tag{A.26}
$$

Write $b_s := \mathbb{E} \operatorname{vec}(G_s G_s^*) \operatorname{vec}(V_s)^*$, which is finite by Lemma 3.2 with $p = 2$. Using the compensation formula, (A.26) and the definition of f_s gives

$$\begin{aligned}
&\mathbb{E}\bigg(\int_0^\Delta f_{s-}\mathrm{d}[L,L^*]_s f_{s-}^*\bigg)\\
&=(\sigma_L+\sigma_W)\int_0^\Delta \mathbb{E}(f_s f_s^*)\mathrm{d}s\\
&=(\sigma_L+\sigma_W)\int_0^\Delta \mathbb{E}(G_{s-}\otimes V_{s-}^{1/2}+V_{s-}^{1/2}\otimes G_{s-})(G_{s-}^*\otimes V_{s-}^{1/2}+V_{s-}^{1/2}\otimes G_{s-}^*)\mathrm{d}s\\
&=(\sigma_L+\sigma_W)\int_0^\Delta (I+K_d)\boldsymbol{Q}b_s(I+K_d)\mathrm{d}s\\
&=(\sigma_L+\sigma_W)(I+K_d)\boldsymbol{Q}\bigg(\int_0^\Delta b_s\mathrm{d}s\bigg)(I+K_d).
\end{aligned}\tag{A.27}$$

Finally, it follows from (A.23) that

$$\int_0^\Delta b_s^*\mathrm{d}s=\int_0^\Delta \mathbb{E}\,\mathrm{vec}(V_s)a_{s-}^*\mathrm{d}s=(\sigma_L+\sigma_W)^{-1}D. \tag{A.28}$$

The result now is a direct consequence of (A.17), (A.23), (A.24), (A.25), (A.27) and (A.28).

Remark A.2. *An inspection of the proofs of Lemmas 4.2 and A.1 shows that the moment assumptions* (*a*.6) *and* (*c*.2) *are only needed to compute expectations of stochastic integrals with the integrator L. If L has paths of finite variation, these expectations can be computed by using the compensation formulas given in Facts A.1 and A.2 without* (*a*.6) *and* (*c*.2)*.*

References

David Applebaum. *Lévy Processes and Stochastic Calculus.* Cambridge University Press, Cambridge, 2009.

Selçuk Bayracı and Gazanfer Ünal. Stochastic interest rate volatility modeling with a continuous-time GARCH(1,1) model. *Journal of Computational and Applied Mathematics*, 259:464–473, 2014.

Anita Behme. Moments of MGOU processes and positive semidefinite matrix processes. *Journal of Multivariate Analysis*, 111:183–197, 2012.

Dennis S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas.* Princeton University Press, Princeton, 2009.

Enrico Bibbona and Ilia Negri. Higher moments and prediction-based estimation for the COGARCH(1,1) model. *Scandinavian Journal of Statistics*, 42(4):891–910, 2015.

Patrick Billingsley. *Probability and Measure.* Wiley, New York, 2008.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31 (3):307–327, 1986.

Richard Bradley. *Introduction to Strong Mixing Conditions*, volume 1. Kendrick Press, Heber City, 2007.

Serge Cohen and Jan Rosinski. Gaussian approximation of multivariate lévy processes with applications to simulation of tempered stable processes. *Bernoulli*, 13(1):195–210, 2007.

Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

Thiago Do Rêgo Sousa. *Simulation-based estimation of time series and stochastic volatility processes.* PhD thesis, Department of Mathematics, Technical University of Munich, 2019. URL `https://mediatum.ub.tum.de/?id=1518760`.

Thiago do Rêgo Sousa, Stephan Haug, and Claudia Klüppelberg. Indirect inference for Lévy-driven continuous-time GARCH models. *Scandinavian Journal of Statistics*, (46):765–801, 2019.

Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

Robert F. Engle and Kenneth F. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150, 1995.

Vicky Fasen. Asymptotic results for sample autocovariance functions and extremes of integrated generalized ornstein–uhlenbeck processes. *Bernoulli*, 16(1):51–79, 2010.

Vicky Fasen, Claudia Klüppelberg, and Alexander Lindner. Extremal behavior of stochastic volatility models. In *Stochastic finance*, pages 107–155. Springer, 2006.

Christian Francq and Jean-Michel Zakoïan. *GARCH Models: Structure, Statistical Inference and Financial Applications.* Wiley, Hoboken, NJ, second edition, 2019.

Dominique M Guillaume, Michel M Dacorogna, Rakhal R Davé, Ulrich A Müller, Richard B Olsen, and Olivier 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.

Stephan Haug, Claudia Klüppelberg, Alexander Lindner, and Matthias Zapp. Method of moment estimation in the COGARCH(1,1) model. *Econometrics Journal*, 10(2):320–341, 2007.

Roger A Horn and Charles R Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, 1991.

Ildar A. Ibragimov and Yuri V. Linnik. *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen, 1971.

Achim Klenke. *Probability Theory: a Comprehensive Course.* Springer, Berlin, 2013.

Claudia Klüppelberg, Ross Maller, and Alexander 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.

Ulrich Krengel. *Ergodic Theorems.* Walter de Gruyter, Berlin, 1985.

Alois Kufner, Oldrich John, and Svatopluk Fučik. *Function spaces.* Noordhoff, Leyden, 1977.

Alexander Lindner and Ross Maller. Lévy integrals and the stationarity of generalised Ornstein-Uhlenbeck processes. *Stochastic Processes and their Applications*, 115(10):1701–1722, 2005.

Jan R Magnus and Heinz Neudecker. The commutation matrix: some properties and applications. *The Annals of Statistics*, 7:381–394, 1979.

Ross A Maller, Gernot Müller, and Alex Szimayer. GARCH modelling in continuous time for irregularly spaced time series data. *Bernoulli*, 14(2):519–542, 2008.

László Mátyás. *Generalized Method of Moments Estimation.* Cambridge University Press, Cambridge, 1999.

Gernot Müller. MCMC estimation of the COGARCH(1,1) model. *Journal of Financial Econometrics*, 8 (4):481–510, 2010.

Christian Pigorsch and Robert Stelzer. On the definition, stationary distribution and second order structure of positive semi-definite Ornstein-Uhlenbeck type processes. *Bernoulli*, 15:754–773, 2009a.

Christian Pigorsch and Robert Stelzer. A multivariate Ornstein-Uhlenbeck type stochastic volatility model. *Technical Report: www.uni-ulm.de/en/mawi/finmath/people/ stelzer/publications.html*, 2009b.

Philip Protter. *Stochastic Integration and Differential Equations.* Springer, New York, 2 edition, 2005.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2017.

Ken-iti Sato. *Lévy processes and Infinitely Divisible Distributions.* Cambridge university press, Cambridge, 1999.

Eckhard Schlemm and Robert Stelzer. Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Levy-driven CARMA processes. *Electronic Journal of Statistics*, 6: 2185–2234, 2012.

Robert Stelzer. Multivariate COGARCH(1,1) processes. *Bernoulli*, 16(1):80–115, 2010.

Robert Stelzer and Johanna Vestweber. Geometric ergodicity of the multivariate COGARCH(1,1) process. *arXiv:1701.07859*, 2019.

Robert Stelzer, Thomas Tosstorff, and Marc Wittlinger. Moment based estimation of supou processes and a related stochastic volatility model. *Statistics & Risk Modeling*, 32(1):1–24, 2015.

Ralph M. Wilcox. Exponential operators and parameter differentiation in quantum physics. *Journal of Mathematical Physics*, 8(4):962–982, 1967.