Bachelor Thesis

Modeling Time-Varying Volatility

Nikolaus Schäfer

2019

Faculty of Mathematics and Computer Science
Heidelberg University

Supervision: Prof. Dr. Rainer Dahlhaus

Declaration of Authorship

I hereby declare that I am the sole author of this bachelor thesis. All sources I used are acknowledged as references. I certify that I have not submitted this work at any other institution. Furthermore, I am aware that this work might be screened electronically for plagiarism.

Heidelberg, August 31, 2019

Nikolaus Schäfer

Abstract

With the introduction of the ARCH and GARCH models in the 1980s (see Engle (1982) and Bollerslev (1986)) volatility modeling has become a vast area of research in financial mathematics. The first part of this thesis introduces these models. The second part deals with a specific modification of the GARCH-model, the GJR-GARCH model of Glosten et al. (1993), which is able to react asymmetrically to positive and negative shocks. Although these models work well for short-term forecast they are unable to incorporate structural changes of volatility over time. Therefore, the third and main part of this thesis discusses the TV-GJR-GARCH model introduced by Amado and Teräsvirta (2013), which is able to model smooth structural changes in volatility. Specifically, the multiplicative decomposition of the variance of the model into a conditional (described by the GJR-GARCH model) and unconditional component is examined. The main focus lies on the quasi-maximum-likelihood estimation of the TV-GJR-GARCH model through maximization by parts (see Song et al. (2005)). A rigorous proof of the consistency and asymptotic normality of the quasi-maximum-likelihood estimator is provided.

Zusammenfassung

Mit der Einführung der ARCH- und GARCH-Modellen in den 1980er Jahren (siehe Engle (1982) and Bollerslev (1986)) wurde die Volatilitätsmodellierung zu einem großen Bereich der Forschung in der Finanzmathematik. Der erste Teil dieser Thesis stellt diese Modelle vor. Der zweite Teil behandelt eine spezifische Modifizierung des GARCH-Modells, das GJR-GARCH-Modell von Glosten et al. (1993), welches asymmetrisch zu positiven und negative Schocks reagieren kann. Obwohl diese Modelle für kurzzeitige Vorhersagen gut funktionieren, können sie strukturelle Veränderungen der Volatilität über die Zeit nicht umfassen. Deshalb behandelt der dritte Teil dieser Thesis das TV-GJR-GARCH-Modell eingeführt von Amado and Teräsvirta (2013), welches glatte strukturelle Veränderungen der Volatilität modellieren kann. Spezifisch die multiplikative Zerlegung der Varianz des Modells in eine bedingte (beschrieben durch das GJR-GARCH-Modell) und eine unbedingte Komponente wird beleuchtet. The Hauptfokus liegt auf der Quasi-Maximum-Likelihood Schätzung des TV-GJR-GARCH-Modells durch Maximierung in Teilen (siehe Song et al. (2005)). Ein gründlicher Beweis der Konsistenz und asymptotischen Normalität des Quasi-Maximum-Likelihood Schätzers ist gegeben.

Contents

1 Introduction

In stock trading, every trader would like to acquire assets that have little risk and a large expected return. Hence, a major area of research in applied mathematics and economics is the accurate modeling of financial time series. A *time series* is a sequence of numerical data points in successive order. In financial studies, a time series often tracks returns of assets over a certain period of time.
One widely analyzed financial time series is the one-period *log-return* of a certain financial asset

$$X_t = \log P_t - \log P_{t-1}, \quad t = 1, 2, 3, ...,$$

where P_t is the price of the asset at time index t. The advantage of considering the log-return instead of the simple return, i.e. $P_t - P_{t-1}$, is that statistical properties of log-returns are more tractable. As we will consider only discrete time series, for the remainder of this thesis t will take values only in either $\mathbb{N}_0$ or $\mathbb{Z}$. Define

$$\mu_t := \mathbb{E}[X_t \mid \mathcal{F}_{t-1}],$$

the conditional mean of X_t given the σ-algebra $\mathcal{F}_{t-1}$ generated by $X_{t-1}, \ldots, X_0$ (the information set available at time $t-1$). Generally, the time series X_t can be divided into a predictable part and an unpredictable part, i.e.

