A.K.A. Kalsbeek

Continuous-time GARCH(1,1) processes

Master thesis, defended on 21 January, 2011

Thesis advisor: Dr. O.W. van Gaans

Mastertrack: Applied Mathematics

Mathematisch Instituut, Universiteit Leiden

Contents

1 Introduction

In practice, financial time series alternate between "quiet periods" and periods of high activity. The frequency of movements is constant over time, but the amplitude seems to be time-varying. This phenomenon is known as volatility clustering. Stochastic volatility processes are used to model the long-range dependence effect evident in financial time series.

One such process is known as the Linear *Generalized autoregressive conditional heteroscedastic* (GARCH) model that was introduced by Bollorsev in 1986 (see [2]). Its typical features are: "a heavy tailed, uncorrelated, but not independent, time-varying volatility and a long-range dependence effect present in the volatility". The properties (not all) and definition of this process are treated in Chapter 2.

The main objective of this thesis is to compare continuous-time GARCH models with discrete-time GARCH models. We will focus on the linear GARCH case, and mainly research two continuous-time models. The first model is derived as a limit from a discrete-time model. This will be done by scaling the parameters properly according to the time-interval, and then sending this time-interval to zero. We will follow Nelson's article dated 1990 (see [7]) and give rigorous proofs for the convergence to the continuous-time model.

The second continuous-time model is an idea of Klüppelberg, Lindner and Maller in 2004 (see [6]). The construction is given in Chapter 4 and is based on intuitive reasons. We will replace the "noise" variables by increments of a (arbitrary) Lévy process. These processes are very flexible, since for any time increment Δt any infinitely divisible distribution can be chosen as the increment distribution over periods of time Δt. On the other hand, they have a simple structure in comparison to general semimartingales, as they have independent strictly stationary increments. In Chapter 5 we will investigate what happens to the striking features that are so distinctive for the original discrete process.

Finally, we discuss some recent developments made by Kallsen and Vesenmayer in 2009 (see [5]). They have looked into a limit procedure for the continuous-time model driven by a Lévy process.

2 Linear GARCH process

2.1 The mathematical build-up

The motivation of this section comes completely from [10]. Many different GARCH-models have been developed in time. In this thesis, we will focus only on the linear GARCH model. Formally, there are two possibilities for defining a linear GARCH process. We will explain one possibility, and shortly mention the other.

Definition 2.1. A GARCH(p, q) process is a martingale difference sequence $X_n : \Omega \to \mathbb{R}$ relative to a given filtration $(\mathcal{F}_n)$, i.e. for every $n \in \mathbb{N}$ holds $X_n = W_n - W_{n-1}$ with $(W_n)_{n\in\mathbb{Z}_{\geq 0}}$ a martingale relative to $\mathcal{F}_n$, and $\mathbb{E}[W_n^2] < \infty$ for all $n \in \mathbb{Z}_{\geq 0}$. Its conditional variance $\sigma_n^2 := \mathbb{E}[X_n^2 | \mathcal{F}_{n-1}]$ satisfies for every $n \in \mathbb{N}$

$$\sigma_n^2 = \beta + \delta_1\sigma_{n-1}^2 + \cdots + \delta_p\sigma_{n-p}^2 + \lambda_1 X_{n-1}^2 + \cdots + \lambda_q X_{n-q}^2, \tag{2.1}$$

where $\beta, \delta_1, \ldots, \delta_p, \lambda_1, \ldots, \lambda_q$ are nonnegative constants.

It is not interesting when the positive square root σ_n equals zero. So we will henceforth assume $\mathbb{P}(\{\sigma_n = 0\}) = 0$ for all $n \in \mathbb{N}$. For the concrete case GARCH(1,1) we will use sufficient conditions for achieving this.

This makes it possible to define $\epsilon_n := X_n/\sigma_n$ for $n = 1, 2, \ldots$. The random variable σ_n^2 is $\mathcal{F}_{n-1}$-measurable and $t \mapsto \sqrt{t}$ is a continuous function on $[0, \infty)$. So σ_n is also $\mathcal{F}_{n-1}$-measurable. The martingale property and the definition of σ_n^2 gives

$$\forall n \in \mathbb{N} : \ \mathbb{E}[\epsilon_n | \mathcal{F}_{n-1}] = 0 \quad \text{and} \quad \mathbb{E}[\epsilon_n^2 | \mathcal{F}_{n-1}] = 1.$$

Often it is assumed that the random variables ϵ_n are i.i.d. and independent of $\mathcal{F}_{n-1}$.

Conversely, one can also define a linear GARCH process by starting with a "scaled martingale difference process" ϵ_n and a predictable process σ_n. Next, the process X_n is given by $X_n = \epsilon_n\sigma_n$. By construction we have that σ_n^2 is the conditional variance of X_n. If the process satisfies (2.1), then it is called a GARCH(p, q) process.

The abbreviation GARCH stands for "Generalized auto-regressive conditional heteroscedastic". If the coefficients $\delta_1, \ldots, \delta_q$ all vanish, then σ_n^2 is a linear function in terms of $X_{n-1}^2, \ldots, X_{n-q}^2$. In this case the model is called an

ARCH(q)-model, from "auto-regressive conditional heteroscedastic". Conditional autoregressive can be explained by the fact that $\sigma_n^2 = \mathbb{E}[X_n^2|\mathcal{F}_{n-1}]$, so in the ARCH($q$)-model equation (2.1) becomes a conditional autoregressive relation. Generalized is just added after extending the equation (2.1) by the terms $\delta_1\sigma_{n-1}^2, \ldots, \delta_p\sigma_{n-p}^2$.

For the origin of heteroscedastic we have to look at the characteristics of a white noise sequence. A white noise series is a discrete time stochastic process (Y_n) with the following properties. The series is second order stationary with mean zero, i.e.

$$\forall n \in \mathbb{Z}_{\geq 0}: \; Y_n \in \mathcal{L}^2, \quad \mathbb{E}[Y_n] = 0 \quad \text{and} \quad \gamma(h) := \operatorname{cov}(Y_{n+h}, Y_n) = \mathbb{E}[Y_{n+h}Y_n]$$

with $h \in \mathbb{Z}_{\geq 0}$. Note that $\gamma(h)$ is well-defined, because by stationarity it is independent of n for a fixed lag h. The distinctive property of a white noise sequence is given in terms of the auto-covariance function. That is, $\gamma(h) = 0$ for $h \neq 0$ and $\gamma(0) := a^2$. Here, a^2 is independent of n by stationarity. We shall speak of a heteroscedastic white noise if the auto-covariances at non-zero lags vanish, but the variances are possibly time-dependent.

Any martingale difference series (X_n) with finite second moments is a (possibly heteroscedastic) white noise series. Namely, the conditional expectation $\mathbb{E}[X_n|\mathcal{F}_{n-1}]$ is a version of the orthogonal projection of X_n onto $\mathcal{L}^2(\Omega, \mathcal{F}_{n-1}, \mathbb{P})$. Hence, $\mathbb{E}[X_n|\mathcal{F}_{n-1}]$ is the least-squares-best $\mathcal{F}_{n-1}$-measurable predictor of X_n. So for $m < n$ holds $\mathbb{E}X_nX_m = 0$, because $\mathbb{E}[X_n|\mathcal{F}_{n-1}] = 0$. In [10] a necessary and sufficient condition is given for when a second order stationary GARCH(p, q) process exists. Namely,

$$\sum_{j=1}^{\max(p,q)} (\delta_j + \lambda_j) < 1. \tag{2.2}$$

2.2 The Markov property

We will henceforth restrict ourself to the simplest non-trivial GARCH-model: "GARCH(1,1)", like in [6], with the following assumption.

Assumption 2.2. *There holds $\delta + \lambda > 0$, and all ϵ_i are non-degenerate random variables with $\mathbb{P}(\{\epsilon_i = 0\}) = 0$.*

We have deleted the unnecessary counter in the parameters. Note that this implies $\mathbb{P}(\{\sigma_n = 0\}) = 0$ for all $n \in \mathbb{N}$. It is possible to see (X_n, σ_n^2) as one process, which under appropriate conditions has the property that it is Markovian. We have to mention that only time-homogeneous Markov processes are considered in my thesis. To prove that it is a Markov process we use the following Lemmas.

Lemma 2.3. *Let h be a random variable and h_r a sequence of random variables, all with $(\Omega, \mathcal{F}, \mathbb{P})$ as their measure space. Assume that $h_r \uparrow h$ as $r \to \infty$. Then for all $y \in \mathbb{R}$ one has*

$$\lim_{r\to\infty} 1_{(-\infty,y]}(h_r) = 1_{(-\infty,y]}(h).$$

Proof. Fix $\omega \in \Omega$ and define $z := h(\omega)$. We have to distinguish two cases.

- If $1_{(-\infty,y]}(z) = 0$, then $z > y$ and $h_r(\omega) > y$ for r big enough. Hence,

$$\lim_{r\to\infty} 1_{(-\infty,y]}(h_r(\omega)) = 0.$$

- If $1_{(-\infty,y]}(z) = 1$, then $h_r(\omega) \leq z \leq y$ for all r. Hence,

$$\lim_{r\to\infty} 1_{(-\infty,y]}(h_r(\omega)) = 1.$$

□

Lemma 2.4. *Let b and ϵ be random variables and b_r a sequence of random variables, all with $(\Omega, \mathcal{F}, \mathbb{P})$ as their measure space. Assume that for all $x \in \mathbb{R}$ we have $\mathbb{P}(\{\epsilon = b\}) = 0$. If $b_r \uparrow b$ as $r \to \infty$, then one has*

$$\lim_{r\to\infty} 1_{(-\infty,b_r]}(\epsilon) \overset{a.s.}{=} 1_{(-\infty,b]}(\epsilon).$$

Moreover, if $b_r \downarrow b$ as $r \to \infty$, then this statement also follows.

Proof. Fix $\omega \in \Omega$. We start by assuming $b_r \uparrow b$ as $r \to \infty$, and distinguish two cases.

- If $1_{(-\infty,b(\omega)]}(\epsilon(\omega)) = 0$, then $\epsilon(\omega) > b(\omega) \geq b_r(\omega)$ for all r. Hence,

$$\lim_{r\to\infty} 1_{(-\infty,b_r(\omega)]}(\epsilon(\omega)) = 0.$$

- If $1_{(-\infty,b(\omega)]}(\epsilon(\omega)) = 1$, then $\epsilon(\omega) \leq b(\omega)$. By assumption we have that $F = \{\omega' \in \Omega : \epsilon(\omega') = b(\omega')\}$ is a null set. So we may almost surely assume $\epsilon(\omega) < b_r(\omega) \leq b(\omega)$ for r big enough. This yields

$$\lim_{r\to\infty} 1_{(-\infty,b_r(\omega)]}(\epsilon(\omega)) = 1.$$

Next, we assume $b_r \downarrow b$ as $r \to \infty$ and keep a $\omega \in \Omega$ fixed. Again we will have to distinguish two cases.

- If $1_{(-\infty, b(\omega)]}(\epsilon(\omega)) = 0$, then $\epsilon(\omega) > b_r(\omega) \geq b(\omega)$ for r big enough. Hence,
$$\lim_{r \to \infty} 1_{(-\infty, b_r(\omega)]}(\epsilon(\omega)) = 0.$$
- If $1_{(-\infty, b(\omega)]}(\epsilon(\omega)) = 1$, then $\epsilon(\omega) \leq b(\omega) \leq b_r(\omega)$ for all r. Thus,
$$\lim_{r \to \infty} 1_{(-\infty, b(\omega)]}(\epsilon(\omega)) = 1.$$

□

Before we are going to apply this Lemma in the proof of our upcoming Theorem, it is convenient to have sufficient conditions on the random variables ϵ and b for $\mathbb{P}(\{\epsilon = b\}) = 0$.

Proposition 2.5. *Let ϵ and b be random variables, both with $(\Omega, \mathcal{F}, \mathbb{P})$ as their measure space. If ϵ is independent of b, and the law Δ_ϵ of ϵ has a density f relative to the Lebesgue measure, i.e. $\frac{d\Delta_\epsilon}{dLeb} = f$. Then $\mathbb{P}(\{\epsilon = b\}) = 0$.*

Proof. Let Δ_b denote the law of b, and $\Delta_{\epsilon,b}$ the (joint) law of the pair (ϵ, b). By independency holds $\Delta_{\epsilon,b} = \Delta_\epsilon \times \Delta_b$. So we have

$$\begin{aligned}
\mathbb{P}(\{\epsilon = b\}) &= \mathbb{E}[1_{\{\epsilon=b\}}] \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{\{\epsilon=b\}} \, d\Delta_\epsilon \, d\Delta_b \\
&= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} 1_{\{x=y\}} f(x) \, dx \Big) \, d\Delta_b(y) \\
&= \int_{\mathbb{R}} 0 \cdot \, d\Delta_b(y) = 0.
\end{aligned}$$

□

In the article of Nelson (see [7]) it is stated that the shifted discrete-time process $(W_n, \sigma_{n+1})_{n \in \mathbb{Z}_{\geq 0}}$ is a Markov process. It is stated without all the necessary conditions and claimed without a proof. We will give a full proof for the fact that our original discrete time process $(W_n, \sigma_n)_{n \in \mathbb{Z}_{\geq 0}}$ is a Markov process. And from this it analogously follows that also the shifted process is Markovian.

Theorem 2.6. *Let all ϵ_n be i.i.d and independent of $\mathcal{F}_{n-1}$. If the law Δ_{ϵ_n} of ϵ_n (independent of n) has a density f relative to the Lebesgue measure, i.e. $\frac{d\Delta_{\epsilon_n}}{dLeb} = f$, then $(X_n, \sigma_n^2)_{n \in \mathbb{Z}_{\geq 0}}$ is a (time-homogeneous) Markov process.*

Proof. In this proof we use "The Standard Machinery". First we take σ_{n+1} to be a simple function, i.e.

$$\sigma_{n+1} = \sum_{j=1}^{m} \alpha_j 1_{F_j}$$

with $\alpha_j \in \mathbb{R}_{>0}$ (recall $\sigma_{n+1} > 0$ a.s.) and $F_j \in \mathcal{F}_n$. Without loss of generality we may assume F_j disjunct and $\cup_{j=1}^{m} F_j = \Omega$. Remember that the Borel σ-algebra $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$ is generated by the π-system (as in [12])

$$\pi(\mathbb{R} \times \mathbb{R}) = \{(-\infty, x] \times (-\infty, y] : x, y \in \mathbb{R}\}.$$

Choose $n \in \mathbb{Z}_{\geq 0}$ and $A := ((-\infty, x], (-\infty, y]) \in \pi(\mathbb{R} \times \mathbb{R})$ arbitrary. Almost surely follows

$$\begin{aligned}
\mathbb{E}[1_A(X_{n+1}, \sigma_{n+1}^2)|\mathcal{F}_n] &= \mathbb{E}[1_{(-\infty,x]}(\epsilon_{n+1}\sigma_{n+1}) \cdot 1_{(-\infty,y]}(\sigma_{n+1}^2)|\mathcal{F}_n] \\
&= \mathbb{E}[\sum_{j=1}^{m} 1_{(-\infty,x]}(\epsilon_{n+1}\alpha_j)1_{F_j}|\mathcal{F}_n] \cdot 1_{(-\infty,y]}(\sigma_{n+1}^2) \\
&= \sum_{j=1}^{m} \mathbb{E}[1_{(-\infty,\frac{x}{\alpha_j}]}(\epsilon_{n+1})|\mathcal{F}_n] \cdot 1_{F_j} \cdot 1_{(-\infty,y]}(\sigma_{n+1}^2) \\
&= \sum_{j=1}^{m} \mathbb{E}[1_{(-\infty,\frac{x}{\alpha_j}]}(\epsilon_{n+1})] \cdot 1_{F_j} \cdot 1_{(-\infty,y]}(\sigma_{n+1}^2),
\end{aligned}$$

where we used the independency and $\mathcal{F}_n$-measurability of σ_{n+1}. On $\mathbb{R}_{>0}$ we define the function

$$g_x(\alpha) := \mathbb{E}[1_{(-\infty,\frac{x}{\alpha}]}(\epsilon_{n+1})].$$

This measurable function is independent of n, because the ϵ_{n+1}'s are identically distributed. Now, the sets F_j are disjoint and together with property (2.1) this gives almost surely

$$\begin{aligned}
\mathbb{E}[1_A(X_{n+1}, \sigma_{n+1}^2)|\mathcal{F}_n] &= \sum_{j=1}^{m} g_x(\alpha_j) \cdot 1_{F_j} \cdot 1_{(-\infty,y]}(\sigma_{n+1}^2) \\
&= g_x(\sigma_{n+1}) \cdot 1_{(-\infty,y]}(\sigma_{n+1}^2) \\
&= g_x(\sqrt{\beta + \delta\sigma_n^2 + \lambda X_n^2}) \cdot 1_{(-\infty,y]}(\beta + \delta\sigma_n^2 + \lambda X_n^2) \\
&=: P_1\Big(\big(X_n(\omega), \sigma_n^2(\omega)\big), A\Big). \qquad (2.3)
\end{aligned}$$

Here, P_1 is a transition kernel on $(\mathbb{R} \times \mathbb{R}, \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}))$, which is independent of n.

