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 Bayract and Unal [2014]; Bibbona and Negri [2015]; Haug et al. [2007]; Klüppelberg et al. [2011]; Müller 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

$$G_t = \int_0^t V_s^{1/2} dL_s, \quad t \geq 0,$$  \hspace{1cm} (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

$$G_i = \int_{(i-1)\Delta}^{i\Delta} V_s^{1/2} dL_s, \quad i = 1, \ldots, n.$$  \hspace{1cm} (1.2)

Therefore, an important question is how to estimate the true parameter $\theta_0$ based on observations $(G_i)_{i=1}^n$. In the univariate case, several methods have been proposed to estimate the parameters of the COGARCH process ([Bayracturk and Unal (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 $(G_i)_{i=1}^\infty$, 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 parameters. In Section 6 we study the finite sample behavior of the estimators in a simulation study. Finally, Section 7 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 $z \in \mathbb{C}, \mathbb{R}(z)$ 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 $S_d$, the (closed) positive semidefinite cone by $S^+_d$ and the (open) positive definite cone by $S^+_d$. 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, $A^\ast$ denotes the transpose of a matrix $A \in M_{m,d}(\mathbb{R})$ and $A_{(i,j)}$ denotes the entry in the $i$th line and $j$th 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 $M_d(\mathbb{R})$ 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 c\(a\)d\(a\)g\(g\) and adapted processes and $(L_t)_{t \in \mathbb{R}^+}$ in $M_{d,r}(\mathbb{R})$ be a semimartingale. We then denote by $\int_0^t A_s dB_s$ the matrix $C_t \in M_{m,u}(\mathbb{R})$ which has $ij$-th entry $\sum_{k=1}^d \sum_{s=1}^r A_{ik,s} - B_{ij,s} L_{kl,s}$. 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 $\langle X_t, Y_t \rangle_{t \in \mathbb{R}^+}$ is defined as the finite variation process in $M_{m,d}(\mathbb{R})$ with $ij$-th entry $\langle X_i, Y_j \rangle_t$ for $t \in \mathbb{R}^+$, $i = 1, \ldots, m$ and $j = 1, \ldots, d$. 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 $Q : M_d(\mathbb{R}) \mapsto M_d(\mathbb{R})$ be the linear map defined by

$$
(QX)(k-1)d+l, (p-1)d+q = X(k-1)d+p, (l-1)d+q \text{ for all } k, l, p, q = 1, \ldots, d,
$$

which has the property that $Q(\text{vec}(X) \text{ vec}(Z)^T) = X \otimes Z$ for all $X, Z \in S_d$ (Pigorsch and Stelzer, 2009b, Theorem 4.3)). Let $K_d$ be the commutation matrix characterized by $K_d \text{ vec}(A) = \text{ vec}(A^\ast)$ for all $A \in M_d(\mathbb{R})$ (see Magnus and Neudecker (1979) for more details). Define $Q \in M_{d^2}(\mathbb{R})$ as the matrix associated with the linear map $\text{vec} \circ Q \circ \text{vec}^{-1}$ on $\mathbb{R}^{d^2}$, and $K_d \in M_{d^2}(\mathbb{R})$ as the matrix associated with the linear map $\text{vec}(K_d \text{ vec}^{-1}(x))$ for $x \in \mathbb{R}^{d^2}$.
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 $\mathbb{E} e^{it^T L_t} = \exp \{ t \psi_L(u) \}$ for $t \in \mathbb{R}^+$ with

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

where $\gamma_L \in \mathbb{R}^d$, $\Gamma_L \in \mathbb{S}_d^+$ and the Lévy measure $\nu_L$ is a non-zero measure on $\mathbb{R}^d$ satisfying $\nu_L(\{0\}) = 0$ and $\int_{\mathbb{R}^d} (\|x\|^2 \wedge 1) \nu_L(dx) < \infty$. We assume w.l.o.g. $L$ to have càdlàg paths. The discontinuous part of the quadratic variation of $L$ is denoted by $([L,L]_t^B)_{t \in \mathbb{R}^+}$ and it is also a Lévy process. It has finite variation, zero drift and Lévy measure $\nu_{[L,L]^B}(B) = \int_{\mathbb{R}^d} I_B (xx^*) \nu_L(dx)$ 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

$$\begin{align*}
dG_t &= V_t^{1/2} dL_t \quad (3.1) \\
V_t &= C + Y_t \quad (3.2) \\
dY_t &= (BY_{t-} + Y_{t-}B^*) dt + AV_{t-}^{1/2} d[L,L]_t^B + V_{t-}^{1/2} A^* \quad (3.3)
\end{align*}$$

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 $(G_i G_i^*)_{i \in \mathbb{N}}$ 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 7.

We group the needed assumptions as follows.

**Assumptions a** (Lévy process).

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

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

(a.3) $\int_{\mathbb{R}^d} x_i x_j x_k \nu_L(dx) = 0$, 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

$$\mathbb{E}[\text{vec}([L, L^*])^\upsilon, \text{vec}([L, L^*])^\upsilon_i] = \rho_L(I_d^2 + K_d + \text{vec}(I_d) \text{vec}(I_d)^\upsilon).$$

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

**Assumptions b** (Parameters).

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

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

$$B \ := \ B \otimes I + I \otimes B + \sigma_L(A \otimes A)$$

$$C \ := \ B \otimes I_{d^2} + I_{d^2} \otimes B + AR,$$

where $A = (A \otimes A) \otimes (A \otimes A)$, $R = \rho_L(Q + K_d Q + I_{d^2})$, and $K_d$ and $Q$ 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 $\mathbb{E}\|Y_0\|^4 < \infty$. 

5
Sufficient conditions for Assumption c are given in \cite[Theorem 4.5]{Stelzer2010}. 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 $(G_t)_{t \in \mathbb{N}}$. First, for a second-order stationary $\mathbb{R}^d$-valued process, its autocovariance function $\text{acov}_X : \mathbb{R} \mapsto M_d(\mathbb{R})$ is denoted by $\text{acov}_X(h) = \text{cov}(X_{h}, X_0) = \mathbb{E}(X_h X_0^*) - \mathbb{E}(X_0)\mathbb{E}(X_0)^* \text{ for } h \geq 0$ and by $\text{acov}_X(h) = (\text{acov}_X(-h))^*$ for $h < 0$. For matrix-valued processes $(Z_t)_{t \in \mathbb{R}}$, we set $\text{acov}_Z = \text{acov}_{\text{vec}(Z)}$.