$$X_t = \mu_t + \epsilon_t. \tag{1.1}$$

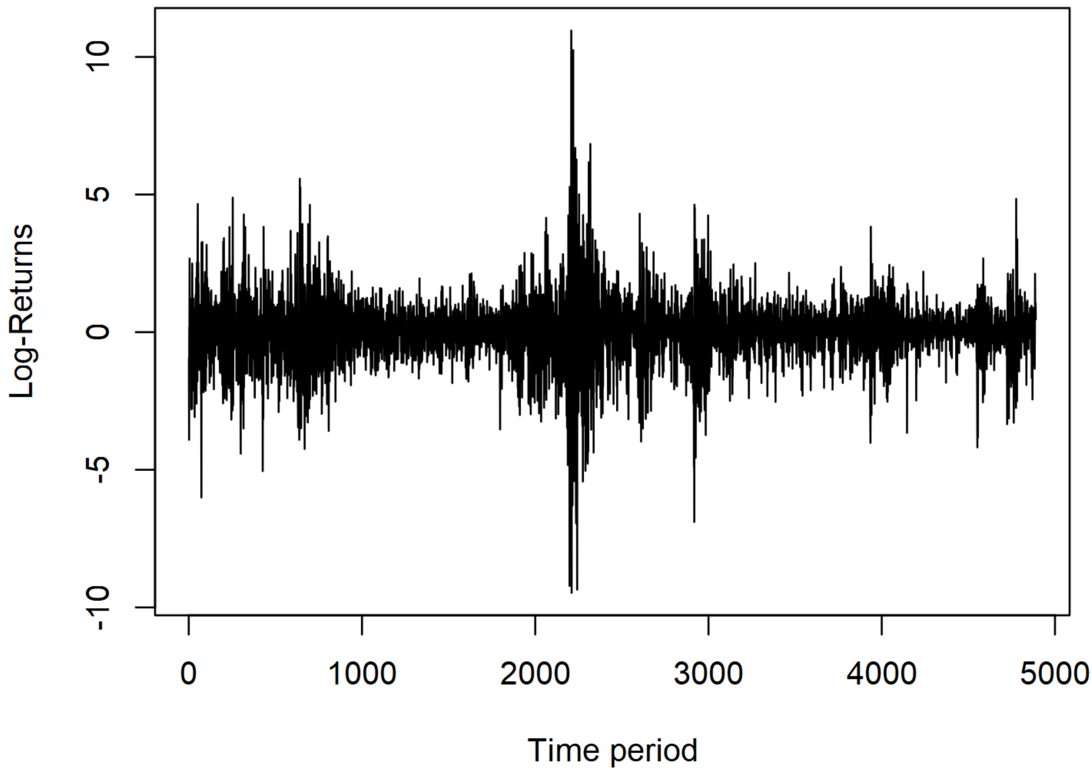
For a stock trader holding an asset at time t, there are two major points of interest. Firstly, the expected return in future periods, $\mathbb{E}[X_{t+k}] (k \in \mathbb{N})$, and secondly the risk associated with holding the asset, i.e. the probability of a negative return in period $t+k$. A common risk measure in financial mathematics is the *volatility*, i.e. the standard deviation of X_t conditional on previous returns

$$\sqrt{\mathrm{Var}[X_t \mid \mathcal{F}_{t-1}]} = \sqrt{\mathbb{E}[\epsilon_t^2 \mid \mathcal{F}_{t-1}]} = \sigma_t.$$