Next, take σ_{n+1} to be a (non-negative) measurable function. From [12] we obtain a sequence of simple functions h_r such that $h_r \uparrow \sigma_{n+1}$ and $h_r^2 \uparrow \sigma_{n+1}^2$ as $r \to \infty$. Each h_r satisfies (2.3) with σ_{n+1} replaced by h_r, and an indicator function is trivially dominated by the measurable constant function 1. The Dominated Convergence Theorem together with the definition of conditional expectation then develops

$$\begin{aligned}
\mathbb{E}[1_A(X_{n+1}, \sigma_{n+1}^2)|\mathcal{F}_n] &= \lim_{r\to\infty} \mathbb{E}[1_A(X_{n+1}, h_r^2)|\mathcal{F}_n] \\
&= \lim_{r\to\infty} P_1\Big(\big(X_n(\omega), h_r^2(\omega)\big), A\Big) \\
&= \lim_{r\to\infty} \Big[\mathbb{E}[1_{(-\infty, \frac{x}{h_r(\omega)}]}(\epsilon_{n+1})] \cdot 1_{(-\infty,y]}(h_r^2(\omega))\Big] \\
&= \lim_{r\to\infty} \mathbb{E}[1_{(-\infty, \frac{x}{h_r(\omega)}]}(\epsilon_{n+1})] \cdot \lim_{r\to\infty} 1_{(-\infty,y]}(h_r^2(\omega))
\end{aligned}$$

Note that Proposition 2.2 gives $\mathbb{P}(\{\epsilon_{n+1} = \sigma_{n+1}\}) = 0$ for all n, because σ_{n+1} is $\mathcal{F}_n$-measurable and ϵ_{n+1} is independent of $\mathcal{F}_n$. So by Lemma 2.3 and Lemma 2.4 we have both point-wise $1_{(-\infty,y]}(h_r^2(\omega)) \to 1_{(-\infty,y]}(\sigma_{n+1}^2(\omega))$ and $1_{(-\infty, \frac{x}{h_r(\omega)}]}(\epsilon_{n+1}) \overset{a.s.}{\to} 1_{(-\infty, \frac{x}{\sigma_{n+1}(\omega)}]}(\epsilon_{n+1})$ as $r \to \infty$, where the last convergence holds for positive and negative x. We again apply The Dominated Convergence Theorem to find

$$\begin{aligned}
\mathbb{E}[1_A(X_{n+1}, \sigma_{n+1}^2)|\mathcal{F}_n] &= g_x(\sigma_{n+1}) \cdot 1_{(-\infty,y]}(\sigma_{n+1}^2) \\
&= P_1\Big(\big(X_n(\omega), \sigma_n^2(\omega)\big), A\Big). \qquad (2.4)
\end{aligned}$$

One can check that on $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$ the function

$$m_G(A) := \mathbb{E}[1_A(X_{n+1}, \sigma_{n+1}^2) \cdot 1_G]$$

is a finite measure for every $G \in \mathcal{F}_n$. So from (2.4) and the Lemma of "The Uniqueness of Extension, π-systems" (see [12]) we obtain

$$\forall G \in \mathcal{F}_n : \ \mathbb{E}\Big[1_G \cdot \mathbb{E}[1_A(X_{n+1}, \sigma_{n+1}^2)|\mathcal{F}_n]\Big] = \mathbb{E}\Big[1_G \cdot P_1\Big(\big(X_n(\omega), \sigma_n^2(\omega)\big), A\Big)\Big].$$

Hence, (2.4) holds for all $A \in \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$ and all $n \in \mathbb{Z}_{\geq 0}$.

For simplicity, we define $b\mathbb{R}$ as the space of bounded, Borel measurable functions $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. As used before, conditional expectation is defined trough integrals. So we can apply "The Standard Machinery" for a second time. One obtains for all $n \in \mathbb{Z}_{\geq 0}$ and for every $f \in b\mathbb{R}$ the equality

$$\mathbb{E}[f(X_{n+1}, \sigma_{n+1}^2)|\mathcal{F}_n] = \int_{\mathbb{R}^2} f(y) P_1\big((X_n, \sigma_n^2), dy\big). \qquad (2.5)$$

Choose $n \in \mathbb{Z}_{\geq 0}$ and $f \in b\mathbb{R}$ arbitrary. We define the transition kernel P_k inductively by

$$P_k(z, B) = \int_{\mathbb{R}^2} P_1(y, B) P_{k-1}(z, dy), \;\; k = 2, 3, \ldots.$$

Consider the function

$$\begin{array}{rcl} z_f : \mathbb{R} \times \mathbb{R} & \to & \mathbb{R} \\ x_1 \times x_2 & \mapsto & \mathbb{E}[f(X_{n+2}, \sigma^2_{n+2}) | \sigma\big(X_{n+1} = x_1, \sigma^2_{n+1} = x_2, \mathcal{F}_n\big)], \end{array}$$

where $\sigma(\cdot)$ denotes the smallest σ-algebra (on Ω) generated by it is argument. Note that $z_f \in b\mathbb{R}$. So, equality (2.5) and the Tower property gives

$$\begin{aligned} \mathbb{E}[f(X_{n+2}, \sigma^2_{n+2}) | \mathcal{F}_n] &= \mathbb{E}\big[\mathbb{E}[f(X_{n+2}, \sigma^2_{n+2}) | \mathcal{F}_{n+1}] | \mathcal{F}_n\big] \\ &= \mathbb{E}[z_f(X_{n+1}, \sigma^2_{n+1}) | F_n] \\ &= \int_{\mathbb{R}^2} z_f(y) P_1\big((X_n, \sigma^2_n), dy\big) \\ &= \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} f(u) P_1(y, du) P_1\big((X_n, \sigma^2_n), dy\big) \\ &= \int_{\mathbb{R}^2} f(y) P_2\big((X_n, \sigma^2_n), dy\big). \end{aligned}$$

Induction develops for all $f \in b\mathbb{R}$, $n \in \mathbb{Z}_{\geq 0}$ and $k \in \mathbb{N}$

$$\mathbb{E}[f(X_{n+k}, \sigma^2_{n+k})(\omega) | \mathcal{F}_n] \;\; = \;\; \int_{\mathbb{R}^2} f(y) P_k\big((X_n, \sigma^2_n), dy\big).$$

□

Remark 2.7. The assumption that all ϵ_n are i.i.d and independent of $\mathcal{F}_{n-1}$ can be relaxed. The fact that ϵ_n is independent of $\mathcal{F}_{n-1}$ and adapted to the filtration already gives that all ϵ_n are independent.

Remark 2.8. If $\mathcal{F}_n = \sigma(\epsilon_n, \epsilon_{n-1}, \ldots)$ (is equal to $\sigma(X_n, X_{n-1}, \ldots)$), then one can prove that all ϵ_n are i.i.d. implies that ϵ_n is independent of $\mathcal{F}_{n-1}$.

Remark 2.9. From equation (2.3) one can deduce that under the same conditions the process $(\sigma^2_n)_{n \in \mathbb{Z}_{\geq 0}}$ is also a (time-homogeneous) Markov process.

Sometimes it is useful to look at W_n instead of X_n, because W_n is a martingale.

Corollary 2.10. *Let all ϵ_n be i.i.d and independent of $\mathcal{F}_{n-1}$. If the law Δ_{ϵ_n} of ϵ_n (independent of n) has a density f relative to the Lebesgue measure, i.e. $\frac{d\Delta_{\epsilon_n}}{dLeb} = f$, then $(W_n, \sigma^2_{n+1})_{n\in\mathbb{Z}_{\geq 0}}$ is a (time-homogeneous) Markov process.*

Proof. We will sketch the proof. The set $\pi(\mathbb{R}\times\mathbb{R})$ is defined as in the proof of Theorem 2.6. Let $n \in \mathbb{Z}_{\geq 0}$ and $A := ((-\infty, x], (-\infty, y]) \in \pi(\mathbb{R}\times\mathbb{R})$. There holds

$$\begin{aligned}
\mathbb{E}[1_A(W_{n+1}, \sigma^2_{n+2})|\mathcal{F}_n] &= \mathbb{E}[1_A(\epsilon_{n+1}\sigma_{n+1} + W_n, \sigma^2_{n+2})|\mathcal{F}_n] \\
&= \mathbb{E}[1_{(-\infty,x]}(\epsilon_{n+1}\sigma_{n+1} + W_n)\cdot 1_{(-\infty,y]}(\sigma^2_{n+2})|\mathcal{F}_n] \\
&= \mathbb{E}\big[1_{(-\infty,y]}(\sigma^2_{n+2})\Big(1_{(-\infty,x]}(\epsilon_{n+1}\sigma_{n+1})\cdot 1_{(-\infty,0]}(W_n) \\
&\qquad + \sum_{n\in\mathbb{N}} 1_{(-\infty,x-n]}(\epsilon_{n+1}\sigma_{n+1})\cdot 1_{(n-1,n]}(W_n)\Big)|\mathcal{F}_n\big],
\end{aligned}$$

where

$$\sigma^2_{n+2} = \beta + \delta\sigma^2_{n+1} + \lambda(\epsilon_{n+1}\sigma_{n+1} - W_n).$$

Recall that σ_{n+1} and W_n are both $\mathcal{F}_n$-measurable. If we use "The Standard Machinery" and start with a simple function on the random variable σ_{n+1}, then analogously as the obtained equation (2.3) we obtain

$$\mathbb{E}[1_A(W_{n+1}, \sigma^2_{n+2})|\mathcal{F}_n] =: \tilde{P}_1\Big(\big(W_n(\omega), \sigma^2_{n+1}(\omega)\big), A\Big), \tag{2.6}$$

with $\tilde{P}_1$ a transition kernel on $(\mathbb{R}\times\mathbb{R}, \mathcal{B}(\mathbb{R})\times\mathcal{B}(\mathbb{R}))$. As in the proof of Theorem 2.6, we invoke our Lemmas and Proposition to conclude that (2.6) holds for an arbitrary (non-negative) σ_{n+1}. The final part of the proof also follows analogously. □

Remark 2.11. We see that under the same conditions of Theorem 2.6 (and Corollary 2.10), the process $(W_n, \sigma^2_n)_{n\in\mathbb{Z}_{\geq 0}}$ is not Markovian.

3 Continuous-time model 1: Diffusion approximation

3.1 Set-up

Many different parameterizations have been made for the function σ^2 instead of (2.1). The GARCH(p,q)- and the ARCH(p)-model are just the most famous models. In 1990 Nelson published the paper *"Arch Models as Diffusion Approximations"* (see [7]). Nelson developed conditions under which our known discrete systems converge in distribution to an Itô process. This was done by looking at the difference equations of our models, and letting the length of the time interval go to zero in an "appropriate way". Obviously, one of the conditions was given in the fact that the discrete time model is Markovian.

In this case it is, for reasons mentioned in Remark 2.10, more convenient to look at W_n instead of X_n. We want to partition the time of the system GARCH$(1,1)$ more and more finely. So, for each $h > 0$ we are going to define a GARCH$(1,1)$ process $(W_{kh}, \sigma^2_{(k+1)h})_{k\in\mathbb{Z}_{\geq 0}}$, where the nonnegative parameters β, δ and λ in equation (2.1) are depending on the time interval h. One must keep in mind how we have defined the original GARCH model. We have chosen to start from defining a martingale difference sequence X_n and $\sigma_n^2 = \mathbb{E}[X_n^2|\mathcal{F}_{n-1}]$, whereupon we defined the noise variable ϵ_n. Let us repeat this with a diffusive scaling for the martingale sequence.

So, for each $k \in \mathbb{N}, h > 0$ consider $(W_{kh}, \sigma^2_{(k+1)h})_{k\in\mathbb{Z}_{\geq 0}}$ given by

$$\begin{aligned} W_{kh} &= W_{(k-1)h} + \sqrt{h}\epsilon_{kh}\sigma_{kh} \\ \sigma^2_{(k+1)h} &= \beta_h + \delta_h\sigma^2_{kh} + \lambda_h\epsilon^2_{kh}\sigma^2_{kh}, \end{aligned} \tag{3.1}$$

with $\sigma^2_{(k+1)h} = \mathbb{E}[X^2_{(k+1)h}|\mathcal{G}_{(k-1)h}]$. Let $\epsilon_{kh} = X_{kh}/\sigma_{kh}$ and $(\mathcal{G}_{kh})_{k\in\mathbb{Z}_{\geq 0}}$ as filtration for the martingale $(W_{kh})_{k\in\mathbb{Z}_{\geq 0}}$. There still holds $\mathbb{E}[W^2_{kh}] < \infty$ for all $k \in \mathbb{Z}_{\geq 0}$, and

$$\forall k \in \mathbb{N}: \ \mathbb{E}[\epsilon_{kh}|\mathcal{G}_{(k-1)h}] = 0 \quad \text{and} \quad \mathbb{E}[\epsilon^2_{kh}|\mathcal{G}_{(k-1)h}] = 1. \tag{3.2}$$

Thus, the dependency of h can only be found in the factor $\sqrt{h}$ and the parameters β_h, δ_h and λ_h.

We need to assume some properties of the initial distribution.

Assumption 3.1. *For each $h > 0$ the initial probability law for each h is given by*

$$\nu(\Gamma) := \mathbb{P}\big[(W_0, \sigma_h^2) \in \Gamma\big], \quad \textit{for any } \Gamma \in \mathcal{B}(\mathbb{R}^2),$$

and $\sigma_h^2 \overset{d}{=} \sigma_0^2$ with $\sigma_h^2 \in \mathcal{L}^4$.

So the initial probability law is the same for each h. The time-difference between W_{kh} and $\sigma^2_{(k+1)h}$ does not create a problem for σ_0^2. The initial moment property is important for future reasons. With all the ingredients in place, we are capable of defining a continuous-time process for each h.

Definition 3.2. Let $h > 0$. The continuous-time GARCH$(1,1)_h$ process $(W_{t,h}, \sigma^2_{t+h,h})_{t\geq 0}$ with filtration $\mathbb{F}_h := (\mathcal{F}_{t,h})_{t\geq 0}$ is given by

$$W_{t,h} = W_{kh}, \quad \sigma^2_{t+h,h} = \sigma^2_{(k+1)h} \quad \text{and} \quad \mathcal{F}_{t,h} = \mathcal{G}_{kh}, \quad t \geq 0,$$

where $kh \leq t < (k+1)h$ for a unique $k \in \mathbb{Z}_{\geq 0}$.