**Proposition 4.1** \cite[Theorems 4.8, 4.11, Corollary 4.19 and Proposition 5.2]{Stelzer2010}. If Assumptions (a.1) (a.5) (b.2) and (c.1) hold, then

\[
\mathbb{E}(\text{vec}(Y_0)) = -\sigma_L B^{-1}(A \otimes A) \text{vec}(C) \quad (4.2)
\]

\[
\text{var}(\text{vec}(Y_0)) = \text{var}(\text{vec}(V_0)) = -C^{-1}[(\sigma_L^2 C(B^{-1} \otimes B^{-1})A + \mathcal{A}R)(\text{vec}(C) \otimes \text{vec}(C)) + (\sigma_L(A \otimes A) \otimes I_d + \mathcal{A}R) \text{vec}(C) \otimes \mathbb{E}(\text{vec}(Y_0)) + (\sigma_L I_d \otimes (A \otimes A) + \mathcal{A}R) \mathbb{E}(\text{vec}(Y_0)) \otimes \text{vec}(C)]
\]

\[
\text{acov}_Y(h) = \text{acov}_Y(h) = e^{Bh} \text{var}(\text{vec}(Y_0))
\]

\[
\mathbb{E}(G_1) = 0
\]

\[
\text{var}(G_1) = (\sigma_L + \sigma_W)\Delta \mathbb{E}(C + Y_0)
\]

\[
\text{acov}_G(h) = 0 \text{ for all } h \in \mathbb{Z} \setminus \{0\}.
\]

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

**Lemma 4.2.** If Assumptions a, b and c hold, then

\[
\text{acov}_{GG^*}(h) = e^{B\Delta h} B^{-1}(I_{d^2} - e^{-B\Delta})(\sigma_L + \sigma_W) \text{var}(\text{vec}(V_0)) \times (e^{B\Delta} - I_d^2)[(\sigma_W + \sigma_L)(B^*)^{-1} - 2((A \otimes A)^*)^{-1}], \quad h \in \mathbb{N},
\]

\[
\mathbb{E} \text{ vec}(G_1 G_1^*) \text{ vec}(G_1 G_1^*)^* = \Delta \rho_L ((Q + K_d Q + I_d^2)(\mathbb{E} \text{ vec}(V_0) \text{ vec}(V_0)^*)) \times (I_{d^2} + K_d) Q(D^*)(I_{d^2} + K_d) + D + D^*,
\]

with,

\[
D := (\sigma_L + \sigma_W)\left(\frac{1}{2}(\sigma_L + \sigma_W)\Delta^2 \mathbb{E} \text{ vec}(V_0) \text{ vec}(V_0)^* + \text{var}(\text{vec}(V_0))\tilde{B}\right)
\]

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

**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 7.2). In the following, we will define the moment based estimator for MUCOGRARCH processes driven by general Lévy processes (without path restrictions). Only in Section 5.3 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 $C_\theta \in S_d^{++}$ 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 $(G_i)_{i=1}^n$ as defined in (1.2) with true parameter $\theta_0 \in \Theta$. We assume that the true $\sigma_L, \sigma_W$ and $\rho_L$ 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 Zakoian 2011, eq. (11.6)].

### 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 $(G_i)_{i=1}^n$ and a fixed $r < n$, the sample moments are defined as

$$\hat{k}_{n,r} = \frac{1}{n} \sum_{i=1}^{n-r} D_i = \frac{1}{n} \sum_{i=1}^{n-r} \begin{pmatrix} \text{vec}(G_i G_i^*) \\ \text{vec}(\text{vec}(G_i G_i^*) \text{vec}(G_i G_i^*)) \\ \vdots \\ \text{vec}(\text{vec}(G_i G_i^*) \text{vec}(G_{i+r} G_{i+r}^*)) \end{pmatrix}. \quad (5.1)$$

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

$$k_{\theta,r} = \begin{pmatrix} E_\theta \text{vec}(G_1 G_1^*) \\ E_\theta \text{vec}(\text{vec}(G_1 G_1^*) \text{vec}(G_1 G_1^*)) \\ \vdots \\ E_\theta \text{vec}(\text{vec}(G_1 G_1^*) \text{vec}(G_{1+r} G_{1+r}^*)) \end{pmatrix}, \quad (5.2)$$

where the expectations are explicitly given by (4.3), (4.4) and (4.5) by replacing $A, B$ and $C$ by $A_\theta, B_\theta$ and $C_\theta$, respectively. Then, the GMM estimator of $\theta_0$ is given by

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

where $\Omega$ is a positive definite weight matrix.

### 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 $(G_i)_{i \in \mathbb{N}}$. 

7
**Assumptions** d (Parameter space and log-price process).

(d.1) The parameter space $\Theta$ is a compact subset of $\mathbb{R}^q$.

(d.2) The true parameter $\theta_0$ lies in the interior of $\Theta$.

(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 $(G_i)_{i \in \mathbb{N}}$ is strictly stationary and exponentially $\alpha$-mixing.

**Assumption** e (Moments).

(e.1) There exists a positive constant $\delta > 0$ such that $E\|G_1\|^{8+\delta} < \infty$.

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 $k_{n,r} \overset{a.s.}{\rightarrow} k_{\theta_0}$ as $n \rightarrow \infty$.

**Proof.** It follows from (d.5) that the log-price process $(G_i)_{i \in \mathbb{N}}$ is ergodic and since both $E\|\text{vec}(G_1 G_1^\ast)\|$ and $E\|\text{vec}(G_1 G_1^\ast)\| \text{vec}(G_{1+h} G_{1+h}^\ast)\|$ 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.

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

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

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

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

**Lemma 5.3.** If Assumptions a, b, c, (d.1), and (d.4) hold, then the map $\Theta \mapsto k_{\theta,r}$ in (5.3) 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 $A_\theta, A_\theta^{-1}, B_\theta, B_\theta^{-1}, e^{-\alpha B_\theta}, \alpha > 0, C_\theta, C_\theta$ and $C_\theta^{-1}$. From (d.4) we obtain the continuous differentiability
For the asymptotic normality of (5.1) we use the Cramér-Wold device and show that

\[
\frac{\partial}{\partial \theta_i} e^{-\alpha B_{\theta}} = - \int_0^\alpha e^{-(\alpha-u)B_{\theta}} \left( \frac{\partial}{\partial \theta_i} B_{\theta} \right) e^{-u B_{\theta}} du. \tag{5.4}
\]

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

\[
\sup_{\theta \in \Theta} \|B_{\theta}\| \leq 2 \left( \sup_{\theta \in \Theta} \|B_{\theta}\| \right) \|I_d\| + \sigma_L \left( \sup_{\theta \in \Theta} \|A_{\theta}\|^2 \right) < \infty.
\]

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

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

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 $\Theta$. \hfill \Box

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

\[
\Sigma_{\theta_0} = \mathbb{E}(F_1 F_1^*) + 2 \sum_{i=1}^{\infty} \mathbb{E}(F_i F_{i+1})^* \tag{5.6}
\]

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

\[
\sqrt{n}(k_{n,r} - k_{\theta_0,r}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\theta_0}), \quad n \to \infty.
\]

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

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

for all vectors $\lambda \in \mathbb{R}^{d^2+(r+1)d}$. Denote by $\alpha G$ the mixing coefficients of $(G_i)_{i \in \mathbb{N}}$. Since each $F_i$ is a measurable function of $G_i, \ldots, G_{i+r}$, it follows from (d.5) and Remark 1.8 of Bradley (2007) that $(\lambda^* F_i)_{i \in \mathbb{N}}$ is $\alpha$-mixing with mixing coefficients satisfying $\alpha_F(n) \leq \alpha G(n-(r+1))$ for all $n \geq r+2$. Therefore, $\sum_{n=0}^{\infty} (\alpha_F(n))^{\frac{1}{2+\epsilon}} < \infty$ 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 $\alpha$-mixing sequences applies, see e.g. Ibragimov and Linnik (1971, Theorem 18.5.3), so that

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

where

\[
\zeta = \mathbb{E} \lambda^* F_1 F_1^* \lambda + 2 \sum_{i=1}^{\infty} \mathbb{E} \lambda^* F_i F_{i+1}^* \lambda.
\]

After rearranging this equation we find (5.6). \hfill \Box
Theorem 5.5. Assume that Assumptions $a, b, c, d$ and (e.1) hold and that the matrix $\Sigma$ in (5.6) is positive definite. Then the GMM estimator defined in (5.3) is asymptotically normal with covariance matrix

$$(J_{\theta_0})^{-1}I_{\theta_0}(J_{\theta_0})^{-1},$$

where $J_{\theta_0} = (\nabla_{\theta} k_{\theta_0,r})^\top \Omega(\nabla_{\theta} k_{\theta_0,r})$ and $I_{\theta_0} = (\nabla_{\theta} k_{\theta_0,r})^\top \Omega \Sigma_{\theta_0} \Omega(\nabla_{\theta} k_{\theta_0,r})$.