Intuitively speaking, the volatility is a measure for the degree of fluctuation in X_t.
It goes without saying that volatility has many applications in the financial world. In options trading, volatility takes up an especially important role. In the well-known Black-Scholes option pricing formula, σ_t plays a major factor in computing the price of a European *call option*. A call option is a contract giving its holder the right, but not the obligation, to buy a fixed number of shares of a specified common stock at a fixed price on a given date. Another

application worth mentioning is that in risk management, volatility modeling supplies a simple approach to computing value at risk of a financial condition. On top of that, it is also possible to specify risk premiums of financial investments as a function of volatility.
Yet, the difficulty of modeling volatility is equally as great as its importance. Although the Black-Scholes model assumes homoscedasticity, a quick look at Figure 1 calls this approach very much into question. [1] The graph displays the log-returns of the DAX of 4889 consecutive days starting January 1st 2000. It is clear that typical time series (as represented in Figure 1) suffer from heteroscedastic effects. This is unsurprising as log-returns are of course vulnerable to economic and political influences. The big oscillations in the middle of the graph of Figure 1, for example, can be accredited to the American financial crisis. This vulnerability to external factors makes an accurate modeling of volatility quite difficult.

Figure 1: `Log-Returns DAX`



Notes: Log-Returns of the DAX between January 1st 2000 and June 10th 2019 (n = 4889).

Engle (1982) introduced the ARCH (autoregressive conditional heteroscedasticity) models in order to provide a solution to this problem. It was the first approach that provided a systematic yet simple framework for volatility modeling and was awarded with the noble prize in 2003.

[1] Homoscedasticity would assume a constant volatility over time. The antonym is heteroscedasticity.

The basic idea, as one can already derive from the name, is that the squared volatility σ_t^2 is underlying an autoregressive structure, $\sigma_t^2 \equiv \sigma_t^2(\mathcal{F}_{t-1})$. ϵ_t from (1.1) is expressed as

$$\epsilon_t = z_t \sigma_t \tag{1.2}$$

to introduce conditional heteroscedasticity, where z_t are i.i.d. random variables with $\mathbb{E}[z_t] = 0$ and $\mathrm{Var}[z_t] = 1$.
Equations (1.1) and (1.2) build the basic framework of every ARCH-model. The difference between the models depends upon the precise definition of σ_t^2. The simplest form is the standard ARCH-model itself, where (1.2) adopts the following representation, called ARCH(q)-times series:

Definition 1.1. `ARCH(`p`)-time series`
A real time-series ($\epsilon_t : t \in \mathbb{Z}$) is called an ARCH($q$) time-series if it is recursively defined as:

$$\begin{aligned} \epsilon_t &= z_t \sigma_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{q} \alpha_1 \epsilon_{t-i}^2, \end{aligned}$$

where $\alpha_0, \ldots, \alpha_q \in \mathbb{R}_{\geq 0}$, $\alpha_q \neq 0$.

Notice that $\alpha_0 > 0$ and $\alpha_i \geq 0$ in order to ensure $\sigma^2 > 0$. Another useful observation is that volatility clustering,[2] as first introduced by Mandelbrot (1963), can be accurately described by this model. Large past shocks $\{X_{t-i}^2\}_{i=1}^q$ tend to be followed by another large shock and vice versa. Tsay (2005) points out, that "tend" is a key word in this context as a large conditional variance σ_t^2 does not directly lead to a large X_t. It only implies that the probability of obtaining a large realization is greater than that of a smaller variance.
Tim Bollerslev (1986) refined the ARCH model to a generalized ARCH (short GARCH) model, which implements lags of σ_t^2 to the computation of σ_t^2 itself.

Definition 1.2. `(GARCH(`p, q`)-time series)`
A real time series ($\epsilon_t : t \in \mathbb{Z}$) is called a GARCH($p, q$)-time series if it is recursively defined as:

$$\begin{aligned} \epsilon_t &= z_t \sigma_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \end{aligned}$$

where ($\sigma_t^2 : t \in \mathbb{Z}$) is a nonnegative time series with $\alpha_0, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p \in \mathbb{R}_{\geq 0}$, $\alpha_q, \beta_p \neq 0$.