As already mentioned in the previous chapter, the process is not interesting when the positive square root $\sigma_{t,h}$ equals zero. Similar to Assumption 2.2 we have the following.

Assumption 3.3. *There holds for each $h > 0$*

$$\nu[(W_{0,h}, \sigma^2_{0,h}) \in \mathbb{R} \times (0, \infty)] = 1$$

and $\delta_h + \lambda_h > 0$. Also, the noise variables ϵ_{kh} are non-degenerate for all $k \in \mathbb{N}$, $h > 0$.

3.2 Technical preliminaries

We have just defined a continuous-time GARCH$(1,1)_h$ process. If this process obeys certain requirements, then GARCH$(1,1)_h$ converges in distribution to a Itô process when $h \downarrow 0$. These requirements are based on Theorem 2.1 in [7] and need some preparation to prove. Therefore, in this section we do some technical preparatory work. We state without proof the following simple Lemma.

Lemma 3.4. *Let $(a_k)_{k\in\mathbb{Z}_{\geq 0}}$ be sequence in $\mathbb{R}$ and $\alpha, b \in \mathbb{R}$.*

(a) If

$$a_{k+1} = \alpha a_k + b, \quad \forall k \in \mathbb{Z}_{\geq 0},$$

then

$$a_k = (a_0 - \frac{b}{1-\alpha})\alpha^k + \frac{b}{1-\alpha}, \quad \forall k \in \mathbb{Z}_{\geq 0}.$$

(b) If

$$a_{k+1} \leq \alpha a_k + b, \quad \forall k \in \mathbb{Z}_{\geq 0},$$

then

$$a_k \leq (a_0 - \frac{b}{1-\alpha})\alpha^k + \frac{b}{1-\alpha}, \quad \forall k \in \mathbb{Z}_{\geq 0}.$$

This recursive result is imporant for proving moment properties of our discrete volatility process.

Lemma 3.5. *Let $R \geq 0$. Let for each $h > 0$ the random variables ϵ_{kh}, with $k \in \mathbb{N}_{\geq 2}$, be i.i.d, independent of $\mathcal{G}_{(k-1)h}$ with $\epsilon_{kh} \in \mathcal{L}^4$. If the limit is*

$$\begin{aligned} \beta &:= \lim_{h \downarrow 0} \frac{\beta_h}{h} \\ \theta &:= \lim_{h \downarrow 0} \frac{1 - \delta_h - \lambda_h}{h} \end{aligned}$$

exist, and

$$\limsup_{h \downarrow 0} \frac{\lambda_h^2}{h} < \frac{2\theta}{\mathbb{E}[\epsilon_{kh}^4] - 1}, \quad k \in \mathbb{N}_{\geq 2},$$

with $\theta \in \mathbb{R}_{>0}$, $\beta \in \mathbb{R}_{\geq 0}$, then $\exists \epsilon^{(3)} \in \mathbb{R}_{>0}$, $\exists K \in \mathbb{R}_{>0}$ such that for all $k \in \mathbb{N}_{\geq 2}$, $0 < h \leq \epsilon^{(3)}$ and all $||(W_{(k-1)h}, \sigma_{kh}^2)(\omega)|| \leq R$ a.s. holds $\max(\mathbb{E}[\sigma_{kh}^2], \mathbb{E}[\sigma_{kh}^4]) < K$.

Proof. Choose $R \geq 0$ arbitrary. By the assumed properties of the noise variable we have the following relation

$$\mathbb{E}[\sigma_{(k+1)h}^2] = \beta_h + (\delta_h + \lambda_h)\mathbb{E}[\sigma_{kh}^2], \quad k \in \mathbb{N}.$$

We define for each $h > 0$

$$a_{k,h} \quad := \quad \mathbb{E}[\sigma_{k,h}^2], \quad k \in \mathbb{N}_{\geq 2},$$

and

$$\begin{aligned} \alpha_h &:= \delta_h + \lambda_h \\ c &:= \mathbb{E}\sigma_0^2 - \frac{\beta_h}{1-\alpha_h} = \mathbb{E}\sigma_h^2 - \frac{\beta_h}{1-\alpha_h}. \end{aligned}$$

By Lemma 3.4 we have

$$a_{k,h} = c \cdot \alpha_h^k + \frac{\beta_h}{1-\alpha_h}, \quad k \in \mathbb{N}_{\geq 2}, \; h > 0.$$

Note that $\alpha_h < 1$ for h small enough, because $\theta > 0$. Also,

$$\lim_{h \downarrow 0} \frac{\beta_h}{1-\alpha_h} = \frac{\beta}{\theta}.$$

Thus, by assumption 3.1 there exists a $\epsilon^{(1)} > 0$ and a certain $K^{(1)} \in \mathbb{R}_{>0}$ such that

$$a_{k,h} \leq K^{(1)},$$

for all $0 < h \leq \epsilon^{(1)}$, $k \in \mathbb{N}_{\geq 2}$ and all $||(W_{(k-1)h}, \sigma_{kh})(\omega)|| \leq R$.

Also by independency we have

$$\begin{aligned} \mathbb{E}\sigma^4_{(k+1)h} &= \mathbb{E}[\Big((\beta_h + \sigma^2_{kh} - \sigma^2_{kh}(1-\delta_h - \lambda_h\epsilon^2_{kh})\Big)^2] \\ &= \beta_h^2 + 2\beta_h(\delta_h+\lambda_h)\sigma^2_{kh} + \mathbb{E}[(\delta_h + \lambda_h\epsilon^2_{kh})^2]\mathbb{E}[\sigma^4_{kh}], \quad k \in \mathbb{N}_{\geq 2}. \end{aligned}$$

For $h > 0$ we define

$$d_h \;:=\; \frac{\beta_h^2 + 2\beta_h(\delta_h+\lambda_h)K^{(1)}}{1-\mathbb{E}[(\delta_h+\lambda_h\epsilon^2_{kh})^2]}.$$

So Lemma 3.4 yields for all $k \in \mathbb{N}_{\geq 2}$

$$\mathbb{E}[\sigma^4_{kh}] \;\leq\; (\mathbb{E}\sigma^4_0 - d_h)\cdot\Big(\mathbb{E}[(\delta_h+\lambda_h\epsilon^2_{kh})^2]\Big)^k + d_h, \quad h > 0. \tag{3.3}$$

This suggests to take a closer look at d_h and at $\mathbb{E}[(\delta_h+\lambda_h\epsilon^2_{kh})^2]$.

The numerator of d_h obviously has $\beta_h^2 + 2\beta_h(\delta_h+\lambda_h)K = O(h)$ as $h \downarrow 0$. For the denominator of d_h we have

$$\begin{aligned} 1-\mathbb{E}[(\delta_h+\lambda_h\epsilon^2_{kh})^2] &= 1-\delta_h^2 - 2\delta_h\lambda_h - \lambda_h^2\mathbb{E}[\epsilon^4_{kh}] \\ &= -\Big((1-\delta_h)^2 - 2(1-\delta_h) + 2\delta_h\lambda_h + \lambda_h^2\mathbb{E}[\epsilon^4_{kh}]\Big). \end{aligned}$$

The bad case scenario would be if denominator tends to zero too fast when $h \downarrow 0$. A sufficient condition to avoid this is as follows

$$\limsup_{h\downarrow 0} \frac{1}{h}\big[(1-\delta_h)^2 - 2(1-\delta_h) + 2\delta_h\lambda_h + \lambda_h^2\mathbb{E}[\epsilon^4_{kh}]\big] \;<\; 0.$$

We have $1-\delta_h = \lambda_h + h\theta + o(h)$ as $h \downarrow 0$. Hence,

$$\begin{aligned} &(1-\delta_h)^2 - 2(1-\delta_h) + 2\delta_h\lambda_h + \lambda_h^2\mathbb{E}[\epsilon^4_{kh}] \\ =\;& (\lambda_h + h\theta)^2 - 2(\lambda_h+h\theta) + 2\delta_h\lambda_h + \lambda_h^2\mathbb{E}[\epsilon^4_{kh}] + o(h) \\ =\;& \lambda_h^2 + 2\lambda_h h\theta - 2h\theta - 2\lambda_h(\lambda_h + h\theta) + \lambda_h^2\mathbb{E}[\epsilon^4_{kh}] + o(h) \\ =\;& \lambda_h^2(\mathbb{E}[\epsilon^4_{kh}] - 1) - 2h\theta + o(h) \end{aligned} \tag{3.4}$$

as $h \downarrow 0$. Our sufficient condition is given by

$$\limsup_{h\downarrow 0} \frac{\lambda_h^2}{h} < \frac{2\theta}{\mathbb{E}[\epsilon_{kh}^4] - 1}, \quad k \in \mathbb{N}_{\geq 2}.$$

Note that independency and non-degeneracy (see Assumption 3.3) of ϵ_{kh} in combination with Jensen's inequality gives $\mathbb{E}[\epsilon_{kh}^4] > 1$. So the lim sup-condition gives a $\gamma > 0$, independent of h, such that for h small enough holds

$$(\mathbb{E}[\epsilon_{kh}^4] - 1)\lambda_h^2 < h(2\theta - \gamma).$$

In combination with (3.4) we develop

$$\lambda_h^2(\mathbb{E}[\epsilon_{kh}^4] - 1) - 2h\theta + o(h) < \gamma_h + o(h)$$

as $h \downarrow 0$. So there exists a $0 < \epsilon^{(2)} \leq \epsilon^{(1)}$ such that

$$\mathbb{E}[(\delta_h + \lambda_h \epsilon_{kh}^2)^2] = \delta_h^2 + \lambda_h^2 \mathbb{E}[\epsilon_{kh}^4] + 2\delta_h \lambda_h < 1, \quad k \in \mathbb{N}_{\geq 2},$$

for all $0 < h \leq \epsilon^{(2)}$.

Since $\epsilon_{kh} \in \mathcal{L}^4$, inequality (3.3) and Assumption 3.1 give us a $0 < \epsilon^{(3)} \leq \epsilon^{(2)}$ and a certain $K^{(2)} \in \mathbb{R}_{>0}$ such that

$$\mathbb{E}[\sigma_{kh}^4] \leq K^{(2)},$$

for all $0 < h \leq \epsilon^{(3)}$, $k \in \mathbb{N}_{\geq 2}$ and all $||(W_{(k-1)h}, \sigma_{kh}^2|| \leq R$. Finally, take $K = \max(K^{(1)}, K^{(2)})$. □

Lemma 3.6. *Let $R \geq 0$. Let $f_1, f_2 : \Omega \to \mathbb{R}_{\geq 0}$ be two measurable functions with bounded expecations and both independent of ϵ_{kh}, for all $k \in \mathbb{N}_{\geq 2}$ and all $h > 0$. If the conditions of Lemma 3.5 hold and the extra requirement*

$$\limsup_{h\downarrow 0} \frac{\lambda_h^2}{h} < \frac{2\theta}{3(\mathbb{E}[\epsilon_{kh}^4] - 1)} \quad \textit{and} \quad \epsilon_{kh} \in \mathcal{L}^8, \quad k \in \mathbb{N}_{\geq 2},$$

then $\exists \epsilon^{(5)} \in \mathbb{R}_{>0}$, $\exists N \in \mathbb{R}_{>0}$ such that for all $k \in \mathbb{N}_{\geq 2}$, $0 < h \leq \epsilon^{(5)}$ and all $||(W_{(k-1)h}, \sigma_{kh}^2)(\omega)|| \leq R$ a.s. holds $\mathbb{E}[f_1 \sigma_{kh}^6] \leq N$ and $\mathbb{E}[f_2 \sigma_{kh}^8] \leq N$.

Proof. The proof is similar to the proof of Lemma 3.5. Hence, it is technical. Choose $R \geq 0$ arbitrary. There holds (please verify) for all $h > 0$ and $k \in \mathbb{N}_{\geq 2}$

$$\begin{aligned} & \mathbb{E}[\sigma_{(k+1)h}^6] \\ = \; & \mathbb{E}[\beta_h^3 + \sigma_{kh}^2(3\beta_h^2\delta_h + 3\beta_h^2\lambda_h) \\ & + \sigma_{kh}^4(3\beta_h\delta_h^2 + 6\beta_h\delta_h\lambda_h + 3\beta_h\lambda_h^2\epsilon_{kh}^4) \\ & + \sigma_{kh}^6(\delta_h^3 + 3\delta_h^2\lambda_h + 3\delta_h\lambda_h^2\epsilon_{kh}^4 + \lambda_h^3\epsilon_{kh}^6)] \end{aligned}$$

and

$$\begin{aligned}
& \mathbb{E}[\sigma^8_{(k+1)h}] \\
= \; & \mathbb{E}[\beta_h^4 + \sigma^2_{kh}(4\beta_h^3\lambda_h + 4\beta_h^3\delta_h) \\
& + \sigma^4_{kh}(6\beta_h^2\lambda_h^2\epsilon^4_{kh} + 6\beta_h^2\delta_h^2 + 12\beta_h^2\delta_h\lambda_h) \\
& + \sigma^6_{kh}(4\beta_h\lambda_h^3\epsilon^6_{kh} + 4\beta_h\delta_h^3 + 12\beta_h\delta_h^2\lambda_h + 12\beta_h\delta_h\lambda_h^2\epsilon^4_{kh}) \\
& + \sigma^8_{kh}(4\delta_h^3\lambda_h + 6\delta_h^2\lambda_h^2\epsilon^4_{kh} + 4\delta_h\lambda_h^3\epsilon^6_{kh} + \lambda_h^4\epsilon^8_{kh} + \delta_h^4)].
\end{aligned}$$

Lemma 3.5 gives a $\epsilon^{(3)} > 0$ and a $K \in \mathbb{R}_{>0}$ such that $\max(\mathbb{E}[\sigma^2_{kh}], \mathbb{E}[\sigma^4_{kh}]) \leq K$ for all $0 < h < \epsilon^{(3)}$, $k \in \mathbb{N}_{\geq 2}$. Take $D := \max(\mathbb{E}[f_1], \mathbb{E}[f_2])$. Now, let us take a closer look at the sixth and eighth moment.