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 $\tilde{\theta}_n \to \theta_0$ 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 $\frac{\partial}{\partial \theta}(k_{\theta_n,r} - k_{\theta_n}) \overset{p}{\to} (k_{\theta_0} - k_{\theta_0})$ 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. 

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 MUCOARCH volatility process which is $\mu$-irreducible with the support of $\mu$ having non-empty interior and aperiodic. Assume that one of the following conditions is satisfied:

(i) setting $p = 1$ there exists $\Xi \in S_+^{d \times +}$ such that

$$\Xi B + B^\top \Xi + \sigma_L A^\top \Xi A \in -S_+^{d \times +},\quad (5.8)$$

(ii) there exist $p \in [1, \infty)$ and $\Xi \in S_+^{d \times +}$ such that

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

where

$$K_{\Xi,B} = \max_{X \in S_+^d, \text{tr}(X) = 1} \frac{\text{tr}\left((\Xi B + B^\top \Xi) X\right)}{\text{tr}(\Xi X)}$$

and

$$K_{\Xi,A} = \max_{X \in S_+^d, \text{tr}(X) = 1} \frac{\text{tr}\left(A^\top \Xi AX\right)}{\text{tr}(\Xi X)}.$$

(iii) there exist $p \in [1, \infty)$ and $\Xi \in S_+^{d \times +}$ such that

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

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 \( \beta \)-mixing. This implies \( \alpha \)-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 \( \beta \)-mixing, then the log-price process \( (G_i)_{i \in \mathbb{N}} \) is stationary, exponentially \( \alpha \)-mixing, and as a consequence also ergodic.

**Proof.** Since \( Y \) is an exponentially \( \beta \)-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 \( \Theta \). 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 \( \theta \) 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 \( A_\theta, B_\theta \), 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 \( \theta \) from the notation, except when explicitly needed. We start with the identifiability of the matrix \( C \).

**Lemma 5.10.** Assume that Assumptions [(a.1)](assumption_a1), [(a.5)](assumption_a5), [(b.2)](assumption_b2), and [(c.1)](assumption_c1) hold and that \( \sigma(B) \subset \{ z \in \mathbb{C} : \Re(z) < 0 \} \). If the matrices \( A \) and \( B \) are known, then \( \mathbb{E}(G_1 G_1^\top) \) uniquely determines \( C \).

**Proof.** Since \( \sigma(B \otimes I + I \otimes B) = \sigma(B) + \sigma(I) \subset \{ z \in \mathbb{C} : \Re(z) < 0 \} \), 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

\[
\text{vec}(C) = (\sigma_L + \sigma_W)^{-1} \Delta^{-1} (B \otimes I + I \otimes B)^{-1} B \text{vec}(\mathbb{E}(G_1 G_1^\top)).
\]

\[ \Box \]
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

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

(5.11)

with

$$
\|X\|_S = \|S^{-1}X(S \otimes S)\|_2, \quad X \in M_d(\mathbb{R}), S \in GL_d(\mathbb{C}),
$$

(5.12)

then the matrix

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

(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 $X = \frac{B}{(\sigma_L + \sigma_W)}, Y = -\frac{1}{2}(A \otimes A)$ and using the definition of $B$ in (4.1) we get

$$
X + Y = \frac{1}{(\sigma_L + \sigma_W)} \left( (B \otimes I + I \otimes B) + \frac{(\sigma_L - \sigma_W)}{2}(A \otimes A) \right).
$$

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 $B$ 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 $X \in S_d$ identifies $A$.

**Proof.** Assume first that $A_{(1,1)} > 0$. For each $i \in \{1, \ldots, d\}$, let $e_i$ be the $i$th 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

$$
(A_{(1,1)}^2, A_{(1,1)}A_{(2,1)}, \ldots, A_{(1,1)}A_{(d,1)}).
$$

(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

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

(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 \( a, b \) and \( c \) and the conditions of Lemma 5.12 hold, that the matrix in \([5.13]\) is invertible, that \( \sigma(\mathcal{B}) \subset \{ z \in \mathbb{C} : -\pi < 3(z)\Delta < \pi, \Re(z) < 0 \} \) and that \( \text{var}(\text{vec}(V_0)) \) is invertible. Define \( M = (e^{B\Delta})^{-1}\text{acov}_{\theta,GG^*}(1) \). Then, \( \text{acov}_{\theta,GG^*}(1) \) and \( \text{acov}_{\theta,GG^*}(2) \) uniquely identify \( \mathcal{B} \) and \( M \).

Proof. Since \( M \) is given in terms \( \mathcal{B} \) and \( \text{acov}_{\theta,GG^*}(1) \), 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 \( S_d \) 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 vec operator and cumbersome notation, we take an abstract point of view. The variance of a random element of \( S_d \) is a symmetric positive semi-definite linear operator from \( S_d \) to itself. Likewise, the auto-covariance of \( G_1^* \) and \( G_{1+k}^* \) is a linear operator from \( S_d \) to itself. The condition that \( \text{var}(\text{vec}(V_0)) \) 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

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

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

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

By the assumptions on the eigenvalues of \( \mathcal{B} \) there is a unique logarithm for \( e^{B\Delta} \) (see [Horn and Johnson 1991 Section 6.4]) or [Schlemm and Stelzer 2012 Lemma 3.11]), so \( B\Delta \) and thus \( \mathcal{B} \) is identified. Finally, note that the matrices in the vec representations are uniquely identified by the employed linear operators on \( S_d \) 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 \( A_\theta \) and \( B_\theta \) satisfy: for some \( k \neq l \in \{1, \ldots, d\} \), \( A_{(k,l),\theta} > 0 \), \( A_{(l,k),\theta} \neq A_{(l,k),\theta} \) and \( B_{(k,k),\theta} = B_0. \) Then \( k_1, \theta \) uniquely identifies \( A_\theta, B_\theta \) and \( C_\theta \).