[2]Volatility clustering refers to the observation that large (small) changes tend to be followed by large (small) changes of either sign (see Mandelbrot (1963)). It now belongs to the "stylized facts" of financial time series.

It is obvious that in the case of $q = 0$, $(\epsilon_t : t \in \mathbb{Z})$ defines just a regular ARCH(p)-time series as in Definition 1.1. One weakness of the GARCH-model is that positive and negative shocks have the same effects on volatility. However, empirically this assumption is often violated. In fact, especially for asset returns it has been determined that negative shocks have a stronger effect on the volatility than positive shocks. Black (1976) equipped this observation with the term *leverage effect*. He hypothesized that negative news imply an increase in leverage through a reduced market value of the firm, which in turn leads to a rise in the volatility of the stock. Although the explanation is merely a hypothesis, the term "leverage effect" has become prevalent to describe this phenomena. In any case, the most important thing is that empirically speaking there is no doubt that the volatility responds asymmetrically to the sign of the shock. Therefore, a number of models have been developed in the past decades to account for this behavior. Most notably the EGARCH (see Nelson (1991)), QGARCH (see Sentana (1995)) or the GJR-GARCH (see Glosten et al. (1993)). Though these models allow the GARCH model to react asymmetrically, one problem remains. They are unable to incorporate structural changes of the volatility over long periods of time.[3] In some empirical contexts the assumption of stationary data (financial return series) is violated. Looking at Figure 1 it seems unreasonable to assume that the mean and variance stay constant over time.
This thesis deals with contexts where the volatility seems to be nonstationary. Specifically, it will deal with a nonstationary GARCH model introduced by Amado and Teräsvirta (2013) which they call the multiplicative time varying GJR-GARCH (TV-GJR-GARCH) model.[4] As the name suggests, the model builds on the above mentioned (stationary) GJR-GARCH model. Therefore, the structure of the thesis is as follows. Section 2 will discuss the GJR-GARCH model. I will provide a proof for its stationarity under certain conditions and calculate the unconditional variance of the (stationary) GJR-GARCH model. The first part of Section 3 then focuses on the above mentioned TV-GJR-GARCH model. In particular, we will see that Amado and Teräsvirta (2013) split the variance into two components, a component which describes conditional heteroscedasticity (modeled by the GJR-GARCH model) and a component which describes nonstationarity. The second part of Section 3 then deals with the estimation of the model. In particular an extensive proof of consistency and asymptotic normality of a (quasi-)maximum-likelihood estimator is provided which was omitted in Amado and Teräsvirta (2013).[5]

[3] It is reasonable to assume that changes in institutions, or political and economic events alter the structure of volatility in the long-run (see e.g. Mikosch and Stărică (2004) for details).

[4] The authors also shortly discuss an additive TV-GJR-GARCH model (i.e. σ_t^2 has an additive structure), however, this thesis will solely focus on the model were σ_t^2 has a multiplicative structure. For simplicity, for the remainder of this thesis I will call the multiplicative TV-GJR-GARCH model just the TV-GJR-GARCH model.

[5] Amado and Teräsvirta (2013) only refer to various literature rather than providing a rigorous proof.

2 The GJR-GARCH Model

This section gives an overview of the GJR-GARCH model as it is the basis for the time-varying model discussed in Section 3. This introductory part of the section will deal with the definition of a GJR-GARCH (p, q)-time series. Section 2.1 will deal with its stationarity. Eventually, Section 2.2 derives two important properties of the stationary GJR-GARCH model used Section 3.
Again, Equations (1.1) and (1.2) build the groundwork.