Sixth moment: We define

$$\begin{aligned}
b \; &:= \; \beta_h^3 + K(3\beta_h^2\delta_h + 3\beta_h^2\lambda_h) \\
& \quad + K(3\beta_h\delta_h^2 + 6\beta_h\delta_h\lambda_h + 3\beta_h\lambda_h^2\mathbb{E}[\epsilon^4_{kh})], \\
\alpha \; &:= \; \delta_h^3 + 3\delta_h^2\lambda_h + 3\delta_h\lambda_h^2\mathbb{E}[\epsilon^4_{kh}] + \lambda_h^3\mathbb{E}[\epsilon^6_{kh}]
\end{aligned}$$

and

$$a_{k,h} := \mathbb{E}[f_1\sigma^6_{k,h}] = \mathbb{E}[f_1] \cdot \mathbb{E}[\sigma^6_{k,h}], \quad h > 0, \; k \in \mathbb{N}_{\geq 2},$$

By Lemma 3.4 we have

$$a_{k,h} \leq (a_{0,h} - \frac{D \cdot b}{1-\alpha})\alpha^k + \frac{D \cdot b}{1-\alpha}, \quad k \in \mathbb{N}_{\geq 2}, \tag{3.5}$$

for all $0 < h \leq \epsilon^{(3)}$. Note that

$$b \;=\; O(3\beta_h\delta_h^2) = O(h)$$

as $h \downarrow 0$. We have

$$\begin{aligned}
\alpha \; &= \; \delta_h^3 + 3\delta_h^2\lambda_h + 3\delta_h\lambda_h^2 + \lambda_h^3 \\
& \quad + 3\delta_h\lambda_h^2(\mathbb{E}[\epsilon^4_{kh}] - 1) + \lambda_h^3(\mathbb{E}[\epsilon^6_{kh}] - 1) \\
&= \; (1 - h\theta)^3 + 3\delta_h\lambda_h^2(\mathbb{E}[\epsilon^4_{kh}] - 1) + \lambda_h^3(\mathbb{E}[\epsilon^6_{kh}] - 1) + o(h) \\
&= \; 1 - 3h\theta + 3\delta_h\lambda_h^2(\mathbb{E}[\epsilon^4_{kh}] - 1) + \lambda_h^3(\mathbb{E}[\epsilon^6_{kh}] - 1) + o(h)
\end{aligned}$$

as $h \downarrow 0$. Moreover, $1 - \alpha = O(h) + o(h)$ as $h \downarrow 0$. Since $\epsilon_{kh} \in \mathcal{L}^6 \subset \mathcal{L}^8$ and the distribution of ϵ_{kh} is independent of h,

$$\alpha \;<\; 1 - 3h\theta + 2h\delta_h\theta + o(h)$$

as $h \downarrow 0$. This yields,

$$\limsup_{h\downarrow 0} \frac{D \cdot b}{1-\alpha} \in \mathbb{R}_{>0}, \quad k \in \mathbb{N}_{\geq 2}.$$

So there exists a $0 < \epsilon^{(4)} \leq \epsilon^{(3)}$ and a $C_1 \in \mathbb{R}$ such that $0 < \alpha < 1$ and $\limsup_{h\downarrow} \frac{D\cdot b}{1-\alpha} < C_1$, for all $0 < h \leq \epsilon^{(4)}$, $k \in \mathbb{N}_{\geq 2}$.

Hence, inequality (3.5) and Assumption 3.1 gives a $N^{(1)} \in \mathbb{R}_{>0}$ such that

$$\mathbb{E}[f_1 \cdot \sigma_{kh}^6] \leq N^{(1)},$$

for all $0 < h \leq \epsilon^{(4)}$, $k \in \mathbb{N}_{\geq 2}$ and all $||(W_{(k-1)h}, \sigma_{kh}^2)(\omega)|| \leq R$.

Eighth moment: Let $N_*^{(1)}$ and $\epsilon_*^{(4)}$ be obtained by using f_2, as respectively $N^{(1)}$ and $\epsilon^{(4)}$ were obtained by using f_1. We define

$$\begin{aligned}
q &:= D \cdot \beta_h^4 + D \cdot K(4\beta_h^3\lambda_h + 4\beta_h^3\delta_h) \\
&\quad + D \cdot K(6\beta_h^2\lambda_h^2\mathbb{E}[\epsilon_{kh}^4] + 6\beta_h^2\delta_h^2 + 12\beta_h^2\delta_h\lambda_h) \\
&\quad N_*^{(1)}\big(4\beta_h\lambda_h^3\mathbb{E}[\epsilon_{kh}^6] + 4\beta_h\delta_h^3 + 12\beta_h\delta_h^2\lambda_h + 12\beta_h\delta_h\lambda_h^2\mathbb{E}[\epsilon_{kh}^4]\big), \\
\xi &:= (4\delta_h^3\lambda_h + 6\delta_h^2\lambda_h^2\mathbb{E}[\epsilon_{kh}^4] + 4\delta_h\lambda_h^3\mathbb{E}[\epsilon_{kh}^6] + \lambda_h^4\mathbb{E}[\epsilon_{kh}^8] + \delta_h^4).
\end{aligned}$$

and

$$d_{k,h} = \mathbb{E}[f_2\sigma_{k,h}^8] = \mathbb{E}[f_2] \cdot \mathbb{E}[\sigma_{k,h}^8], \quad h > 0,\ k \in \mathbb{N}_{\geq 2}.$$

By Lemma 3.4 we have

$$d_{k,h} \leq (d_{0,h} - \frac{q}{1-\xi})\xi^k + \frac{q}{1-\xi}, \quad k \in \mathbb{N}_{\geq 2}, \tag{3.6}$$

for all $0 < h \leq \epsilon_*^{(4)}$. Recall (from the proof of Lemma 3.5) that $\mathbb{E}[\epsilon_{kh}^4] > 1$. So here, our lim sup-condition gives a $\gamma > 0$, independent of h, such that for h small enough holds

$$(\mathbb{E}[\epsilon_{kh}^4] - 1)\lambda_h^2 < h(\frac{2}{3}\theta - \gamma). \tag{3.7}$$

For the base ξ follows

$$\begin{aligned}
\xi &= (\delta_h + \lambda_h)^4 + 6\delta_h^2\lambda_h^2(\mathbb{E}[\epsilon_{kh}^4] - 1) + 4\delta_h\lambda_h^3(\mathbb{E}[\epsilon_{kh}^6] - 1) + \lambda_h^4(\delta_h^4\mathbb{E}[\epsilon_{kh}^8] - 1) \\
&= 1 - 4h\theta + 6\delta_h^2\lambda_h^2(\mathbb{E}[\epsilon_{kh}^4] - 1) + o(h) \\
&< 1 - h(4\theta + 4\delta_h^2\gamma - 4\delta_h^2\theta) + o(h),
\end{aligned}$$

as $h \downarrow 0$ due to (3.7). Hence,

$$1 - \xi > h(4\theta + 4\delta_h^2\gamma - 4\delta_h^2\theta)$$

for h small enough. Also, we have that $q = O(N_*^{(1)} 4\beta_h \delta_h^3) = O(h)$ as $h \downarrow 0$. This yields

$$\limsup_{h\downarrow 0} \frac{q}{1-\xi} \in \mathbb{R}, \quad k \in \mathbb{N}_{\geq 2},$$

So there exists a $0 < \epsilon^{(5)} \leq \epsilon_*^{(4)}$ and a $C_2 \in \mathbb{R}$ such that $0 < \xi < 1$ and $\frac{q}{1-\xi} < C_2$, for all $0 < h \leq \epsilon^{(5)}$, $k \in \mathbb{N}_{\geq 2}$.

Inequality (3.6) and Assumption 3.1 gives a $N^{(2)} \in \mathbb{R}_{>0}$ such that

$$\mathbb{E}[f_2 \cdot \sigma_{kh}^8] \leq N^{(2)},$$

for all $0 < h \leq \epsilon^{(5)}$, $k \in \mathbb{N}_{\geq 2}$ and all $||(W_{(k-1)h}, \sigma_{kh}^2)(\omega)|| \leq R$.

We conclude by taking $N = \max(N^{(1)}, N^{(2)})$. □

Remark 3.7. The condition in terms of "lim sup" can be interpreted as a restriction on the noise variable ϵ_{kh} in combination with λ. Namely, it is tail can't be too "fat" and/or the contribution of the noise can't be too large. For the standard normal distribution as noise variable, our strongest condition transforms into $\limsup_{h\downarrow 0} \frac{\lambda_h^2}{h} < \frac{1}{3}\theta$, which is certainly an acceptable requirement. Also, $\theta > 0$ is reasonable if we keep the necessary and sufficient condition (2.2) for existence of a second order stationary process in mind.

Remark 3.8. Note that for obtaining $\mathbb{E}[f_1 \sigma_{kh}^6] \leq N^{(1)}$ the weaker assumption $\limsup_{h\downarrow 0} \frac{\lambda_h^2}{h} < \frac{\theta}{\mathbb{E}[\epsilon_{kh}^4]-1}$ was enough. Namely, use a γ analogous as in (3.7).

We mention that Corollary 2.10 gives, under appropriate conditions, for each h a collection of (homogeneous) transition kernels $(\tilde{P}_{kh})_{k\in\mathbb{N}}$ for the process $(W_{kh}, \sigma_{(k+1)h}^2)_{k\in\mathbb{Z}_{\geq 0}}$. Let $\lfloor t/h \rfloor$ denote the integer part of t/h, i.e., the largest integer k such that $k \leq t/h$. Then, according to Definition 3.2, we also have for each h a collection of (homogeneous) transition kernels $(\tilde{P}_{t,h})_{t\geq 0}$ given by

$$\forall t \geq 0: \ P_{t,h} := \tilde{P}_{h\lfloor t/h \rfloor}. \tag{3.8}$$

Henceforth, it is easier that we use the following operator notation for a measurable function f on $\mathbb{R}^2$

$$P_{s,h}(f)(x) := \int_{\mathbb{R}^2} f(y) \ P_{s,h}(x, dy), \quad h \geq 0, \ s \geq 0,$$

with transition kernel $P_{s,h}$ as in (3.8). And x^T denotes the transpose of a vector x (in $\mathbb{R}^2$). The following Lemma gives expressions for the drift en second moment per unit of time.

Lemma 3.9. *We let the following functions be given on* $\mathbb{R}^2$

$$\begin{aligned} g^{(k)}(y) &= \big(y - (W_{(k-1)h}, \sigma^2_{kh})\big)^T \\ w^{(k)}(y) &= \big(y - (W_{(k-1)h}, \sigma^2_{kh})\big)^T \cdot \big(y - (W_{(k-1)h}, \sigma^2_{kh})\big), \quad k \in \mathbb{N}. \end{aligned}$$

Let $R \geq 0$ *and for each* $h > 0$ *the random variables* ϵ_{kh}*, with* $k \in \mathbb{N}_{\geq 2}$*, be i.i.d, independent of* $\mathcal{G}_{(k-1)h}$*. If* $\forall k \in \mathbb{N}_{\geq 2}$ *and* $h > 0$ *the law* $\Delta_{\epsilon_{kh}}$ *of* ϵ_{kh} *has a density with respect to the Lebesgue measure, then holds*

$$\frac{1}{h} P_{h,h}(g^{(k)}(x) = \begin{pmatrix} 0 & \frac{\beta_h - \sigma^2_{kh}(1-\delta_h-\lambda_h)}{h} \end{pmatrix},$$

and

$$\begin{aligned} &\frac{1}{h} P_{h,h}(w^{(k)}(x)) \\ = &\begin{pmatrix} \sigma^2_{kh} & 0 \\ 0 & \frac{\beta_h^2}{h} - \frac{2\beta_h\sigma^2_{kh}}{h}(1-\delta_h-\lambda_h) + \frac{\sigma^4_{kh}}{h}(1-\delta_h-\lambda_h)^2 + \frac{\lambda_h^2\sigma^4_{kh}(M-1)}{h} \end{pmatrix}, \end{aligned}$$

where $x := (W_{(k-1)h}, \sigma^2_{kh})$ *such that* $||x|| \leq R$*, and where* $M := \mathbb{E}[\epsilon^4_{kh}]$*.*

Proof. Let $R \geq 0$ and $h > 0$ small enough (such that we can invoke Lemma 3.5). Fix $k \in \mathbb{N}_{\geq 2}$ such that for our random starting point $x = (W_{(k-1)h}, \sigma^2_{kh})$ holds $||x|| \leq R$. Let the indexes i and j in the functions $g_i^{(k)}$ and $w_{i,j}^{(k)}$ denote the matrix entry, with $i, j \in \{1, 2\}$. Conditioned on information at time $(k-1)h$, the martingale property tells us that

$$\frac{1}{h} P_{h,h}(g_1^{(k)})(x) = 0.$$

We use $X_{kh} = \sigma_{kh}\epsilon_{kh}$ in combination with (3.2) to obtain

$$\begin{aligned} \frac{1}{h} P_{h,h}(g_2^{(k)})(x) &= \frac{\mathbb{E}[\beta_h + \sigma^2_{kh}(\delta_h + \lambda_h\epsilon^2_{kh} - 1)|\mathcal{G}_{(k-1)h}]}{h} \\ &= \frac{\beta_h - \sigma^2_{kh}(1-\delta_h-\lambda_h)}{h}, \end{aligned}$$

by $\mathcal{G}_{(k-1)h}$ measurability of σ_{kh}. So we have an expression for the drift per unit of time. Now, we look at the second moment per unit of time. The independence property of the noise variables and direct computation tells us the following

$$\begin{aligned} &\frac{1}{h} P_{h,h}(w_{1,1}^{(k)})(x) \\ = &\frac{\beta_h^2}{h} - \frac{2\beta_h\sigma^2_{kh}}{h}(1-\delta_h-\lambda_h) + \frac{\sigma^4_{kh}}{h}\mathbb{E}[(\delta_h + \lambda_h\epsilon^2_{kh} - 1)^2|\mathcal{G}_{(k-1)h}] \\ = &\frac{\beta_h^2}{h} - \frac{2\beta_h\sigma^2_{kh}}{h}(1-\delta_h-\lambda_h) + \frac{\sigma^4_{kh}}{h}(1-\delta_h-\lambda_h)^2 + \frac{\lambda_h^2\sigma^4_{kh}(M-1)}{h}, \end{aligned}$$

$$\begin{aligned}
\frac{1}{h}P_{h,h}(w_{2,2}^{(k)})(x) &= \mathbb{E}[X_{kh}^2|\mathcal{G}_{(k-1)h}] \\
&= \sigma_{kh}^2
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{h}P_{h,h}(w_{1,2}^{(k)}(x)) &= \mathbb{E}[\frac{X_{kh}(\beta_h+\sigma_{kh}^2(\delta_h+\lambda_h\epsilon_{kh}^2-1)}{\sqrt{h}}|\mathcal{G}_{(k-1)h}] \\
&= \frac{\beta_h+\sigma_{kh}^2(\delta_h+\lambda_h-1)}{\sqrt{h}}\mathbb{E}[X_{kh}|\mathcal{G}_{(k-1)h}] \\
&= 0.
\end{aligned}$$

□

Remark 3.10. In different notation, the same statements would hold without existence of a density function. In that case, the process has not shown to be Markovian, but we could still write the conditional expectation in full.

3.3 Diffusion approximation of GARCH$(1,1)$

As mentioned, Theorem 2.1 in [7] states conditions (denoted by assumptions 2 through 5 in [7]) under which our discrete Markov process converges in distribution if the time step h goes to zero. To check the formal setup as in [7], we notice that the paths of $(W_{t,h},\sigma_{t+h,h}^2)_{t\geq 0}$ are right-continuous with finite left limit is at each $t>0$. Let $D:=D([0,\infty),\mathbb{R}^2)$ be the space of mappings from $[0,\infty)$ into $\mathbb{R}^2$ that are continuous from the right with finite left limit is. Endow D with the Skorohod metric in that it becomes a metric space (see [1]). We can see the GARCH$(1,1)_h$ process as a D-valued random variable. For the Theorem, we need $||\cdot||$ to be the Euclidean norm on $\mathbb{R}^2$.