Proof. Recall that we omit \( \theta \) 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

\[
\begin{pmatrix}
2B_{(1,1)} + A^2_{(1,1)} & B_{(1,2)} + A_{(1,1)}A_{(1,2)} & B_{(1,2)} + A_{(1,1)}A_{(1,2)} & A^2_{(1,2)} \\
B_{(2,1)} + A_{(1,1)}A_{(2,1)} & B_{(1,1)} + B_{(2,2)} + A_{(1,1)}A_{(2,2)} & A_{(2,2)} & A_{(2,2)} \\
B_{(2,1)} + A_{(1,1)}A_{(2,1)} & A_{(1,1)}A_{(2,1)} & B_{(1,1)} + B_{(2,2)} + A_{(1,1)}A_{(2,2)} & B_{(1,2)} + A_{(1,2)}A_{(2,2)} \\
A^2_{(2,1)} & B_{(2,1)} + A_{(2,1)}A_{(2,2)} & B_{(2,1)} + A_{(2,1)}A_{(2,2)} & 2B_{(2,2)} + A^2_{(2,2)}
\end{pmatrix}
\]

Using the entry at position (1, 4) and the fact that \( A_{(1,2)} > 0 \), we 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 \( A_{(1,2)} \neq A_{(1,2)} \) and \( B_{(1,2)} = B_{(2,1)} \) to write \( A_{(1,1)} = (B_{(1,2)} - B_{(2,1)})/(A_{(1,2)} - A_{(2,1)}) \). Similarly, we use the entries at positions (3, 4), (4, 3) to get
Finally, since the matrix $A_{(2,2)} = (B_{(3,4)} - B_{(4,3)}/(A_{(1,2)} - A_{(2,1)})$. 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 $B$ from (4.1) in the following block form:

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

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

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

Since $A_{(1,2)} > 0$ we can identify it from (5.18), because $B^{(1,2)}(1,2) = A^{2}_{(1,2)}$. Then we use the off-diagonal entries of the matrix $B^{(1,2)}$ 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

$$\begin{cases} B^{(k,k)}_{(1,2)} = B_{(1,2)} + A_{(k,k)}A_{(1,2)}, & k = 1, \ldots, d. \\ B^{(k,k)}_{(2,1)} = B_{(2,1)} + A_{(k,k)}A_{(2,1)} \end{cases}$$ (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

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

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

In Lemma 5.14 we identify the matrices $A$ and $B$ only from $B$ 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 $E\text{vec}(\text{vec}(G_1G_1^*)\text{vec}(G_1G_1^*)^*)$ 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 $(Q + KdQ + I_d)$ at the right hand side of (4.5) is not one-to-one in the space $E\text{vec}(V_0)\text{vec}(V_0)^*$. In the end, in order to use the moment conditions $E\text{vec}(G_1G_1^*)\text{vec}(G_1G_1^*)^*$, we need to assume that the matrices $B \otimes I + I \otimes B$ and $A \otimes A$ commute (see [Do Rêgo Sousa, 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 $\theta$ 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 $\Theta$.

First, we define

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

(5.20)

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

(5.21)

Consider now the following group of assumptions:

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

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

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

(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 $A_\theta$ and $B_\theta$ satisfy: for some $k \neq l \in \{ 1, \ldots, d \}$, $A_{(k,l),\theta} > 0$, $A_{(l,k),\theta} \neq A_{(l,k),\theta}$ and $B_{(k,l),\theta} = B_{(l,k),\theta}$.

(f.5) The matrix $\vartheta_0(\text{vech}(V_0))$ is invertible.

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

(f.7) There exists $\Xi_\theta \in S_d^{++}$ such that, condition (5.8) holds with $A, B$ replaced by $A_\theta, B_\theta$.

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

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

(5.22)

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

**Assumptions g** (MUCOGARCH process at $\theta_0$).

(g.1) The MUCOGARCH volatility process $Y$ is stationary, $\mu$-irreducible with the support of $\mu$ 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
Corollary 5.16 (Consistency of the GMM estimator - a). Let \( \mathbb{S}_d \in \mathbb{S}_d \) denote the usual trace functional) defines a scalar product on \( \mathbb{R}^d \) if it has finite second moments and

\[
\lim_{t \to \infty} \mathbb{E} \langle X_t \rangle = \mu, \quad \lim_{t \to \infty} \text{var}(\langle X_t \rangle) = \Sigma
\]

\[
\lim_{t \to \infty} \sup_{h \to 0} \left\{ \left\| \text{cov}(\langle X_{t+h} \rangle, \langle X_t \rangle) - f(h) \right\| \right\} = 0.
\]

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.

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 \( B_0, B_0, C_0 \) 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 \( \text{tr}(Y_t Y) \) (tr denoting the usual trace functional) defines a scalar product on \( \mathbb{S}_d \) via \( \text{tr}(Y_t Y) = \text{vec}(Y_t^*) \text{vec}(Y_t) \) it follows that

\[
\mathbb{E}\|Y_t\|^2 = \text{tr}(Y_t^* Y_t) = \text{vec}(Y_t^*) \text{vec}(Y_t) = \sum_{i,j}^d EY_{t,ij}^2 = \sum_{i,j}^d \text{var}(Y_{t,ij}) + \sum_{i,j}^d (EY_{t,ij})^2
\]

\[
\text{tr}(\text{var}(Y_t)) + \|\mathbb{E}(Y_t)\|^2, \quad t > 0.
\]

Since both maps \( t \mapsto \mathbb{E}\|Y_t\|^2 \) and \( t \mapsto \text{var}(Y_t) \) are continuous (Stelzer 2010 eqs. (4.7) and (4.16)), it follows from (5.23) that \( \lim_{t \to \infty} \mathbb{E}\|Y_t\|^2 < \infty \). 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 7.2. 

□
Recall that for the asymptotic normality result, we need to ensure that the stationarity distribution of the MUCO\-GARCH 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.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 $E\|Y_0\|^p < \infty$ 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).

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 $L$ is chosen as $N(0,1/4I_2)$ and the jump rate is 4, so that $\text{var}(L_1) = 2I_2$ and

$$E[\text{vec}([L, L^*])^2] = 1/4(I_4 + K_2 + \text{vec}(I_2)\text{vec}(I_2)^*)$$

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 $A_\theta$, $B_\theta$ and $C_\theta$ are defined as:

$$A_\theta = \begin{pmatrix} \theta^{(1)} & \theta^{(2)} \\ \theta^{(3)} & \theta^{(4)} \end{pmatrix}, \quad B_\theta = \begin{pmatrix} \theta^{(5)} & \theta^{(6)} \\ \theta^{(6)} & \theta^{(7)} \end{pmatrix} \quad \text{and} \quad C_\theta = \begin{pmatrix} \theta^{(8)} & \theta^{(9)} \\ \theta^{(10)} & \theta^{(11)} \end{pmatrix},$$

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 MUCO\-GARCH model in dimension $d = 2$. We fix $L_t = L_t^0 + \sqrt{\sigma_W} W_t$ for $t \in \mathbb{R}^+$ where $L^0$ is a bivariate compound Poisson process (CPP), $W$ is a standard bivariate Brownian motion, independent of $L^0$ and $\sigma_W \geq 0$ is fixed. We choose $L^0$ as a CPP, since it allows to simulate the MUCO\-GARCH 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 $L^0$ 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 $L^0$ 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)).
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 \( G = (G_t)_{t=1}^n \) 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 \( \sigma = 1 \),
\[
A_{\theta_0} = \begin{pmatrix}
0.85 & 0.10 \\
-0.10 & 0.75
\end{pmatrix}, \quad
B_{\theta_0} = \begin{pmatrix}
-2.43 & 0.05 \\
0.05 & -2.42
\end{pmatrix}
\quad \text{and} \quad
C_{\theta_0} = \begin{pmatrix}
1 & 0.5 \\
0.5 & 1.5
\end{pmatrix}.
\] (6.1)