Definition 2.1.
A real time series $(\epsilon_t : t \in \mathbb{Z})$ is called a GJR-GARCH(p, q)-time series if it is recursively defined as:

$$\begin{aligned} \epsilon_t &= z_t \sigma_t, & (2.1) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{q} (\alpha_i + \lambda_i I_{t-i}) \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, & (2.2) \end{aligned}$$

where

$$I_{t-i} := \begin{cases} 0 & \text{if } \epsilon_{t-i} > 0 \\ 1 & \text{if } \epsilon_{t-i} < 0, \end{cases}$$

and $(\sigma_t^2 : t \in \mathbb{Z})$ is a nonnegative time series with $\alpha_0, \ldots, \alpha_q, \lambda_1, \ldots, \lambda_q, \beta_1, \ldots, \beta_p \in \mathbb{R}_{\geq 0}$, $\alpha_q, \beta_p \neq 0$. $(z_t : t \in \mathbb{Z})$ are i.i.d. normally distributed random variables with $\mathbb{E}[z_t] = 0$ and $\mathrm{Var}[z_t] = 1$. Furthermore, z_t is independent of ϵ_{t-s}, $s > 0$.

Notice that the indicator variable I_{t-i} is responsible for the asymmetric shocks as negative past shocks lead to a higher future volatility than positive past shocks.

2.1 Stationarity

Define

$$r := \max\{p, q\}$$

Notice that Equation (2.2) can be rewritten as

$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{r} c_k(z_{t-k}) \sigma_{t-k}^2 \tag{2.3}$$

where

$$c_k(z_{t-k}) = \beta_k + (\alpha_k + \lambda_k I(z_{t-k} < 0))z_{t-k}^2,$$

$k = 1, \ldots, r$ and $\alpha_i, \lambda_i, \beta_j = 0$ for $i > q$ and $j > p$.
Equation (2.3) in turn can be rewritten as

$$\boldsymbol{Y}_t = \boldsymbol{A}_t \boldsymbol{Y}_{t-1} + \boldsymbol{B}_t \tag{2.4}$$

where

$$\boldsymbol{Y}_t = (\sigma_t^2, \sigma_{t-1}^2, \ldots, \sigma_{t-r+1}^2)^t \in \mathbb{R}^r \text{ and } \boldsymbol{B}_t = (\alpha_0, 0, \ldots, 0)^t \in \mathbb{R}^r$$

$$\boldsymbol{A}_t = \begin{pmatrix} c_1(z_{t-1}) & \cdots & c_{r-1}(z_{t-r+1}) & c_r(z_{t-r})) \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \in \mathbb{R}^{r \times r}.$$

Thus, there is a unique stationary solution of the process described in (2.1) and (2.2) if and only if the stochastic recurrence equation (SRE) descried in (2.4) has a unique stationary nonnegative solution such that $\boldsymbol{Y}_t$ is $\mathcal{F}_{t-1}$-measurable.

Chen and An (1998) use a similar approach to provide a sufficient condition for stationarity of the GARCH model. The following theorem and its proof extend their work to the GJR-GARCH model.

Theorem 2.2. *(Stationarity of GJR-GARCH(p, q))*
Let $\boldsymbol{Y}_t$ be a process as defined in (2.4). Assume $\mathbb{E}[|z_t|^2] < \infty$, $\alpha_0 < \infty$, and $\mathbb{E}[c_k(z_t)] < \infty$ where $k = 1, \ldots, r$ and $\alpha_i, \lambda_i, \beta_j = 0$ for $i > q$ and $j > p$. Then $\boldsymbol{Y}_t$ (and thus the GJR-GARCH model in (2.1 and (2.2)) has a unique strictly stationary solution if and only if

$$\sum_{i=1}^{q}(\alpha_i + \lambda_i/2) + \sum_{j=1}^{p}\beta_j < 1. \tag{2.5}$$

Moreover, this solution is $\mathcal{F}_{t-1}$-measurable and is given by

$$\boldsymbol{Y}_t = \boldsymbol{B}_t + \sum_{k=1}^{\infty} \left(\prod_{j=0}^{k-1} \boldsymbol{A}_{t-j} \right) \boldsymbol{B}_{t-k}.^{6} \tag{2.6}$$