Theorem 3.11. *Let for each $h>0$ the random variables ϵ_{kh}, with $k\in\mathbb{N}_{\geq 2}$, be i.i.d, independent of $\mathcal{G}_{(k-1)h}$ and $\epsilon_{kh}\in\mathcal{L}^8$. Let $M:=\mathbb{E}[\epsilon_{kh}^4]$. Assume that for all $k\in\mathbb{N}_{\geq 2}$ and h small enough the law $\Delta_{\epsilon_{kh}}$ of ϵ_{kh} (independent of k) is independent of h and has a density with respect to the Lebesgue measure. Also, assume that the limit is*

$$\beta := \lim_{h\downarrow 0}\frac{\beta_h}{h} \tag{3.9}$$

$$\theta := \lim_{h\downarrow 0}\frac{1-\delta_h-\lambda_h}{h} \tag{3.10}$$

$$\alpha^2 := \lim_{h\downarrow 0}\frac{M-1}{h}\lambda_h^2 \tag{3.11}$$

exist, and

$$\limsup_{h\downarrow 0} \frac{\lambda_h^2}{h} < \frac{2\theta}{3(M-1)},$$

with $\theta, \alpha^2 \in \mathbb{R}_{>0}$ *and* $\beta \in \mathbb{R}_{\geq 0}$*. Then, for all* $t \in \mathbb{R}_{\geq 0}$ *we have that* $\lim_{h\downarrow 0}(W_{t,h}, \sigma_{t,h}^2) \stackrel{d}{=} (W_t, \sigma_t^2)$*, where the process* $(W_t, \sigma_t^2)_{t\geq 0}$ *satisfies*

$$\begin{aligned} W_t &= W_0 + \int_0^t \sigma_s^2 \, dB_{1,s}, \\ \sigma_t^2 &= \sigma_0^2 + \int_0^t (\beta - \theta\sigma_s^2) ds + \int_0^t \alpha\sigma_s^2 dB_{2,s}, \end{aligned}$$

and

$$\mathbb{P}[(W_0, \sigma_0^2) \in \Gamma] = \nu(\Gamma) \qquad \textit{for any } \Gamma \in \mathcal{B}(\mathbb{R}^2),$$

where $(B_{1,t})_{t\geq 0}$ *and* $(B_{2,t})_{t\geq 0}$ *are two independent Brownian motions. A weak solution of* (W_t, σ_t^2) *exists and is distributionally unique. Finally,* (W_t, σ_t^2) *remains finite in finite time intervals almost surely, i.e., for all* $T > 0$*,*

$$\mathbb{P}[\sup_{0\leq t\leq T} ||(W_t, \sigma_t^2)|| < \infty] = 1.$$

Proof. The discrete-time processes $\{(W_{kh}, \sigma_{(k+1)h}^2)_{k\in\mathbb{Z}_{\geq 0}}\}_h$ will be our main sequence of interest, because it is Markovian. If the limit result is proved for this sequence, then follows $\lim_{h\downarrow 0}(W_{t,h}, \sigma_{t+h,h}^2) \stackrel{d}{=} \lim_{h\downarrow 0}(W_{t,h}, \sigma_{t,h}^2)$ for all $t \in \mathbb{R}_{\geq 0}$. This is justified by the fact that Theorem 2.1 in [7] shows that the sample paths are continuous with probability 1. This is also seen in our statement. The stochastic integrals given in this Theorem are continuous-time processes, because the Brownian Motions are continuous. Also $\sigma_t^2 \in \mathcal{L}^1$ by Lemma 3.5, so the above Lebesgue-Stieltjes integral is (absolutely) continuous. In other words, once we have proved the limit result for the sequence $\{(W_{kh}, \sigma_{(k+1)h}^2)_{k\in\mathbb{Z}_{\geq 0}}\}_h$ the proof is complete. Let us check the four conditions.

Condition 1: Let $\epsilon^{(3)}, K$ and $\epsilon^{(5)}, N$ be as respectively in Lemma 3.5 and Lemma 3.6. Choose $R \geq 0$ and $T \geq 0$ arbitrary. Fix $0 < h \leq \epsilon^{(5)} \leq \epsilon^{(3)}$, and fix our random starting point $x := (W_{(k-1)h}, \sigma_{kh}^2)$, for a certain $k \in \mathbb{N}$, such that $||x(\omega)|| \leq R$ almost everywhere. Let our time t be given such that $0 \leq t \leq T$. Let l be uniquely given by $(l+1)h \leq t < (l+2)h$. Without loss of generality we may assume $l \in \mathbb{Z}_{\geq 0}$.

The first part of Condition 1 deals with a technical requirement in terms of a fourth moment. It implies that the sample paths of the limit process are continuous with probability one. This is intuitively seen by the fact that

$$1_{(a,\infty)}(||y-x||) \le \frac{1}{a^4}||y-x||^4, \quad a \in \mathbb{R}_{>0}.$$

So we define the functions

$$f_i(y) := |(y-x)_i|^4, \quad i = 1, 2$$

on $\mathbb{R}^2$. We look at

$$\begin{aligned} P_{h,h}(f_1)(x) &= h^2 \mathbb{E}[\epsilon_{kh}^4 \sigma_{kh}^4 | \mathcal{G}_{(k-1)h}], \\ P_{h,h}(f_2)(x) &= \mathbb{E}[\Big(\beta_h - \sigma_{kh}^2(1-\delta_h - \lambda_h \mathbb{E}[\epsilon_{kh}^2])\Big)^4 | \mathcal{G}_{(k-1)h}]. \end{aligned}$$

Because the conditional expectation is defined through integrals, we take expectations. First, we invoke Lemma 3.5 to obtain $\mathbb{E}[P_{h,h}(f_1)(x)] \le h^2 \mathbb{E}[\epsilon_{kh}^4]K$. Observe that $\epsilon_{kh}^4 \sigma_{kh}^4 \ge 0$, so

$$\forall G \in \mathcal{G}_{(k-1)h}: \quad 0 \le \mathbb{E}[1_G \cdot P_{h,h}(f_1)(x)] \le h^2 \mathbb{E}[\epsilon_{kh}^4]K$$

and

$$\lim_{h \downarrow 0} \frac{1}{h} P_{h,h}(f_1)(x) \le \lim_{h \downarrow 0} h \mathbb{E}[\epsilon_{kh}^4]K = 0$$

Second, for simplicity we define

$$B(m) := (1-\delta_h - \lambda_h \epsilon_{kh}^2)^m, \quad m \in \mathbb{N}.$$

with

$$\begin{aligned} \mathbb{E}|B(m)| &= \sum_{k=0}^{m} \binom{m}{k} (1-\delta_h)^k \lambda_h^{m-k} \mathbb{E}[\epsilon_{kh}^{2(m-k)}] \\ &= O(h^{m\frac{1}{2}}), \end{aligned}$$

as $h \downarrow 0$. We investigate

$$\begin{aligned} \mathbb{E}[P_{h,h}(f_2)(x)] &= \mathbb{E}[\beta_h^4 - 4\beta_h^3 \sigma_{kh}^2 B(1) + 6\beta_h^2 \sigma_{kh}^4 B(2) \\ &\quad -4\beta_h \sigma_{kh}^6 B(3) + \sigma_{kh}^8 B(4)] \\ &\le \mathbb{E}[\beta_h^4 + 6\beta_h^2 \sigma_{kh}^4 B(2) + \sigma_{kh}^8 B(4)] \end{aligned}$$

We may assume that for all $h \le \epsilon^{(5)}$ that $\mathbb{E}[B(4)] \le D$ for a certain $D \in \mathbb{R}_{\ge 0}$. We invoke Lemmas 3.5 and 3.6 to obtain

$$\mathbb{E}[P_{h,h}(f_2)(x)] \le \mathbb{E}[\beta_h^4 + 6\beta_h^2 B(2)K + NB(4)].$$

Without loss of generality we have for all $h \leq \epsilon^{(5)}$ that $\mathbb{E}[P_{h,h}(f_2)(x)] \leq h^2 M$ for a certain $M \in \mathbb{R}_{\geq 0}$. Because $\Big(\beta_h - \sigma_{kh}^2(1 - \delta_h - \lambda_h \mathbb{E}[\epsilon_{kh}^2])\Big)^4 \geq 0$ follows

$$\forall G \in \mathcal{G}_{(k-1)h} : \quad 0 \leq \mathbb{E}[1_G \cdot P_{h,h}(f_2(x))] \leq h^2 M,$$

and

$$\lim_{h \downarrow 0} \frac{1}{h} P_{h,h}(f_2)(x) \quad \leq \quad \lim_{h \downarrow 0} hM = 0.$$

Note that estimates for $\mathbb{E}[P_{h,h}(f_i(x))]$, $i = 1, 2$, were independent of t and holds for all $||x|| \leq R$ by our technical Lemmas. So the speed of convergence to 0 was in both cases independent of t and x. Hence,

$$\lim_{h \downarrow 0} \sup_{||x|| \leq R, 0 \leq t \leq T} P_{t,h}(f_i(x)) = 0, \quad i = 1, 2.$$

Condition 1 also deals with the drift and the second moment per unit of time. It requires that the drift and second moment per unit of time converges uniformly on compact sets to well-behaved functions (of time t and state x). For $t \in \mathbb{R}_{\geq 0}$ we define the drift vector

$$b\big([w(\omega, t), s(\omega, t)], t\big) \quad := \quad \begin{pmatrix} 0 & \beta - \theta s(\omega, t) \end{pmatrix}$$

and diffusion matrix

$$a\big([w(\omega, t), s(\omega, t)], t\big) \quad := \quad \begin{pmatrix} s(\omega, t) & 0 \\ 0 & \alpha^2 s(\omega, t)^2 \end{pmatrix},$$

where w and s are measurable functions

$$\begin{aligned} w : \Omega \times \mathbb{R}_{\geq 0} &\to \mathbb{R} \\ (\omega, t) &\mapsto w(\omega, t) \end{aligned}$$

and

$$\begin{aligned} s : \Omega \times \mathbb{R}_{\geq 0} &\to \mathbb{R}_{\geq 0} \\ (\omega, t) &\mapsto s(\omega, t). \end{aligned}$$

Let the functions $g^{(k)}(y)$ and $w^{(k)}(y)$ be as in Lemma 3.9. Use (3.9), (3.10), (3.11) and Lemma 3.5 to obtain

$$
\begin{aligned}
\lim_{h\downarrow 0}\frac{1}{h}P_{t,h}(g_1^{(k)}(x)) &= 0\\
\lim_{h\downarrow 0}\frac{1}{h}P_{t,h}(g_2^{(k)}(x)) &= \beta-\theta\lim_{h\downarrow 0}P_{h,h}(P_{h,h}(\cdots(\sigma_{kh}^2)\cdots))\\
\lim_{h\downarrow 0}\frac{1}{h}P_{t,h}(w_{1,1}^{(k)}(x)) &= \lim_{h\downarrow 0}P_{h,h}(P_{h,h}(\cdots(\sigma_{kh}^2)\cdots))\\
\lim_{h\downarrow 0}\frac{1}{h}P_{t,h}(w_{2,2}^{(k)}(x)) &= \alpha^2\lim_{h\downarrow 0}P_{h,h}(P_{h,h}(\cdots(P_{h,h}(\sigma_{kh}^4)\cdots))\\
\lim_{h\downarrow 0}\frac{1}{h}P_{t,h}(w_{1,2}^{(k)}(x)) &= \lim_{h\downarrow 0}\frac{1}{h}P_{t,h}(w_{2,1}^{(k)}(x))=0.
\end{aligned}
$$

Note that speed of convergence in the limit is (3.9), (3.10), (3.11) is independent of t and x and holds for all $||x||\le R$. Hence,

$$
\begin{aligned}
\lim_{h\downarrow 0}\sup_{||x||\le R,0\le t\le T}||P_{t,h}(g(x))-b([W_t,\sigma_t^2],t)|| &= 0\\
\lim_{h\downarrow 0}\sup_{||x||\le R,0\le t\le T}||P_{t,h}(w(x))-a([W_t,\sigma_t^2],t)|| &= 0.
\end{aligned}
$$

Condition 2: This condition requires that the diffusion matrix a has a well-behaved matrix square root r. We define the matrix

$$
r\big([w(\omega,t),s(\omega,t)],t\big) \;:=\; \begin{pmatrix}\sqrt{s(\omega,t)} & 0\\ 0 & \alpha s(\omega,t)\end{pmatrix},
$$

where w and s are as before in the proof. Obviously, for all $[w(\omega,t)s(\omega,t)],t$ holds

$$
a\big([w(\omega,t),s(\omega,t)],t\big)=r\big([w(\omega,t)s(\omega,t)],t\big)r\big([w(\omega,t)s(\omega,t)],t\big)^T.
$$

The function r is measurable, because w and s are. Note that it is continuous as function from w and s.

Condition 3: The third condition concerns the behavior of the initial distribution of our discrete time process $\{(W_{kh},\sigma^2_{(k+1)h})_{k\in\mathbb{Z}_{\ge 0}}\}_h$, when taking the limit. This is not a concern, because Assumption 3.3 tells us that the initial probability law is given by ν for every $h>0$.

Condition 4: So far Theorem 2.1 in [7] suggests a limit diffusion of the form

$$\begin{aligned} dW_t &= \sigma_t dB_{1,t}, \\ d\sigma_t^2 &= (\beta - \theta\sigma_t^2)dt + \alpha\sigma_t^2 dB_{2,t}, \\ \mathbb{P}[(W_0, \sigma_0^2) \in \Gamma] &= \nu(\Gamma) \quad \text{for any } \Gamma \in \mathcal{B}(\mathbb{R}^2). \end{aligned}$$

At this point, there are two things that can go wrong. First, ν, a and b may not uniquely define a limit process. Second, a limit process may not exist, because when taking together ν, a and b may imply that the process explodes with strict positive probability to infinity in finite time. In [7] one can find conditions which are sufficient to exclude these possibilities.

It helps to define

$$V_t := \log(\sigma_t^2), \quad t > 0.$$

For our candidate limit diffusion we rewrite

$$\begin{aligned} dW_t &= \exp(\frac{V_t}{2})dB_{1,t}, \\ \mathbb{P}[(W_0, \exp(V_0)) \in \Gamma] &= \nu(\Gamma) \quad \text{for any } \Gamma \in \mathcal{B}(\mathbb{R}^2). \end{aligned}$$

and an application of Itô's Lemma gives

$$\begin{aligned} dV_t &= \alpha dB_{2,t} + (\frac{\beta}{\sigma_s^2} - \theta)dt - \frac{1}{2}d(\alpha^2 t) \\ &= (\beta\exp(-V_t) - \theta - \frac{\alpha^2}{2})dt + \alpha dB_{2,t}. \end{aligned}$$

So we define a new drift vector b' and diffusion matrix a' by

$$b'\big([w(\omega,t), v(\omega,t)], t\big) := \begin{pmatrix} 0 & \beta\exp(-v(\omega,t)) - (\theta + \frac{\alpha^2}{2}) \end{pmatrix}$$

and diffusion matrix

$$a'\big([w(\omega,t), v(\omega,t)], t\big) := \begin{pmatrix} \exp(\frac{v(\omega,t)}{2}) & 0 \\ 0 & \alpha \end{pmatrix},$$

where w is as in condition 1 and 2, and v is a measurable function from $\Omega \times \mathbb{R}_{\geq 0}$ to $\mathbb{R}_{\geq 0}$.

Recall that a (symmetric) matrix is positive definite if all eigenvalues of the matrix are positive. Note that the eigenvalues of a' (as function from w and v) are given by $\exp(\frac{v(\omega,t)}{2}) > 0$ and $\alpha > 0$. Hence, condition B (in the Appendix of [7]) for distributional uniqueness holds.