**Example 6.2.** We fix \( \sigma = 0 \), \( A_{\theta_0} \) and \( C_{\theta_0} \) are as in Example 6.1 and \( B_{\theta_0} = \frac{1}{4} \begin{pmatrix}
-2.43 & 0.05 \\
0.05 & -2.42
\end{pmatrix} \).
\] (6.2)

For the chosen Lévy process here, Assumption (g.1) is satisfied. In Example 6.1, \( \theta_0 \) is chosen in such a way that the asymptotic normality of \( \hat{\theta}_n \) can be verified. Then, in Example 6.2 we rescale \( B_{\theta_0} \) 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 \( n \in \{1000, 5000, 10000, 20000, 50000, 100000\} \). 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 \( \theta_0 \). We only consider estimators based on the identity matrix for the weight matrix \( \Omega \) 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 \( A_{\theta_0} \) are such that Assumptions \( b \) and (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. We use Corollary 5.17 to ensure asymptotic normality. For our choice of \( \theta_0 \) we have that \( B_{\theta_0} \) 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 \( p = 4.001 \),
\[
\int_{\mathbb{R}^2} \left(1 + \alpha \left\| \text{vec}(yy^*) \right\|_{S_{\theta_0}}^p - 1\right) \nu_L(dy) + 2p \max\{\Re(\sigma(B_{\theta_0}))\} = -0.024 < 0.
\] (6.3)

Therefore, (g.2) is also valid and Corollary 5.17 gives asymptotic normality of the GMM estimator. 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 Figures 1 and 2, where we excluded those paths for which the algorithm did not converge successfully (around 10 percent of the paths of length \( n = 1000 \) and less than 3 percent for larger \( n \)). Figures 1 and 2 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 \( B_{\theta_0} \) is more difficult than the other parameters, specially for \( n \in \{1000, 5000, 10000\} \).

Figures 3 and 4 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. We also restrict ourselves to \( n \in \{5000, 20000, 100000\} \), since they already allow us to analyse the convergence to the normal distribution. Here we do not exclude those paths for which the algorithm did not converge (these are denoted by large red points in the normal QQ-plots in Figures 3 and 4). 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 \( B_{\theta_0} \) deviate from the ones of a normal distribution for values of \( n \in \{5000, 20000\} \), but they get closer to a normal distribution for \( n = 100000 \). The left tail of the plots for \( A_{(2,1),\hat{\theta}_n} \) in Figure 3 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 \( n = 5000 \), there are very large negative outliers for the estimates of the diagonal entries of \( B_{\theta_0} \), which affects the bias substantially.
Figure 1: Example 1: Estimated absolute bias of $\hat{\theta}_{n,r}$. The colors green, blue and red correspond to $r = 2, 5$ and 10, respectively.
Figure 2: Example 1: Estimated standard deviation (std) of $\hat{\theta}_{n,r}$. The colors green, blue and red correspond to $r = 2, 5$ and 10, respectively.
Figure 3: Example 1: Normal QQ-plots of $\hat{\theta}_{n,10}$ for $\theta_0$ as in (6.2). The red dots are values for which the algorithm did not converge.
Figure 4: Example 1: Normal QQ-plots of $\hat{\theta}_{n,10}$ for $\theta_0$ as in (6.2). The red dots are values for which the algorithm did not converge.
Figure 5: Example 2: Estimated mean absolute error of $\hat{\theta}_{n,r}$. The colors green, blue and red correspond to $r = 2, 5$ and $10$, respectively.
Figure 6: Example 2: Normal QQ-plots of $\hat{\theta}_{n,10}$ for $\theta_0$ as in (6.2). The red dots are values for which the algorithm did not converge.
Figure 7: Example 2: Normal QQ-plots of $\hat{\theta}_{n,10}$ for $\theta_0$ 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 $\sigma(B_{\theta_0}/4+$
Proof. First notice that \( \text{vec}(B_{\theta_0}^*/4 + \sigma_L A_{\theta_0}^* A_{\theta_0}) = \{-0.594, -0.619\} \in (-\infty, 0) + i\mathbb{R} \). 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 5-7. The estimation of the entries of \( B_{\theta_0} \) does not seem to be substantially more difficult than the entries of \( A_{\theta_0} \) and \( C_{\theta_0} \), as observed in the previous example. Also, the estimated mean absolute error 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 \( A_{(2,1)} \), and \( C_{(2,1)} \) also show some deviation from the normal distribution.

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.

7 Proofs

7.1 Auxiliary results

Several results related to the algebra of multivariate stochastic integrals will be used here, for which we refer to Lemma 2.1 in [Behme 2012]. Furthermore, we need the following.

Fact 7.1 ([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

\[
\mathbb{E} \int_0^\Delta X_{s-} dL_s = \int_0^\Delta \mathbb{E}(X_{s-}) \mathbb{E}(L_1) ds.
\]

Fact 7.2. Let \((A_t)_{t \in \mathbb{R}^+}\) in \(M_d(\mathbb{R})\), \((B_t)_{t \in \mathbb{R}^+}\) in \(M_{d^2}(\mathbb{R})\) be adapted càdlàg 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,

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

Proof. First notice that \(\text{vec}([L, L]_s)\) is an \(\mathbb{R}^{d^2}\)-valued Lévy process with finite variation. Then it follows from Fact 7.1 that

\[
\text{vec} \left( \mathbb{E} \int_0^t A_s d(\text{vec}([L, L]_s)) B_s \right) = \mathbb{E} \int_0^t (B_s \otimes A_s) d(\text{vec}([L, L]_s)) = \int_0^t \mathbb{E}(B_s \otimes A_s) \text{vec}([L, L]_1) ds = (\sigma_W + \sigma_L) \int_0^t \mathbb{E}(B_s \otimes A_s) \text{vec}(I_d) ds = (\sigma_W + \sigma_L) \text{vec} \left( \int_0^t \mathbb{E}(A_s I_d B_s) ds \right).
\]
so the result follows by an application of \( \text{vec}^{-1} \).

**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 thatLemma 7.1.

that \( p \) both

Since (\( \text{Proof.} \))

Integration by parts formula (Stelzer, 2010, p. 111) gives

A integration by parts formula, the formula \( d(\text{vec}(G_t)) \)

where the finiteness is due to the fact that the integrand is locally bounded, and thus, also

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

\[
d(\text{vec}(Y_s)) = \tilde{C} \text{vec}(Y_s)ds + (A \otimes A)(V_{s-}^{1/2} \otimes V_{s-}^{1/2})d \text{vec}([L, L]^0_s). \tag{7.5}
\]
Using (7.5), the bilinearity of the quadratic covariation process, \cite[Lemma 3.2, Facts 7.1, (a.3), (7.3) and the Itô isometry]{Behme2012} we obtain

\[
[\text{vec}(Y), (\text{vec}(A))^*]_\Delta = \left[ \int_0^\Delta \tilde{C} \text{vec}(Y_s) \, ds + \int_0^\Delta (A \otimes A)(V_s^{-1/2} \otimes V_s^{-1/2}) \text{vec}([L, L]^\otimes) \, ds, \int_0^\Delta dL_s^* (G_s^* \otimes V_s^{-1/2}) \right]_\Delta
\]

\[
= \int_0^\Delta (A \otimes A)(V_s^{-1/2} \otimes V_s^{-1/2}) \text{vec}([L, L]^\otimes), L^*_s (G_s^* \otimes V_s^{-1/2}) \right]_\Delta.
\]