Proof. Before we begin with the actual proof. It is necessary to make the following observations:

$$\begin{aligned} \mathbb{E}[c_k(z_{t-k})] &= \beta_k + (\alpha_k + \lambda_k \underbrace{\mathbb{E}[I(z_{t-k} < 0)]}_{=\mathbb{P}(z_{t-k}<0)})\mathbb{E}[z_{t-k}^2] \\ &= \beta_k + \alpha_k + \lambda_k/2, \end{aligned} \tag{2.7}$$

since $\{z_t\}$ is a sequence of i.i.d. normally distributed random variables with $\mathbb{E}[z_t] = 0$ and $\text{Var}[z_t] = 1$. Thus,

$$\mathbb{E}[\boldsymbol{A}_t] = \begin{pmatrix} \beta_1 + (\alpha_1 + \lambda_1/2) & \cdots & \beta_{r-1} + (\alpha_{r-1} + \lambda_{r-1}/2) & \beta_r + (\alpha_r + \lambda_r/2) \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \equiv \boldsymbol{A} \tag{2.8}$$

is independent of t. Similarly,

$$\mathbb{E}[\boldsymbol{B}_t] = (\alpha_0, 0, \ldots, 0)^T \equiv \boldsymbol{B} \tag{2.9}$$

is also independent of t. By the definition of $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ it is easy to see that both $\{\boldsymbol{A}_t\}$ and $\{\boldsymbol{B}_t\}$ are sequences of independent, nonnegative random vectors and $\boldsymbol{A}_{t-j}$ is independent of $\boldsymbol{B}_{t-k}$ for $k \neq j$. Therefore, we have

$$\mathbb{E}[\boldsymbol{Y}_t] = \boldsymbol{B}(1 + \sum_{k=1}^{\infty} \left(\prod_{j=0}^{k-1} \mathbb{E}[\boldsymbol{A}_{t-j}] \right) = \boldsymbol{B}(1 + \sum_{k=1}^{\infty} \boldsymbol{A}^k)$$

"$\Leftarrow$": First, we show that if (2.5) holds $\sum_{k=1}^{\infty} \left(\prod_{j=0}^{k-1} \boldsymbol{A}_{t-j} \right) \boldsymbol{B}_{t-k}$ converges almost surely.
By the definition of $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ it is easy to see that both $\{\boldsymbol{A}_t\}$ and $\{\boldsymbol{B}_t\}$ are sequences of independent nonnegative random vectors and $\boldsymbol{A}_{t-j}$ is independent of $\boldsymbol{B}_{t-k}$ for $k \neq j$. Therefore

[6] Hereby $\prod_{l=1}^{0} \cdot = 1$.

we have

$$\mathbb{E}\left[\left(\prod_{j=0}^{k-1}\boldsymbol{A}_{t-j}\right)\boldsymbol{B}_{t-k}\right]=\left(\prod_{j=0}^{k-1}\mathbb{E}[\boldsymbol{A}_{t-j}]\right)\mathbb{E}[\boldsymbol{B}_{t-k}]=\boldsymbol{A}^k\boldsymbol{B}$$

by the virtue of (2.8) and (2.9). Define $c_i := \beta_i + (\alpha_i + \lambda_i/2)$, $i = 1, \ldots, r$, where $\alpha_i, \lambda_i = 0$ for $i > q$ and $\beta_i = 0$ for $i > q$. Then, the characteristic polynomial of $\boldsymbol{A}$ is given by:

$$\begin{aligned}
\det(\mu\boldsymbol{I}-\boldsymbol{A}) &= \det\begin{pmatrix} \mu-c_1 & -c_2 & \cdots & \cdots & -c_r \\ -1 & \mu & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & \mu \end{pmatrix} \\
&= \underbrace{(\mu-c_1)\det\begin{pmatrix} \mu & & & \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ & & -1 & \mu \end{pmatrix}}_{=\mu^{r-1}} - (-1)\det\begin{pmatrix} -c_2 & \cdots & \cdots & -c_r \\ -1 & \mu & & \\ & \ddots & \ddots & \\ & & -1 & \mu \end{pmatrix} \\
&= \mu^r - \mu^{r-1}\cdot c_1 + \underbrace{(c-2)\det\begin{pmatrix} \mu & & & \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ & & -1 & \mu \end{pmatrix}}_{=\mu^{r-2}} + \det\begin{pmatrix} -c_3 & \cdots & \cdots & -c_r \\ -1 & \mu & & \\ & \ddots & \ddots & \\ & & -1 & \mu \end{pmatrix} \\
&\vdots \\
&= \mu^r - \sum_{i=1}^{r} c_i\mu^{r-i} = \mu^r\left(1-\sum_{i=1}^{r} c_i\mu^{-i}\right)
\end{aligned} \tag{2.10}$$

Let $\rho(\boldsymbol{A})$ denote the spectral radius of $\boldsymbol{A}$, i.e. $\sup\{|\mu| \colon \mu \in \sigma(\boldsymbol{A})\}$, where μ is an eigenvalue of $\boldsymbol{A}$ and $\sigma(\boldsymbol{A})$ the spectrum of $\boldsymbol{A}$. Following (2.10), it is evident, that $\rho(\boldsymbol{A}) < 1$ if and only if $\sum_{i=1}^{q}(\alpha_i + \lambda_i/2) + \sum_{j=1}^{p}\beta_j < 1$. Johnson and Horn (1985) showed that if $\rho(\boldsymbol{A}) < 1$, the geometric series $\sum_{k=1}^{\infty} = \boldsymbol{A}^k$ converges, i.e.

$$\sum_{k=1}^{\infty} \boldsymbol{A}^k < \infty.$$

Therefore

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{j=0}^{k-1} \boldsymbol{A}_{t-j}\right] \boldsymbol{B}_{t-k} < \infty,$$

and hence

$$\sum_{k=1}^{\infty} \left(\prod_{j=0}^{k-1} \boldsymbol{A}_{t-j}\right) \boldsymbol{B}_{t-k} < \infty.$$

Now, the vector-valued stochastic process $\{\boldsymbol{Y}_t\}$ defined by (2.6) is obviously strictly stationary and can be rewritten as

$$\begin{aligned}
\boldsymbol{Y}_t &= \boldsymbol{B}_t + \boldsymbol{A}_t \left[\boldsymbol{B}_{t-1} + \sum_{k=2}^{\infty} \left(\prod_{j=0}^{k-1} \boldsymbol{A}_{t-j}\right) \boldsymbol{B}_{t-k}\right] \\
&= \boldsymbol{B}_t + \boldsymbol{A}_t \left[\boldsymbol{B}_{t-1} + \sum_{l=1}^{\infty} \left(\prod_{j=0}^{l-1} \boldsymbol{A}_{t-1-l}\right) \boldsymbol{B}_{t-1-l}\right] \\
&= \boldsymbol{B}_t + \boldsymbol{A}_t \boldsymbol{Y}_{t-1}.
\end{aligned}$$

This almost completes the "if"-part of the theorem. The uniqueness of the strictly stationary solution will be proven in the end as we will use techniques introduced in the proof of the "only if" part of the theorem below.

"$\Rightarrow$": Conversely, assume (2.4) has a strictly stationary solution. Using Equation (2.4), we have

$$\begin{aligned}
\boldsymbol{Y}_0 &= \boldsymbol{A}_0 \boldsymbol{Y}_{-1} + \boldsymbol{B}_0 \\
&= \boldsymbol{B}_0 + \boldsymbol{A}_0 \boldsymbol{B}_{-1} + \boldsymbol{A}_0 \boldsymbol{A}_{-1} \boldsymbol{Y}_{-2} \\
&\vdots \\
&= \boldsymbol{B}_0 + \sum_{k=1}^{n-1} \left(\prod_{j=0}^{k-1} \boldsymbol{A}_{-j}\right) \boldsymbol{B}_{-k} + \left(\prod_{j=0}^{n-1} \boldsymbol{A}_{-j}\right) \boldsymbol{Y}_{-n} \qquad (2.11)
\end{aligned}$$