Next, we check the non-explosiveness condition. Take the nonnegative function

$$\varphi[(w,v),t] = K + f(w)|w| + f(v)\exp(|v|),$$

where $K \in \mathbb{R}_{>0}$ and

$$f(x) = \begin{cases} \exp(\frac{-1}{|x|}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

One can check that we have the following identities

$$\begin{aligned}
\frac{\partial^2 \varphi}{\partial w^2} &= \begin{cases} \exp(\frac{-1}{w}) \cdot \frac{1}{w^3} & \text{if } w > 0 \\ \exp(\frac{1}{w}) \cdot \frac{-1}{w^3} & \text{if } w < 0, \end{cases} \\
\frac{\partial \varphi}{\partial v} &= \begin{cases} \exp(v - \frac{1}{v}) \cdot (\frac{1}{v^2} + 1) & \text{if } v > 0 \\ \exp(\frac{1}{v} - v) \cdot (-\frac{1}{v^2} - 1) & \text{if } v < 0, \end{cases} \\
\frac{\partial^2 \varphi}{\partial v^2} &= \begin{cases} \exp(v - \frac{1}{v}) \cdot (1 + \frac{2}{v^2} - \frac{2}{v^3} + \frac{1}{v^4}) & \text{if } v > 0 \\ \exp(\frac{1}{v} - v) \cdot (1 + \frac{2}{v^2} + \frac{2}{v^3} + \frac{1}{v^4}) & \text{if } v < 0. \end{cases}
\end{aligned}$$

If we use the definition of the derivative (the difference quotient), then one also obtains that φ is twice differentiable in zero. Note that

$$\lim_{||(w,v)|| \to \infty} \inf_{0 \le t \le T} \varphi((w,v),t) = \infty, \quad T > 0,$$

so φ is a *Liapunov function* (see [7]). There exists a $R > 0$ such that for $|w| \geq R$ and $|v| \geq R$ we have

$$\begin{aligned}
\frac{\partial^2 \varphi}{\partial w^2} &\leq C_1 \\
\frac{\partial \varphi}{\partial v} &\leq \text{sign}(v) \cdot \exp(|v|) + C_2 \\
\frac{\partial^2 \varphi}{\partial v^2} &\leq \exp(|v|) + C_3,
\end{aligned}$$

for certain constants $C_1, C_2, C_3 \in \mathbb{R}_{\geq 0}$, and where sign is the "sign" function. There holds for $|w|, |v| \geq R$

$$\begin{aligned}
& \frac{1}{2}\exp(\frac{v}{2})\frac{\partial^2 \varphi}{\partial w^2} + [\beta\exp(-v) - (\theta + \frac{\alpha^2}{2})]\frac{\partial \varphi}{\partial v} + \frac{1}{2}\alpha\frac{\partial^2 \varphi}{\partial v^2} \\
\leq\ & [\beta\exp(-v) - (\theta + \frac{\alpha^2}{2})]\text{sign}(v) \cdot \exp(|v|) + C_2\beta\exp(-v) + \frac{1}{2}\alpha\exp(|v|) \\
& + \frac{1}{2}\exp(\frac{v}{2})C_1 - C_2(\theta + \frac{\alpha^2}{2}) + \frac{1}{2}\alpha C_3 \qquad (3.12)
\end{aligned}$$

and

$$D \cdot \varphi \;\; = \;\; DK + D\exp(\frac{-1}{|w|})|w| + D\exp(\frac{-1}{|v|})\exp(|v|), \quad D \in \mathbb{R}_{\geq 0}.$$

One sees that we can pick constants $D, K \in \mathbb{R}_{>0}$ independent of t such that

$$\text{RHS of (3.12)} \leq D \cdot \varphi$$

for all (w, v). By Assumption 3.3, the continous Mapping Theorem (see [1]) gives distributional uniqueness and non-explosiveness for our candidate limit diffusion. □

We have seen that a weak solution of the stochastic differential equation exists. Only, we must check that the assumptions on the parameters are not conflicting. So that it is not an empty statement. The next example gives parameters that obey these conditions.

Example 3.12. For each $h > 0$ let $(\epsilon_{kh})_{k\in\mathbb{N}}$ be a sequence of independent standard normal distributed random variables, with $(\mathcal{G}_{kh})$ the generated filtration. Set the nonnegative constants, dependent of $h > 0$, as follows

$$\begin{aligned} \beta_h &= \beta h \\ \delta_h &= 1 - \lambda(h/2)^{1/2} - \theta h \\ \lambda_h &= \lambda(h/2)^{1/2}, \end{aligned}$$

with $\theta \in \mathbb{R}_{>0}$, $\beta > 0$ and $0 \leq \lambda < \sqrt{\theta}$.

4 Continuous-time model 2: Noise variables replaced by increments of a Lévy process

4.1 Motivation for using a Lévy process

We have seen that, under appropriate conditions, the linear GARCH(1,1) model converges in distribution to a bivariate diffusion process. This was given in terms of two independent Brownian motions. So if we condition on the past, then the variance of the displacement, over a small time-interval, made by σ_t^2 is independent of W_t. In practice, you want some dependency between the "direction" X_t (determined by W_t) and the conditional variance σ_t^2. For instance, in periods of high volatility we maybe want a higher probability of going downwards than upwards based on experiences. So we want to loosen this independence property. This is where a Lévy process comes in place. It will be used as only source of randomness instead of two independent Brownian Motions. Namely, the noise variables ϵ_j are replaced by increments of a Lévy process. This construction comes from [6].

To do this, we first look closer at our discrete-time model. One has for $n \in \mathbb{N}$

$$\begin{aligned}
\sigma_n^2 &= \beta + \delta\sigma_{n-1}^2 + \lambda X_{n-1}^2 \\
&= \beta + (\delta + \lambda\epsilon_{n-1}^2)\sigma_{n-1}^2 \\
&= \beta + (\delta + \lambda\epsilon_{n-1}^2)(\beta + [\delta + \lambda\epsilon_{n-2}^2]\sigma_{n-2}^2) \\
&= \beta(1 + \delta + \lambda\epsilon_{n-1}^2) + (\delta + \lambda\epsilon_{n-1}^2)(\delta + \lambda\xi_{n-2}^2)(\beta + [\delta + \lambda\epsilon_{n-3}^2]\sigma_{n-3}^2) \\
&\vdots \\
&= \beta \sum_{i=0}^{n-1} \Big(\prod_{j=i+1}^{n-1} (\delta + \lambda\epsilon_j^2) \Big) + \sigma_0^2 \prod_{j=0}^{n-1} (\delta + \lambda\epsilon_j^2), \qquad (4.1)
\end{aligned}$$

where $\prod_n^{n-1}(\delta + \lambda\epsilon_n^2) := 1$. Of major importance are conditions under which the model converges in distribution to finite random variables (i.e. has a finite stable distribution). Namely, this result will be used to motivate our continuous-time model. For this, the last term in (4.1) plays a significant role. Similarly as in [6] we have the following Theorem.

Theorem 4.1. *Let all ϵ_n be i.i.d. and independent of $\mathcal{F}_{n-1}$. If $\lim_{n\to\infty} \prod_{i=1}^n (\delta + \lambda\epsilon_i^2) \overset{a.s.}{=} 0$, then we have $\lim_{n\to\infty} \sigma_n^2 \overset{d}{=} \sigma^2$ and $\lim_{n\to\infty} X_n \overset{d}{=} X$ for finite random variables σ^2 and X. Also, $\sigma^2 \overset{d}{=} \beta + (\delta + \lambda\epsilon_1^2)\sigma^2$ with σ^2 independent of ϵ_1. Conversely, if $\lim_{n\to\infty} \prod_{i=1}^n (\delta + \lambda\epsilon_i^2) \overset{a.s.}{\neq} 0$, then $\sigma_n^2 \overset{\mathbb{P}}{\to} \infty$ and $|X_n| \overset{\mathbb{P}}{\to} \infty$ as $n \to \infty$.*

Proof. In the notation of [3], pick $M_1 = 1$, $M_j = (\delta + \lambda\epsilon_{j-1}^2)$ for $j = 2, 3, \dots$ and $Q_i = 1$ for $i = 1, 2 \dots$ and $R_0 = \sigma_0^2(\delta + \lambda\epsilon_0^2)/\beta$. Then we have (according to page 1196 in [3])

$$\begin{aligned}
\frac{\sigma_{n+1}^2}{\beta} := R_n(R_0) &= \sum_{i=1}^{n} Q_i \prod_{j=i+1}^{n} M_j + R_0 \prod_{j=1}^{n} M_j \\
&= \sum_{i=1}^{n} \prod_{j=i+1}^{n} (\delta + \lambda\epsilon_{j-1}^2) + \frac{\sigma_0^2(\delta + \lambda\epsilon_0^2)}{\beta} \prod_{j=2}^{n} (\delta + \lambda\epsilon_{j-1}^2) \\
&= \sum_{i=0}^{n-1} \prod_{j=i+1}^{n-1} (\delta + \lambda\epsilon_j^2) + \frac{\sigma_0^2}{\beta} \prod_{j=0}^{n-1} (\delta + \lambda\epsilon_j^2)
\end{aligned}$$

Similarly as in the article we define

$$\pi_n := \begin{cases} \prod_{j=1}^{n} M_j, & n = 1, 2, \dots, \\ 1, & n = 0, \end{cases}$$

so that

$$\begin{aligned}
Z_\infty &:= \sum_{k=1}^{\infty} \pi_{k-1} Q_k \\
&= \sum_{k=1}^{\infty} \pi_{k-1} \\
&= \sum_{k=0}^{\infty} \prod_{j=1}^{k} (\delta + \lambda\epsilon_j^2).
\end{aligned}$$

Assumption 2.2 develops that $-\log(M_i)$ is finite for all $i \in \mathbb{N}$. Now it is justified to use an application of Theorem 2.1 of [3], with $\lim_{n\to\infty} \pi_n \overset{a.s.}{=} 0$ and $Q_i = 1$ for all i, to conclude

$$\frac{\sigma_{n+1}^2}{\beta} \overset{a.s.}{\to} Z_\infty \text{ as } n \to \infty,$$

where Z_∞ is absolutely convergent. All the ϵ_n's are have the same distributon, so we derive $\sigma_{n+1}^2 \overset{a.s.}{\to} \sigma^2 \overset{d}{=} \beta + (\delta + \lambda\epsilon_1^2)\sigma^2$ as $n \to \infty$. By absolute convergence follows that σ^2 is a finite random variable, and independence of ϵ_n gives that σ^2 is independent of ϵ_1^2. This gives $X_n \overset{d}{\to} X$ as $n \to \infty$, with $X \overset{d}{=} \sigma\epsilon_1$ a finite random variable. If $\lim_{n\to\infty} \prod_{i=1}^{n} (\delta + \lambda\epsilon_i^2) \overset{a.s.}{\neq} 0$, then Theorem 2.1 of [3] shows that $\sigma_n \overset{\mathbb{P}}{\to} \infty$, and then $|X_n| \overset{\mathbb{P}}{\to} \infty$ as $n \to \infty$. $\square$

Remark 4.2. In [6] there are necessary and sufficient conditions given under which $\lim_{n\to\infty}\prod_{i=1}^{n}(\delta+\lambda\epsilon_i^2)\overset{a.s.}{=}0$ holds. Keeping Assumption 2.2 in mind, these conditions are as follows

(i) If $\delta > 0$ and $\lambda \geq 0$, then there must hold

$$\mathbb{E}[|\log(\delta+\lambda\epsilon_1^2)|] < \infty \quad \text{and} \quad \mathbb{E}\log(\delta+\lambda\epsilon_1^2) < 0. \tag{4.2}$$

(ii) If $\delta = 0$ and $\lambda > 0$, then either (4.2) or $\mathbb{E}[(\log(\lambda\epsilon_1^2)^-] = \infty$ in combination with

$$\int_0^\infty x\Big(\int_0^x \mathbb{P}\{\log(\lambda\epsilon_1^2) < y\}dy\Big)^{-1} d\mathbb{P}\{\log(\lambda\epsilon_1^2) \leq x\} < \infty$$

must hold.

Condition (i) is also known in terms of the *top Lyapounov exponent* (see [10]).

Theorem 4.1 motivates us to take a closer look at $\beta\cdot Z_\infty$. Note that taking sums is a special type of integration. Namely,

$$\begin{aligned}\beta\sum_{i=0}^{n-1}\Big(\prod_{j=i+1}^{n-1}(\delta+\lambda\epsilon_j^2)\Big) &= \beta\int_0^n \prod_{j=\lfloor s\rfloor+1}^{n-1}(\delta+\lambda\epsilon_j^2)\,ds \\ &= \beta\int_0^n e^{\sum_{j=\lfloor s\rfloor+1}^{n-1}\log(\delta+\lambda\epsilon_j^2)}\,ds,\end{aligned} \tag{4.3}$$

where both integrations are with respect to the Lebesgue measure on $\mathbb{R}_{\geq 0}$ (as before $\lfloor\cdot\rfloor$ denotes the integer part). This suggests replacing the noise variables ϵ_j by increments of a Lévy process. Because, they are strictly stationary and independent.

We proceed from the representation (4.3). Note that for $0 \leq s < 1$ there holds

$$\begin{aligned}\sum_{j=\lfloor s\rfloor+1}^{n-1}\log(\delta+\lambda\epsilon_j^2) &= (n-1)(\log(\delta)-\log(\delta)) + \sum_{j=\lfloor s\rfloor+1}^{n-1}\log(\delta+\lambda\epsilon_j^2) \\ &= (n-1)\log(\delta) + \sum_{j=\lfloor s\rfloor+1}^{n-1}\log(1+\frac{\lambda}{\delta}\epsilon_j^2).\end{aligned} \tag{4.4}$$

Henceforth, to avoid dividing by zero we will, on top of Assumption 2.2 and keeping Remark 4.2 in mind, assume the following.

Assumption 4.3. *There holds* $0 < \delta < 1$.

4.2 GARCH(1,1) process driven by a Lévy process

In this section we will derive a continuous-time GARCH(1,1) process. As in our original discrete setup, a single source of randomness suffices. The jumps of a (any) Lévy process shall replace the noise variables.

Let $(L_t)_{t\geq 0}$ be a (càdlàg) Lévy process on $\mathbb{R}$ with jumps $\Delta L_t := L_t - L_{t^-}$, a filtration $(\mathcal{F}_t)$ which satisfies the "usual conditions", and ν^L as it is Lévy measure. For future reasons we will assume the following.

Assumption 4.4. *The Lévy process L does not jump at the starting point, i.e. $\Delta L_0 = 0$.*

We recall some of it is properties (see [13]). There holds

$$\int_{\mathbb{R}\setminus\{0\}} (x^2 \wedge 1)\nu^L(dx) < \infty, \tag{4.5}$$

by the Lévy-Itô decomposition. Since L is càdlàg,

$$\sum_{\substack{0\leq s\leq t,\\ |\Delta L_s|\geq\epsilon}} \Delta L_s$$

is a finite sum for all $\epsilon > 0$, and the set

$$\{t \geq 0 : \Delta L_t \neq 0\}$$

is at most countable. The random measure μ^L associated with the jumps ΔL_t is a Poisson random measure on $\mathbb{R}_{\geq 0} \times \mathbb{R} \setminus \{0\}$ with intensity measure $\text{Leb} \otimes \nu^L$. A measurable function f is called μ^L-integrable if for for every $t \geq 0$,

$$\int_{(0,t]} \int_{\mathbb{R}\setminus\{0\}} |f(x)|\, \mu^L(ds, dx) = \sum_{s\leq t} |f(\Delta L_s)| \cdot 1_{\{\Delta L_s \neq 0\}} < \infty,$$

almost surely. It is a fact that a measurable function f is μ^L-integrable if and only if $\int_{\mathbb{R}\setminus\{0\}} (|f| \wedge 1)\, d\nu^L < \infty$.

With (4.3) and (4.4) in mind, we define a càdlàg process $(X_t)_{t\geq 0}$ by

$$-X_t := t\log(\delta) + \sum_{0\leq s\leq t} \log(1 + \frac{\lambda}{\delta}(\Delta L_s)^2).$$

Note that

$$-X_t = t\log(\delta) + \sum_{0\le s\le t} \log(1+\frac{\lambda}{\delta}(\Delta L_s)^2)\cdot 1_{\{\Delta L_s\neq 0\}} \tag{4.6}$$

holds, so it is in fact a countable sum. This will be used to define our continuous-time volatility process. We see that only $\delta > 0$ is allowed. Thus our continuous-time GARCH does not contain a continuous-time ARCH as a submodel. Suppose we want to accommodate the case $\delta = 0$, then we have to go back to (4.3) and X_t should be taken as

$$-X_t = -\sum_{0<s\le t} \log(\lambda(\Delta L_s)^2)1_{\{\Delta L_s\neq 0\}}, \quad t\ge 0.$$

Note that this is only a well-defined (Lévy) process, if L is compound Poisson.