(7.6)

Recall that for arbitrary matrices \(M \in M_{m,n}(\mathbb{R})\) and \(N \in M_{k,l}(\mathbb{R})\) it holds \(\|A \otimes B\|_2 = \|A\|_2 \|B\|_2\) \cite[Fact 9.9.61]{Bernstein2009}. This together with the Hölder inequality with \((3/4 + 1/4 = 1)\) and Lemma 3.2 with \(p = 4\) gives

\[
\int_0^\Delta \mathbb{E}\|V_s^{-1/2} \otimes V_s^{-1/2}\|_2^2 \|G_s^* \otimes V_s^{-1/2}\|_2^2 \, ds = \int_0^\Delta \mathbb{E}\|V_s^{-1/2}\|_2^3 \|G_s^* - 1\|_2^2 \, ds
\]

\[
= \int_0^\Delta \mathbb{E}\|V_s^{-1/2}\|_2^{3/4} (\mathbb{E}\|G_s^* - 1\|_2^{1/4})^2 \, ds < \infty.
\]

Thus, applying expectations at both sides of (7.6) gives

\[
\mathbb{E}[\text{vec}(Y), (\text{vec}(A))^*]_\Delta = 0.
\]

(7.7)

Let \(l := \mathbb{E}\text{vec}(Y_s) (\text{vec}(A_s))^*\) and notice that it follows from Lemma 3.2 and the Cauchy-Schwarz inequality that \(\mathbb{E}\|l\| < \infty\) and \(s \mapsto \mathbb{E}\|l_s\|\) is locally bounded. Use \(\text{(7.5), (7.7), the compensation formula, (Bernstein2009, Proposition 7.1.9)}, \text{vec}(V_s) \text{vec}(A_{-s}) = l_s\), the Itô isometry and \((a.2)\) to get

\[
l_\Delta = \mathbb{E} \int_0^\Delta \text{vec}(Y_s) (\text{vec}(A_{-s}))^* \, ds
\]

\[
= \mathbb{E} \int_0^\Delta \left[ \tilde{C} \text{vec}(Y_s) \, ds + (A \otimes A)(V_s^{-1/2} \otimes V_s^{-1/2}) \text{vec}([L, L]^\otimes) \right] (\text{vec}(A_{-s}))^* \, ds
\]

\[
= \tilde{C} \int_0^\Delta \mathbb{E} \text{vec}(Y_s) (\text{vec}(A_{-s}))^* \, ds
\]

\[
+ \sigma_L \int_0^\Delta \mathbb{E} [(A \otimes A)(V_s^{-1/2} \otimes V_s^{-1/2}) \text{vec}(I_d)] (\text{vec}(A_{-s}))^* \, ds
\]

\[
= (\tilde{C} + \sigma_L (A \otimes A)) \int_0^\Delta l_s \, ds.
\]

Solving the matrix-valued integral equation in (7.8) and using that \(A_0 = 0\) implies \(l_0 = 0\), gives \(l_s = 0\) for all \(s \geq 0\) \cite{Haug2007}. Thus, it follows from (7.4) - (7.8) that

\[
\text{cov}(\text{vec}(Y_\Delta), \text{vec}(A_\Delta)) = 0,
\]

(7.9)

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

\[
\text{vec}(C_\Delta) = \int_0^\Delta V_{-s} \text{vec}([L, L^*]_s) = \int_0^\Delta V_{-s} \text{vec}([L, L^*]_s^\otimes)
\]

\[
+ \sigma_W \int_0^\Delta (V_{s-}^{-1/2} \otimes V_{s-}^{-1/2}) \text{vec}(I_d) \, ds = \int_0^\Delta V_{-s} \text{vec}([L, L^*]_s^\otimes) + \sigma_W \int_0^\Delta \text{vec}(V_{s-}) \, ds.
\]

(7.10)
Using the compensation formula, Fact 7.1 and the stationarity of \((V_s)_{s \in \mathbb{R}_+}\) we get
\[
\mathbb{E} \int_0^\Delta V_s - \text{d} \text{vec}([L, L^*]_s) = (\sigma_W + \sigma_L) \int_0^\Delta \mathbb{E} V_s - \text{vec}(I_d) \text{d}s = \Delta (\sigma_W + \sigma_L) \mathbb{E} \text{vec}(V_0).
\] (7.11)
Additionally, it follows from Lemma 3.2 that \(\mathbb{E} \| \text{vec}(V_s) \text{vec}(Y_\Delta)^* \| < \infty\) for all \(s \geq 0\) and that \(s \mapsto \mathbb{E} \| \text{vec}(V_s) \text{vec}(Y_\Delta)^* \|\) is locally bounded. Then,
\[
\mathbb{E} \left( \int_0^\Delta \text{vec}(V_{s-}) \text{d} \text{vec}(Y_\Delta)^* \right) = \int_0^\Delta \mathbb{E} \text{vec}(V_{s-}) \text{vec}(Y_\Delta)^* \text{d}s
\]
\[
= \Delta \text{vec}(C) \mathbb{E} (\text{vec}(Y_\Delta)^*) + \int_0^\Delta \mathbb{E} \text{vec}(Y_s) \text{vec}(Y_\Delta)^* \text{d}s.
\]
Now it follows from the invertibility of \((A \otimes A)\) and from the second equation following (3.5) in Stelzer (2010) that
\[
\int_0^\Delta V_s - \text{d} \text{vec}([L, L^*]_s) = (A \otimes A)^{-1} \left( \text{vec}(Y_\Delta) - \text{vec}(Y_0) - \int_0^\Delta (B \otimes I + I \otimes B) \text{vec}(Y_{s-}) \text{d}s \right).
\] (7.12)

The representation in (7.12) gives
\[
\mathbb{E} \left[ \left( \int_0^\Delta V_s - \text{d} \text{vec}([L, L^*]_s) \right) \text{vec}(Y_\Delta)^* \right]
\]
\[
= \mathbb{E} \left[(A \otimes A)^{-1} \left( \text{vec}(Y_\Delta) - \text{vec}(Y_0) - \int_0^\Delta (B \otimes I + I \otimes B) \text{vec}(Y_{s-}) \text{d}s \right) \text{vec}(Y_\Delta)^* \right]
\]
\[
= (A \otimes A)^{-1} \left[ \mathbb{E} \text{vec}(Y_\Delta) \text{vec}(Y_\Delta)^* - \mathbb{E} \text{vec}(Y_0) \text{vec}(Y_\Delta)^* \right.
\]
\[
- (B \otimes I + I \otimes B) \int_0^\Delta \mathbb{E} \text{vec}(Y_{s-}) \text{vec}(Y_\Delta)^* \text{d}s \right].
\] (7.13)