Let us state some facts of X_t (result from [6]).

Proposition 4.5. *The process $(X_t)_{t\ge 0}$ is a Lévy process of bounded variation with drift $b = -\log\delta$, Gaussian component $a = 0$ and Lévy measure ν^X given by*

$$\nu^X((0,\infty)) = 0$$

and

$$\nu^X((-\infty,-x]) = \nu^L(\{y\in\mathbb{R} : |y| \ge \sqrt{(e^x-1)\delta/\lambda}\}) \quad \textit{for } x > 0.$$

Proof. Observe that the process $\sum_{0\le s\le t}\log(1+\frac{\lambda}{\delta}(\Delta L_s)^2)$ is of bounded variation, because it is increasing in t. The process X_t inherit is the Levy property of L_t. That it only makes negative jumps is clear. Use properties of a Poisson process to obtain for $x > 0$

$$\begin{aligned}
\nu^X(-\infty,-x] &= \mathbb{E}[\sum_{0<s\le 1} 1_{\{-\log(1+\frac{\lambda}{\delta}(\Delta L_s)^2\le -x\}}] \\
&= \mathbb{E}[\sum_{0<s\le 1} 1_{\{|\Delta L_s|\ge\sqrt{(e^x-1)\frac{\delta}{\lambda}}\}}] \\
&= \nu^L(\{y : |y| \ge \sqrt{(e^x-1)\frac{\delta}{\lambda}}\}).
\end{aligned}$$

Particularly, there holds

$$\int_{\mathbb{R}\setminus\{0\}} |x|\wedge 1\ \nu^X(dx) \;=\; \int_{\{|y|\le\sqrt{(e-1)\frac{\delta}{\lambda}}\}} \log(1+\frac{\lambda}{\delta}y^2)\ \nu^L(dy) < \infty,$$

by (4.5). The Lévy-Itô decomposition gives that $(X_t)_{t\ge 0}$ is a Lévy process of bounded variation with drift $b = -\log\delta$ and Gaussian component $a = 0$. $\square$

Remark 4.6. The fact that $\int_{\mathbb{R}\setminus\{0\}} |x| \wedge 1 \, \nu^X(dx) < \infty$ shows that the countable sum $\sum_{0 \le s \le t} \log(1 + \frac{\lambda}{\delta}(\Delta L_s)^2)$ is absolutely convergent.

Remark 4.7. It is worth mentioning that X_t is a càdlàg semimartingale, because any Lévy process is a semimartingale (see [8]).

Analogously to (4.1) we define the following continuous-time volatility process.

Definition 4.8. Let $\beta > 0$ and let σ_0 be a finite random variable independent of $(L_t)_{t \ge 0}$. A left-continuous with finite right limit is (called càglàd) volatility process $(\sigma_t^2)_{t \ge 0}$ is given by

$$\sigma_t^2 \;:=\; \beta \int_0^t e^{-X_{t-}+X_s} ds + e^{-X_{t-}} \sigma_0^2, \quad t \ge 0.$$

Note that the Lebesgue-Stieltjes integral in the definition is well-defined, because X_t is a process of bounded variation as stated in Proposition 4.5.

For $s \le t$ we look at

$$-X_{t-} + X_s \;=\; (t-s)\log(\delta) + \sum_{s<u\le t} \log(1 + \frac{\lambda}{\delta}(\Delta L_u)^2),$$

so σ_t^2 can be written as

$$\begin{aligned} \sigma_t^2 &= \beta\Big(\int_0^t e^{(t-s)\log(\delta)+\sum_{s<u\le t}\log(1+\frac{\lambda}{\delta}(\Delta L_u)^2)} du\Big) + e^{-X_{t-}} \sigma_0^2 \\ &= \beta \int_0^t \delta^{t-s} \prod_{s<u\le t} (1 + \frac{\lambda}{\delta}(\Delta L_u)^2) du + e^{-X_{t-}} \sigma_0^2. \end{aligned}$$

Here, one sees the resemblance with (4.3) when it is combined with (4.4).

We want to define a continuous-time process (G_t) similar to the martingale difference sequence (X_n) in the discrete case.

Definition 4.9. We define a COGARCH process $(G_t)_{t \ge 0}$ as the càdlàg process satisfying the stochastic differential equation

$$G_t = \int_0^t \sigma_s \, dL_s, \quad t \ge 0, \quad G_0 = 0.$$

This is well-defined by [8], because from Remark 4.7 it follows that σ_t is a càglàd semimartingale. Observe that G_t only jumps if L_t does, so $\Delta G_t = \sigma_t \Delta L_t$. Recalling from the discrete case that $X_n = \sigma_n \epsilon_n$, one might suggest that for small $h > 0$ the process $G_{t+h} - G_t$ will in some sense take the place of X_n.

So fare we have only defined the continuous-time volatility process in combination with the COGARCH process. This was based on intuitive reasons. The question that arises is: "Is this definition the right choice based on hard mathematical reasons"? As in the discrete case we want some kind of regressive- and feedback-relation for the volatility process. Moreover, the conditional variance of the difference COGARCH process must in some sense be equal to σ_t. We will observe all this in the next chapter together with some further results from [6].

5 Behaviour of the continuous-time model driven by a Lévy process

5.1 The volatility process

In Chapter 4 we defined stochastic differential equations defining process G_t and σ_t. This Chapter will investigate these processes, so that it can truly can be called a continuous-time GARCH process. We will research if the distinguish features in the discrete case are also present in G_t and σ_t. Therefore, we have to derive a stochastic differential equation for σ_t^2 (result from [6]). We shall need the following.

Lemma 5.1. *There holds*

$$e^{-X_t} = 1 + \log\delta \int_0^t e^{-X_u} du + \frac{\lambda}{\delta} \sum_{0<s\le t} e^{-X_{s-}} (\Delta L_s)^2.$$

Proof. We want to use Itô's famous formula (see [4]). Therefore

$$\begin{aligned} e^{-X_t} &= e^{t\log\delta} \prod_{0<s\le t} (1 + \frac{\lambda}{\delta}(\Delta L_s)^2) \\ &= e^{K_t} S_t, \end{aligned}$$

where $K_t := t\log\delta$ and $S_t := \prod_{0<s\le t}(1 + \frac{\lambda}{\delta}(\Delta L_s)^2)$ for $t \ge 0$. For the application of the formula we define the function $f(k,s) := e^k s$, which is infinite continuously differentiable in all it is arguments. Hence, Itô's formula for non-continuous semimartingales develops

$$\begin{aligned} e^{-X_t} &= f(K_t, S_t) \\ &= e^{-X_0} + \int_0^t e^{K_u} S_u \, dK_u + \int_0^t e^{K_u} \, dS_u \\ &\quad + \sum_{s\le t} \Big(e^{K_s} S_s - e^{K_{s-}} S_{s-} - K_{s-} e^{K_{s-}} S_{s-} \Delta K_s - e^{K_{s-}} \Delta S_s \Big) \\ &= 1 + \log\delta \int_0^t e^{-X_u} du + \int_0^t e^{u\log\delta} \, d(\prod_{0<s\le u} (1 + \frac{\lambda}{\delta}(\Delta L_s)^2) \\ &\quad + \sum_{s\le t} \Big(e^{K_s} \Delta S_s - e^{K_s} \Delta S_s \Big) \\ &= 1 + \log\delta \int_0^t e^{-X_u} du + \int_0^t e^{u\log\delta} \, d(\prod_{0<s\le u} (1 + \frac{\lambda}{\delta}(\Delta L_s)^2), \quad (5.1) \end{aligned}$$

where we used that K_t is continuous. First we restrict ourselves to jumps only bigger than some $\epsilon > 0$. So we replace S_u by $S_u^{(\epsilon)} := \prod_{\substack{0<s\leq u,\\ \Delta L_s>\epsilon}}(1+\frac{\lambda}{\delta}(\Delta L_s)^2)$. Then, there are only a finite number of jumps (bigger than ϵ), say n, on the interval $(0,t]$. We denote these jump times by $t_1, t_2, \ldots, t_n$. Observe that, in this case, we have for the integrator in latter integral that the function value is given by

$$\begin{aligned}
\text{On } (0, \frac{t_1}{2}] &: \quad 1 \\
\text{On } (\frac{t_1}{2}, t_1] &: \quad (1+\frac{\lambda}{\delta}\Delta L_{t_1}^2) \\
\text{On } ((t_1, t_2] &: \quad (1+\frac{\lambda}{\delta}\Delta L_{t_1}^2)(1+\frac{\lambda}{\delta}\Delta L_{t_2}^2) \\
\vdots & \qquad\quad \vdots \\
\text{On } ((t_{n-1}, t_n] &: \quad \prod_{k=1}^{n}(1+\frac{\lambda}{\delta}\Delta L_{t_k^-}^2)(1+\frac{\lambda}{\delta}\Delta L_{t_n}^2).
\end{aligned}$$

Let μ be the Lebesgue-Stieltjes measure associated with $S_u^{(\epsilon)}$, and we define the funtion $g_n(u) := \sum_{k=1}^n 1_{(t_{k-1},t_k]}(u)\cdot e^{t_k \log\delta} + 1_{[0,t_0]}(u)\cdot e^{t_k\log\delta}$ with $t_0 = \frac{t_1}{2}$. Thus,

$$\begin{aligned}
\int_0^t e^{K_u}\, dS_u^{(\epsilon)} &= \int_{\mathbb{R}_{\geq 0}} g_n \, d\mu \\
&= e^{t_1^- \log\delta}(1+\frac{\lambda}{\delta}\Delta L_{t_1^-}^2)(\frac{\lambda}{\delta}\Delta L_{t_2^-}^2) + \cdots \\
&\quad + e^{t_n^- \log\delta}\prod_{k=1}^{n}(1+\frac{\lambda}{\delta}\Delta L_{t_k^-}^2)\frac{\lambda}{\delta}\Delta L_{t_n}^2 \\
&= \frac{\lambda}{\delta}\sum_{\substack{0<s\leq t,\\ \Delta L_s>\epsilon}} e^{-X_{s^-}}(\Delta L_s)^2.
\end{aligned}$$

Therefore, (5.1) and (4.6) tell us

$$\begin{aligned}
e^{-X_t} &= 1 + \log\delta\int_0^t e^{-X_u}du + \lim_{\epsilon\downarrow 0}\frac{\lambda}{\delta}\sum_{\substack{0<s\leq t,\\ \Delta L_s>\epsilon}} e^{-X_{s^-}}(\Delta L_s)^2 \\
&= 1 + \log\delta\int_0^t e^{-X_u}du + \frac{\lambda}{\delta}\sum_{0<s\leq t} e^{-X_{s^-}}(\Delta L_s)^2.
\end{aligned}$$

□

The previous lemma will be of use in our next theorem. We are going to denote $[X,Y]_t$ as the covariation of two semimartingales (perhaps non-continuous) X_t and Y_t.

Theorem 5.2. *The process $(\sigma_t^2)_{t\geq 0}$ satisfies the stochastic differential equation*

$$\sigma_t^2 = \beta t + \log\delta \int_0^t \sigma_s^2 ds + \frac{\lambda}{\delta} \sum_{0<s<t} \sigma_s^2 (\Delta L_s)^2 + \sigma_0^2, \quad t \geq 0$$

Proof. We define $V_t = e^{-X_t}$ and $W_t = \int_0^t e^{X_s} ds$ for $t > 0$. Integration by parts gives

$$\begin{aligned}
& V_t W_t \\
= & \int_{0^+}^t V_{s^-} dW_s + \int_{0^+}^t W_{s^-} dV_s + [V_\cdot, W_\cdot]_t \\
= & \int_{0^+}^t e^{X_{s^-}} d\left(\int_0^s e^{X_y} dy\right) + \int_{0^+}^t \left(\int_0^{s^-} e^{X_u}\, du\right) d(e^{-X_s}) + [e^{-X_\cdot}, \int_0^\cdot e^{X_s} ds]_t \\
= & \int_{0^+}^t e^{X_{s^-}} d\left(\int_0^s e^{X_y} dy\right) + \int_{0^+}^t \left(\int_0^{s} e^{X_u}\, du\right) d(e^{-X_s}) + [e^{-X_\cdot}, \int_0^\cdot e^{X_s} ds]_t,
\end{aligned}$$

because the integrator u is continuous. Note that X_t is càdlàg so all the integrals are well-defined. By associativity of the stochastic integral and (4.6) we have for the first term

$$\begin{aligned}
\int_{0^+}^t e^{X_{s^-}} d\left(\int_0^s e^{X_y} dy\right) &= \int_{0^+}^t e^{-X_{s^-}} e^{X_s} ds \\
&= \int_{0^+}^t e^{-X_s} e^{X_s} ds \\
&= t,
\end{aligned}$$

and for the last term we do some rewriting to conclude

$$\begin{aligned}
[e^{-X_\cdot}, \int_0^\cdot e^{X_s} ds]_t &= [\int_0^t e^{-X_s}\, d1_{[t,\infty)}(s), \int_0^\cdot e^{X_s} ds]_t \\
&= \int_0^t 1\, d[1_{[t,\infty)}, s] = 0.
\end{aligned}$$

Definition 4.8 and our previous result develop

$$\begin{aligned}
\sigma_{t^+}^2 &= \beta \int_0^t e^{-X_t+X_s} ds + e^{-X_t}\sigma_0^2 \\
&= \beta \cdot \Big(t + \int_{0^+}^t (\int_0^s e^{X_u}\ du) e^{-X_{s^-}} e^{X_{s^-}}\ d(e^{-X_s})\Big) + e^{-X_t}\sigma_0^2 \\
&= \beta \cdot \Big(t + \int_{0^+}^t (\int_0^s e^{-X_{s^-}} e^{X_u}\ du) e^{X_{s^-}}\ d(e^{-X_s})\Big) + e^{-X_t}\sigma_0^2 \\
&= \beta t + \int_{0^+}^t (\sigma_s^2 - e^{-X_{s^-}}\sigma_0^2) e^{X_{s^-}}\ d(e^{-X_s}) + e^{-X_t}\sigma_0^2 \\
&= \beta t + \int_0^t \sigma_s^2 e^{X_{s^-}}\ d(e^{-X_s}) + \sigma_{0^+}^2, \quad t > 0.
\end{aligned}$$

We use lemma 5.1 and assumption 4.4 to obtain

$$\begin{aligned}
& \sigma_{t^+}^2 \\
=\ & \sigma_{0^+}^2 + \beta t + \int_0^t \sigma_s^2 e^{X_{s^-}} d(e^{-X_s}) \\
=\ & \sigma_{0^+}^2 + \beta t + \log\delta \int_0^t \sigma_s^2 e^{X_{s^-}} e^{-X_s}\ ds + \int_0^t e^{X_{s^-}} e^{-X_{s^-}}\ d(\frac{\lambda}{\delta} \sum_{0<s\le t} \sigma_s^2 (\Delta L_s)^2) \\
=\ & \sigma_{0^+}^2 + \beta t + \log\delta \int_0^t \sigma_s^2\ ds + \frac{\lambda}{\delta} \sum_{0<s\le t} \sigma_s^2 (\Delta L_s)^2,
\end{aligned}$$

because X_t has only countable many discontinuities by (4.6). Assumption 4.4 gives the final answer

$$\sigma_t^2 = \sigma_0^2 + \beta t + \log\delta \int_0^t \sigma_s^2\ ds + \frac{\lambda}{\delta} \sum_{0<s<t} \sigma_s^2 (\Delta L_s)^2.$$

□

In resemblance, for the discrete-time model we have (write σ_n to indicate that we are in the discrete case)

$$\sigma_{n+1}^2 - \sigma_n^2 = \beta - (1-\delta)\sigma_n^2 + \lambda\sigma_n^2\epsilon_n^2, \quad n \in \mathbb{Z}_{\ge 0},$$

which by summation yields

$$\sigma_n^2 = \beta n - (1-\delta)\sum_{i=0}^{n-1} \sigma_i^2 + \lambda \sum_{i=0}^{n-1} \sigma_i^2 \epsilon_i^2 + \sigma_0^2.$$

Thus, the continuous-time model has the same feedback and autoregressive relation as in the discrete case, only the parameters are shifted. If we use for both models the same starting distribution, then

$$(\delta, \lambda) \quad \mapsto \quad (\log(\delta) + 1, \frac{\lambda}{\log(\delta) + 1}),$$

where δ, λ denote the variables in the discrete-case. This property should not be taken lightly. These feedback and autoregressive properties are important features of the volatility process.