Using the definition of \(C_\Delta\) in (7.2), together with (7.10), (7.11) and (7.13) gives
\[
\text{cov}(\text{vec}(C_\Delta), \text{vec}(Y_\Delta)) = (A \otimes A)^{-1} \left[ \mathbb{E} \text{vec}(Y_\Delta) \text{vec}(Y_\Delta)^* - \mathbb{E} \text{vec}(Y_0) \text{vec}(Y_\Delta)^* \right.
\]
\[
- (B \otimes I + I \otimes B) \left( \int_0^\Delta \mathbb{E} \text{vec}(Y_{s-}) \text{vec}(Y_\Delta)^* \text{d}s \right)
\]
\[
+ \Delta \sigma_W \text{vec}(C)(\mathbb{E} \text{vec}(Y_0)^*) + \sigma_W \int_0^\Delta \mathbb{E} \text{vec}(Y_{s-}) \text{vec}(Y_\Delta)^* \text{d}s
\]
\[
- \Delta \sigma_W + \sigma_L \mathbb{E} \text{vec}(V_0) \text{vec}(Y_\Delta)^* \right]
\]
\[
= [\sigma_W I_d - (A \otimes A)^{-1}(B \otimes I + I \otimes B)] \int_0^\Delta \mathbb{E} \text{vec}(Y_s) \text{vec}(Y_\Delta)^* \text{d}s
\]
\[
+ (A \otimes A)^{-1} \left[ \text{var}(\text{vec}(Y_0)) - \text{cov}(\text{vec}(Y_0), \text{vec}(Y_\Delta)) \right] - \Delta \sigma_L \text{vec}(C) \mathbb{E} \text{vec}(Y_0)^* \right.
\]
\[
- \Delta (\sigma_W + \sigma_L) \mathbb{E} \text{vec}(V_0) \text{vec}(Y_\Delta)^* \right],
\]
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
\[
\int_0^\Delta \mathbb{E} \text{vec}(Y_s) \text{vec}(Y_\Delta)^* \text{d}s = \int_0^\Delta e^{B(\Delta - s)} \text{var}(\text{vec}(Y_0)) \text{d}s + \Delta \mathbb{E} \text{vec}(Y_0) \text{vec}(Y_\Delta)^* \right.
\]
\[
= B^{-1}(e^{B\Delta} - I_d) \text{var}(\text{vec}(Y_0)) + \Delta \mathbb{E} \text{vec}(Y_0) \text{vec}(Y_\Delta)^* \right),
\] (7.14)
and second that

$$\text{var}(\text{vec}(Y_0)) - \text{cov}(\text{vec}(Y_0), \text{vec}(Y_\Delta)) = -(e^{B\Delta} - I_{d^2})\text{var}(\text{vec}(Y_0)).$$  \hspace{1cm} (7.15)$$

Substituting $B \otimes I + I \otimes B = B - \sigma_L(A \otimes A)$, using (7.14), (7.15), (4.1) and the formula for $E \text{vec}(Y_0)$ in (4.2) gives

$$\text{cov}(\text{vec}(C_\Delta), \text{vec}(Y_\Delta)) = -\Delta \text{E} \text{vec}(Y_0)\text{vec}(Y_\Delta)^* .$$

Finally, the result of the Lemma follows from (7.2), (7.9), (7.16) and the fact that

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

\[ \square \]

**7.2 Proof of Lemma 4.2**

(i) The proof of Lemma 4.2 (i) follows directly from Lemma 7.1 combined with (5.7) in [Stelzer 2010].

(ii) Denoting by $\| \cdot \|_F$ the Frobenius norm we have by Lemma 3.2(b) with $p = 2$

$$\text{E}\| \text{vec}(G_1 G_1^*) \text{vec}(G_1 G_1^*)^* \|_2 = \text{E}\| \text{vec}(G_1 G_1^*) \|_2^2 = \text{E}\| G_1 G_1^* \|_F^4 < \infty ,$$

$$= \text{tr}(G_1 G_1^* G_1 G_1^*) = \text{E}\| G_1 \|_2^4 < \infty .$$

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

$$a_{\Delta} a_{\Delta}^* = \int_0^\Delta a_{s-}d(a_{s}^*) + \int_0^\Delta da_s(a_{s-}^*) + [a, a^*]_{\Delta}$$

$$= \left( \int_0^\Delta da_s(a_{s-}^*) \right)^* + \int_0^\Delta da_s(a_{s-}^*) + [a, a^*]_{\Delta}$$

hence we only need to prove that the random variables

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

31
have finite expectations and compute them in closed form. From \(7.2\), Lemma 2.1(vi) in Behme (2012) and the symmetry of \((V_t)_{t \in \mathbb{R}_+}\) it follows that

\[
\text{da}_t = d(\text{vec}(G_t G_t^*)) = d(\text{vec} \left( \int_0^t V_{s-}^{1/2} dL_s G_{s-}^* + \int_0^t G_{s-} dL_s V_{s-}^{1/2} + \int_0^t V_{s-}^{1/2} d\|L, L^*\|_s V_{s-}^{1/2} \right))
\]

\[
= d \left( \int_0^t (G_{s-} \otimes V_{s-}^{1/2}) dL_s + \int_0^t (V_{s-}^{1/2} \otimes G_{s-}) dL_s + \int_0^t (V_{s-}^{1/2} \otimes V_{s-}^{1/2}) d\text{vec}([L, L^*]_s) \right)
\]

\[
= (G_{t-} \otimes V_{t-}^{1/2} + V_{t-}^{1/2} \otimes G_{t-}) dL_t + (V_{t-}^{1/2} \otimes V_{t-}^{1/2}) d\text{vec}([L, L^*]_t), \quad t \geq 0.
\] (7.18)

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

\[
\int_0^\Delta \mathbb{E}\|G_{s-} \otimes V_{s-}^{1/2}\|_2^2 \|a_{s-}\|_2^2 ds = \int_0^\Delta \mathbb{E}\|G_{s-}\|_2^2 \|V_{s-}^{1/2}\|_2^2 \|\text{vec}(G_{s-} G_{s-}^*)\|_2^2 ds \quad (7.19)
\]

\[
= \int_0^\Delta \mathbb{E}\|G_{s-}\|_2^2 \|V_{s-}^{1/2}\|_2^2 \|G_{s-} G_{s-}^*\|_2^2 ds \leq \int_0^\Delta \left( \mathbb{E}\|G_{s-}\|_2^4 (\mathbb{E}\|V_{s-}\|_2^4) \right)^{1/2} \mathbb{E}\|G_{s-}\|_2^4 ds,
\]

which is finite by Lemma 3.2 with \(p = 4\). Additionally, similar calculations and Lemma 3.2 with \(p = 2\) shows that \(\mathbb{E}(\|V_{s-}^{1/2} \otimes V_{s-}^{1/2}\|_2^4) \|a_{s-}\|_2 \leq (\mathbb{E}\|V_{s-}\|_2^4) \|G_{s-}\|_2^4 < \infty\) for all \(s > 0\) and the map \(s \mapsto \mathbb{E}(\|V_{s-}^{1/2} \otimes V_{s-}^{1/2}\|_2^4)\) is locally bounded. Thus it follows from (7.18), the Itô isometry, the fact that \([L, L^*]_t = [L, L^*]_t^2 + \sigma_w I dt\) and fact 7.2 that

\[
\mathbb{E} \int_0^\Delta d\text{a}_s(a_{s-}^*)
\]

\[
= \mathbb{E} \left( \int_0^\Delta (G_{s-} \otimes V_{s-}^{1/2} + V_{s-}^{1/2} \otimes G_{s-}) dL_s a_{s-}^* + \int_0^\Delta (V_{s-}^{1/2} \otimes V_{s-}^{1/2}) d(\text{vec}([L, L^*]_s) a_{s-}^*) \right)
\]

\[
= (\sigma_L + \sigma_W) \left( \int_0^\Delta \mathbb{E}(V_{s-}^{1/2} \otimes V_{s-}^{1/2}) d(\text{vec}(I d) a_{s-}^*) \right)
\]

\[
= (\sigma_L + \sigma_W) \int_0^\Delta \mathbb{E}(\text{vec}(V_{s-}) a_{s-}^*) d\text{s}.
\] (7.20)