5.2 The COGARCH process

As mentioned before, we need some conditional variance relation for G_t and our volatility process σ_t. For studying our defined COGARCH process (G_t) we need some notation. Let (b, a^2, ν^L) be the characteristic triplet for our (arbitrary) Lévy process L_t. For $t \geq 0$ we define

$$\begin{aligned}
B_{t,\epsilon} &:= \sum_{s \leq t} \Delta L_s \cdot 1_{\{\epsilon < |\Delta L_s| \leq 1\}}, \quad 0 < \epsilon < 1, \\
C_t &:= \lim_{\epsilon \downarrow 0} (B_{t,\epsilon} - \mathbb{E} B_{t\epsilon}), \\
A_t &:= bt + \sum_{s \leq t} \Delta L_s 1_{\{|\Delta L_s| > 1\}}, \\
M_t &:= aW_t + C_t, \quad W_t \text{ a Brownian Motion,}
\end{aligned}$$

such that the Lévy-Itô decomposition tells us

$$L_t \;=\; A_t + M_t, \quad t \geq 0.$$

Here, A_t is of bounded variation and M_t is the Brownian motion plus a martingale part (see [13]). Note that the covariation process of C_t is given by

$$[C_\cdot, C_\cdot]_t = \lim_{\epsilon \downarrow 0} \sum_{s \leq t} (\Delta L_s)^2 \cdot 1_{\{\epsilon < |\Delta L_s| \leq 1\}}, \quad 0 < \epsilon < 1.$$

Theorem 5.3. *There holds*

$$\lim_{h \downarrow 0} \frac{var(G_{t+h} - G_t | \mathcal{F}_t)}{h} = (a^2 + c)\sigma_t^2 \quad a.s, \quad t \geq 0,$$

where $c := \lim_{h \downarrow 0} \frac{[C_\cdot, C_\cdot]_{t+h} - [C, C]_t}{h}$ *is independent of* t.

Proof. Let $t \geq 0$ be given. Recall $[W_., W_.]_t = t$, so

$$\begin{aligned} [G_., G_.]_t &= [L_., L_.]_t \\ &= [M_., M_.]_t \\ &= a^2 t + [C_., C_.]_t. \end{aligned}$$

This is well-defined, because $C_t \in \mathcal{L}^2$ is a martingale. The conditional variance of $G_{t+h} - G_t$ is defined by

$$\operatorname{var}(G_{t+h} - G_t|\mathcal{F}_t) = \mathbb{E}[(G_{t+h} - G_t)^2|\mathcal{F}_t] - (\mathbb{E}[G_{t+h} - G_t|\mathcal{F}_t])^2.$$

We have

$$\begin{aligned} & \mathbb{E}[G_{t+h} - G_t|\mathcal{F}_t] \\ = & \mathbb{E}[\int_t^{t+h} \sigma_s \, dA_s|\mathcal{F}_t] \\ = & \mathbb{E}[b \cdot \int_t^{t+h} \sigma_s \, ds|\mathcal{F}_t] + \mathbb{E}[\int_t^{t+h} \sigma_s \, d(\sum_{u \leq s} \Delta L_u 1_{\{|\Delta L_u| > 1\}})|\mathcal{F}_t]. \end{aligned}$$

The number of jumps bigger than 1 are finite. So for $h > 0$ small enough follows

$$\mathbb{E}[G_{t+h} - G_t|\mathcal{F}_t] = \mathbb{E}[b \cdot \int_t^{t+h} \sigma_s \, ds|\mathcal{F}_t],$$

and using the definition of the Lebesgue-Stieltjes integral we obtain

$$\lim_{h \downarrow 0} \frac{(\mathbb{E}[G_{t+h} - G_t|\mathcal{F}_t])^2}{h} = 0 \quad \text{a.s.}$$

Through the Itô isometry we develop

$$\begin{aligned} \mathbb{E}[(G_{t+h} - G_t)^2|\mathcal{F}_t] &= \mathbb{E}[(\int_t^{t+h} \sigma_s \, dL_s)^2|\mathcal{F}_t] \\ &= \mathbb{E}[(\int_t^{t+h} \sigma_s \, dM_s)^2|\mathcal{F}_t] + \mathbb{E}[(\int_t^{t+h} \sigma_s \, dA_s)^2|\mathcal{F}_t] \\ &= \mathbb{E}[\int_t^{t+h} \sigma_s^2 \, d[M_., M_.]_s|\mathcal{F}_t] + \mathbb{E}[(\int_t^{t+h} \sigma_s \, dA_s)^2|\mathcal{F}_t] \\ &= \mathbb{E}[\int_t^{t+h} a^2 \sigma_s^2 \, ds + \int_t^{t+h} \sigma_s^2 \, d[C_., C_.]_s|\mathcal{F}_t] \\ &\quad + \mathbb{E}[(\int_t^{t+h} \sigma_s \, dA_s)^2|\mathcal{F}_t]. \end{aligned}$$

Hence,

$$\lim_{h \downarrow 0} \frac{\mathrm{var}(G_{t+h} - G_t | \mathcal{F}_t)}{h} = (a^2 + c)\sigma_t^2 \quad \text{a.s.},$$

where $c = \lim_{h\downarrow 0} \frac{[C.,C.]_{t+h} - [C,C]_t}{h}$. Note that c is independent of t, because the increments of a Lévy process are strictly stationary. □

In the discrete case we had that the conditional variance of X_n was equal to $\mathbb{E}[X_n^2 | \mathcal{F}_{n-1}] = \sigma_n^2$. Keeping in mind that the time difference is h instead of 1, the conditional variance of $G_{t+h} - G_t$ corresponds, for small h, up to a constant $a^2 + c$ compared to the discrete case.

5.3 Further results

This section we will state some further results, concerning the COGARCH and corresponding volatility process, that are obtained in [6]. It will confirm even more that this model preservers all stylized features of the discrete model. In Remark 4.2 necessary and sufficient conditions where given under which σ_n^2 and X_n converge in distribution to respectively finite random variables σ^2 and X. For X_n, it was a consequence of the convergence of σ_n^2 to a finite random variable. The next theorem tells us a convergence result for the continuous-time process.

Theorem 5.4. *Suppose*

$$\int_{\mathbb{R}} \log(1 + (\frac{\lambda}{\delta} y^2)\nu_L(dy) < -\log \delta. \tag{5.2}$$

Then $\sigma_t^2 \xrightarrow{d} \sigma_\infty^2$, as $t \to \infty$, for a finite random variable σ_∞^2 satisfying

$$\sigma_\infty^2 \stackrel{d}{=} \beta \int_0^\infty e^{-X_t} dt.$$

Conversely, if (5.2) does not holds, then $\sigma_t^2 \xrightarrow{\mathbb{P}} \infty$ as $t \to \infty$.

Proof. See [6]. □

Note that (5.2) incorporates the requirement that the integral is finite, because $0 < \delta < 1$ by Assumption 4.3. Also, the proof shows that the above improper integral exists as a finite random variable a.s. In comparison with condition (i) in Remark 4.2, condtion (5.2) differs only in the measure used for the integration, which can be explained by the difference of the noise variables.

We have that σ_t^2 is Markovian and further that, if the process is started at $\sigma_0^2 \stackrel{d}{=} \sigma_\infty^2$, then it is strictly stationary.

Theorem 5.5. *The squared volatility process* $(\sigma_t^2)_{t\geq 0}$ *is a homogeneous Markov process. Moreover, if the limit* σ_∞^2 *in Theorem 5.4 exists and* $\sigma_0^2 \stackrel{d}{=} \sigma_\infty^2$*, independent of* $(L_t)_{t\geq 0}$*, then* $(\sigma_t^2)_{t\geq 0}$ *is strictly stationary.*

Proof. See [6] □

For the process $G_t = \int_0^t \sigma_s dL_s,\ t \geq 0$, note that for any $0 \leq y < t$,

$$G_t = G_y + \int_{y^+}^t \sigma_s dL_s, \quad t \geq 0.$$

Here, $(\sigma_s)_{y<s\leq t}$ depends on the past until time y only through σ_y, and the integrator is independent of this past. From the previous Theorem we thus obtain:

Corollary 5.6. *The bivariate process* $(\sigma_t, G_t)_{t\geq 0}$ *is Markovian. If* $(\sigma_t^2)_{t\geq 0}$ *is the strictly stationary version of the process with* $\sigma_0^2 \stackrel{d}{=} \sigma_\infty^2$*, then* $(G_t)_{t\geq 0}$ *is a process with strictly stationary increments.*

Thus as in the discrete case the processes $(\sigma_t)_{t\geq 0}$ and $(\sigma_t, G_t)_{t\geq 0}$ are Markov process (when started in σ_∞^2).

As was mentioned after we defined G_t and what Theorem 5.3 confirms, we have to look at the moments of the increments of G_t in arbitrary time intervals. Consequently, for $r > 0$ set

$$G_t^{(r)} := G_{t+r} - G_t.$$

There exists the following result.

Theorem 5.7. *Suppose* $(L_t)_{t\geq 0}$ *is a quadratic pure jump process with* $\mathbb{E}L_1^2 < \infty$, $\mathbb{E}L_1 = 0$, *and* $\log \mathbb{E}[e^{-X_1}] < 0$. *Let* $(\sigma_t^2)_{t\geq 0}$ *be the strictly stationary volatility process with* $\sigma_0^2 \stackrel{d}{=} \sigma_\infty^2$. *Then for any* $t \geq 0$ *and* $h \geq r > 0$,

$$\begin{aligned}
\mathbb{E}[G_t^{(r)}] &= 0, \\
\mathbb{E}[(G_t^{(r)})^2] &= \frac{\beta r}{-\log \mathbb{E}[e^{-X_1}]}\mathbb{E}L_1^2, \\
cov(G_t^{(r)}, G_{t+h}^{(r)}) &= 0.
\end{aligned}$$

Proof. See [6]. $\square$

This uncorrelated property is concordance with the discrete-time model. Note also that $\mathbb{E}[(G_t^{(r)})^2]$ is independent of t. Here, L_t is a pure jump process. So Theorem 5.3 yields $(G_t^{(r)}) \approx r\sigma_t^2$ for r small.

6 Recent developments

Our COGARCH in combination with the corresponding volatility process is a continuous-time variant of the original GARCH process. At least, we have suggested that this is the case. We have derived a continuous-time process, which captures all the same stylized facts that are present in the discrete-time GARCH. Just like the bivariate diffusion model, we want to approximate our new process arbitrarily close to a GARCH process. In other words, we want to have a limit result as before.

Recently, in the paper of Kallsen and Vesenmayer (see [5]) it is shown that $(G_t, \sigma_t^2)_{t\geq 0}$ can indeed be obtained as a limit in law of a sequence of GARCH$(1,1)$ models. In contrast to our diffusion approximation, this result is obtained by a different limiting procedure. Whereas the diffusion result is developed through rescaling the size of the innovations, Kallsen and Vesenmayer apply some sort of random thinning. This is done by decreasing the probability of the nontrivial innovations. Here, the differential characteristics of a semimartinale X play an important role. If the characteristics converge, and some other condition holds, then the corresponding sequence of processes also converges weakly. They also conjecture, by a heuristical argument, that the bivariate diffusion process and the COGARCH process (in combination with his volatility process) are probably the only continuous-time limit is of GARCH.

7 Conclusion

We started bij looking into the discrete GARCH model. There we investigated the Markov property and what conditions are needed. In many articles it was stated that it was Markov without proof and without assuming conditions that were seen necessary in our analysis. After that, we have studied two different continuous-time models. First, we have derived a result using diffusion approximation. In the limit we obtained a Itô process that has a weak solution to the stochastic differential equation. This solution was unique in law, existsed and was continuous in probability. In addition to Nelson, we have completely proved and stated all necessary assumptions needed to achieve this. The stochastic differential equation was given in terms of two independent Brownian Motions. The "direction" and the conditional variance of the displacement, over a small time-interval, is determined by these two independent Brownian Motions. Also, jumps are not present in the bivariate diffusion.

In our second model we relaxed this independence property and made jumps possible, because sometimes one needs some dependency between the direction and the volatility. This was done by replacing the noise variables by increments of a Lévy process, and we acquired a continuous-time volatility process of bounded variation. Next to that, we defined a continuous-time GARCH (called COGARCH) process as a solution of a stochastic differential equation. This COGARCH process only jumps if the corresponding Lévy process does. For the volatility process we proved that the same important feedback and autoregressive properties hold. If we condition it on the past, then the the COGARCH process is in some sense equal to the volatility process. It is worth mentioning that this is a property that is given in the original definition of the linear GARCH model. Also, some other important properties stayed intact in the continuous case, such as uncorrelated increments and the Markov property for the volatility process. Also the bivariate process was Markovian when started in the strictly stationary distribution given by σ_∞^2.

Finally, we have given an important feature of the COGARCH process in combination with his volatility process. Namely, it is shown by Kallsen and Vesenmayer (see [5]) that the COGARCH process (in combination with the volatility process) can be obtained as limit in law of a sequence of GARCH$(1,1)$ models.

References

[1] Billingsley, P., 1968, Convergence of probability measures, Wiley, New York.

[2] Bollorslev, T., 1986, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics **31**, 307-327.

[3] Goldie, C.M., Maller, R.A., 2000, *Stability of perpetuities*, Annals of Probability **28**, 1195-1218.

[4] Jacod, J., Shityaev, A.N., 2003, *Limit theorems for stochastic processes*, second edition, Springer-Verlag Berlin Heidelberg New York.

[5] Kallsen, J., Vesenmayer, B., 2009, *COGARCH as a continuous-time limit of GARCH*$(1,1)$, Stochastic Processes and their Applications **119**, 74-98.

[6] Klüppelberg, C., Lindner, A., Maller, R., 2004, *A continuous-time GARCH process driven by a Lévy process: stationarity and second-order behaviour*, Journal Applied Probability **41**, 601-622.

[7] Nelson, D., 1990, *Arch models as diffusion approximations*, Journal of Econometrics **45**, 7-38.

[8] Protter, P.E., 2005, *Stochastic integration and differential equations*, second edition, Springer Berlin Heidelberg New Yourk.

[9] Spreij, P., 2010, *Stochastic integration*, UVA University Amsterdam, lecture notes.

[10] Van der Vaart, A.W., 2010, *Time Series*, VU University Amsterdam, lecture notes.

[11] Spieksma, F., 2010, *An Introduction to Stochastic Processes in Continuous Time*, an adaption of the original text written by Zanten, H., Leiden University, lecture notes.

[12] Williams, D., 2007 (tenth printing), *Probability with Martingales*, Cambridge University Press.

[13] Zanten, H., 2007, *An Introduction to Stochastic Processes in Continuous Time*, VU University Amsterdam, lecture notes.