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

\[
\int_0^\Delta \mathbb{E} a_{s-}^* d\text{s} = \int_0^\Delta \left( \text{vec}((\sigma_L + \sigma_W) s \text{E} V_0) \right)^* d\text{s} = \frac{1}{2} (\sigma_L + \sigma_W) \Delta^2 \mathbb{E} \text{vec}(V_0)^*.
\] (7.21)

Since we assumed here that all hypothesis for using Lemma 7.1 are valid, we can use (7.1) with \(\Delta = s\) to get

\[
\int_0^\Delta \text{cov}(\text{vec}(Y_{s-}), a_{s-}) d\text{s}
\]

\[
= \text{var}(\text{vec}(Y_0)) \left( \int_0^\Delta (e^{\mathcal{B}^t s} - I dt) d\text{s} \right) \left[ (\sigma_W + \sigma_L)(\mathcal{B}^*)^{-1} - 2((A \otimes A)^*)^{-1} \right]
\]

\[
= \text{var}(\text{vec}(Y_0)) \mathcal{B}^{-1}.
\] (7.22)
where $\tilde{B}$ is defined in (4.7). Using (7.20), (c.1) (7.21), (7.22) gives

$$\int_0^\Delta \mathbb{E} \text{vec}(V_{s-})a_{s-}^*ds = \int_0^\Delta \text{cov}(\text{vec}(V_{s-}), a_{s-})ds + (\mathbb{E} \text{vec}(V_0)) \int_0^\Delta \mathbb{E}(a_{s-}^*)ds$$

$$= \int_0^\Delta \text{cov}(\text{vec}(Y_{s-}), a_{s-})ds + (\mathbb{E} \text{vec}(V_0)) \int_0^\Delta \mathbb{E}(a_{s-}^*)ds$$

$$= \frac{1}{2}(\sigma_L + \sigma_W)|\Delta|^2 \mathbb{E} \text{vec}(V_0)E \text{vec}(V_0)^* + \text{var}(\text{vec}(V_0))\tilde{B}$$

$$= (\sigma_L + \sigma_W)^{-1}D,$$

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 $V_{s-} = V_{s-}^{1/2} \otimes V_{s-}^{-1/2}.$ Using (7.2), Lemma 2.1(vi) in Behme (2012) and the symmetry of $V_{s-}^{1/2}$ gives

$$[a, a^*]_\Delta$$

$$= \left[ \text{vec} \left( \int_0^\Delta V_{s-}^{1/2}dL_sG_{s-}^* + \int_0^\Delta G_{s-}dL_sV_{s-}^{1/2} + \int_0^\Delta V_{s-}^{1/2}d[L, L^*]sV_{s-}^{1/2} \right), \right]$$

$$\left( \text{vec} \left( \int_0^\Delta V_{s-}^{1/2}dL_sG_{s-}^* + \int_0^\Delta G_{s-}dL_sV_{s-}^{1/2} + \int_0^\Delta V_{s-}^{1/2}d[L, L^*]sV_{s-}^{1/2} \right) \right)_\Delta$$

$$= \int_0^\Delta f_s-d[L, L^*]s^* + \int_0^\Delta f_s-d[L, \text{vec}([L, L^*])^*]_sV_{s-}$$

$$+ \int_0^\Delta \text{vec}([L, L^*]), L^*]_s^*s^* + \int_0^\Delta \text{vec}([L, L^*], L^*]_s^*V_{s-}$$

$$:= I_1 + I_2 + I_3 + I_4.$$

By Lemma 3.2 with $p = 2$ and similar calculations as in (7.19) it follows that $\mathbb{E}\|V_{s-}\|f_{s-}\| < \infty$ for all $s > 0$ and the map $s \mapsto \mathbb{E}\|V_{s-}\|f_{s-}\|$ 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 $\mathbb{E}\|V_{s-}\|^2 < \infty$ for all $s > 0$ and local boundedness of the map $s \mapsto \mathbb{E}\|V_{s-}\|^2.$ 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) gives

$$\mathbb{E}I_4 = \mathbb{E}\left( \int_0^\Delta \text{vec}([L, L^*]), \text{vec}([L, L^*])^*V_{s-} \right)$$

$$= \mathbb{E}\left( \int_0^\Delta \text{vec}([L, L^*])^\phi, \text{vec}([L, L^*])^\phi V_{s-} \right)$$

$$= \int_0^\Delta \mathbb{E}(V_{s-}\rho_L[I_d + K_d + \text{vec}(I_d)\text{vec}(I_d)^*]V_{s-})ds$$

$$= \rho_L\int_0^\Delta (Q + K_dQ + I_d^2)\mathbb{E}(\text{vec}(V_s)\text{vec}(V_s)^*)ds$$

$$= \Delta \rho_L(Q + K_dQ + I_d^2)\mathbb{E}(\text{vec}(V_0)\text{vec}(V_0)^*).$$
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,

$$(A \otimes B + B \otimes A)(A \otimes B + B \otimes A)^* = (A \otimes B + K_d(A \otimes B))(A \otimes B + K_d(A \otimes B))^* = (I + K_d)(A \otimes B)(A^* \otimes B)(I + K_d) = (I + K_d)Q \text{vec}(A A^*) \text{vec}(B^2)(I + K_d).$$

(7.26)

Write $b_s := \mathbb{E} \text{vec}(G_s G_s^*) \text{vec}(V_s)^*$, which is finite by Lemma 3.2 with $p = 2$. Using the compensation formula, (7.26) and the definition of $f_s$ gives

$$\mathbb{E}\left(\int_0^\Delta f_s d[L, L^*]_s f_s^*\right) = (\sigma_L + \sigma_W)\int_0^\Delta \mathbb{E}(f_s f_s^*) ds = (\sigma_L + \sigma_W)\int_0^\Delta \mathbb{E}(G_s - \otimes V_s^{1/2} + V_s^{1/2} \otimes G_s -)(G_s^* - \otimes V_s^{1/2} + V_s^{1/2} \otimes G_s^*) ds$$

$$= (\sigma_L + \sigma_W)\int_0^\Delta (I + K_d)Q b_s (I + K_d) ds$$

$$= (\sigma_L + \sigma_W)(I + K_d)Q \left(\int_0^\Delta b_s ds\right)(I + K_d).$$

Finally, it follows from (7.23) that

$$\int_0^\Delta b^*_s ds = \int_0^\Delta \mathbb{E} \text{vec}(V_s)a^*_s ds = (\sigma_L + \sigma_W)^{-1} D.$$

(7.28)

The result now is a direct consequence of (7.17), (7.23), (7.24), (7.25), (7.27) and (7.28).

**Remark 7.2.** An inspection of the proofs of Lemmas 4.2 and 7.1 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 7.1 and 7.2 without (a.6) and (c.2).

**References**