At this point it is important to note that all $\boldsymbol{A}_n$, $\boldsymbol{B}_n$ and $\boldsymbol{Y}_n$ are nonnegative, $\{\boldsymbol{A}_t\}$ is a sequence of independent random matrices, $\boldsymbol{A}_{n-j}$ and $\boldsymbol{B}_{n-k}$ are independent for $k \neq j$, and $\mathbb{E}[\boldsymbol{Y}_0] < \infty$. Taking the expectation on both sides of (2.11) gives us

$$\mathbb{E}[\boldsymbol{Y}_0] \geq \sum_{k=1}^{n-1} \mathbb{E}\left[\prod_{j=0}^{k-1} \boldsymbol{A}_{-j}\right] \boldsymbol{B}_{-k} = \sum_{k=1}^{n-1} \boldsymbol{A}^k \boldsymbol{B}$$

$$\begin{aligned} &\Rightarrow \quad \sum_{k=1}^{n} \boldsymbol{A}^k \boldsymbol{B} < \infty \\ &\Rightarrow \quad \lim_{n\to\infty} \boldsymbol{A}^n \boldsymbol{B} = 0. \end{aligned} \tag{2.12}$$

Let $\{\delta_i | i = 1, \ldots, r\}$ be the canonical basis of $\mathbb{R}^r$, i.e. $\delta_i = (\delta_{i,1}, \ldots, \delta_{i,r})^t$, where $\delta_{ij} = 0$, for $i \neq j$, $\delta_{ii} = 1$. If we can prove that for $1 \leq i \leq r$,

$$\lim_{n\to\infty} \boldsymbol{A}^n \delta_i = 0, \tag{2.13}$$

then (2.13) implies that $\lim\limits_{n\to\infty} \boldsymbol{A}^n = 0$, which in turn implies $\rho(\boldsymbol{A}) < 1$ (see Lemma 4.1 in the appendix). As we showed before, the latter is equivalent to $\sum\limits_{i=1}^{q} \alpha_i + \lambda_i/2 + \sum\limits_{j=1}^{p} \beta_j < 1$, which would complete the "only if" part of the lemma.

In fact, since $\boldsymbol{B} = \alpha_0 \delta_1$ and $0 < \alpha_0 < \infty$, by the definition of $\boldsymbol{A}$ and (2.12), (2.13) holds for $i = 1$.

Notice that for $i = r$ we have $\boldsymbol{A}\delta_r = c_r \delta_1$ by the definition of $\boldsymbol{A}$, c_i is again defined as $c_i := \beta_i + (\alpha_i + \lambda_i/2)$, $i = 1, \ldots r$. Therefore, we can write

$$\lim_{n\to\infty} \boldsymbol{A}^n \delta_r = \lim_{n\to\infty} \boldsymbol{A}^{n-1} c_r \delta_1 = 0.$$

The last equality follows from the fact that by definition of a GJR-GARCH process $0 < c_i < \infty$ for $i = 1, \ldots, r$ and that (2.13) holds for $i = 1$. We now turn to the case of $2 \leq i < r$. Here it is easy to see that we have

$$\boldsymbol{A}\delta_i = c_i \delta_1 + \delta_{i+1}.$$

To see that (2.13) holds even for $2 \leq i < r$, we perform the following backward recursion:

$$
\begin{aligned}
\lim_{n\to\infty} &= \lim_{n\to\infty} \boldsymbol{A}^{n-1}(c_i\delta_1 + \delta_{i+1}) \\
&= \underbrace{\lim_{n\to\infty} \boldsymbol{A}^{n-1} c_i \delta_1}_{\longrightarrow 0} + \underbrace{\lim_{n\to\infty} \boldsymbol{A}^{n-1}\delta_{i+1}}_{\lim_{n\to\infty} \boldsymbol{A}^{n-2}(c_{i+1}\delta_1 + \delta_{i+2})} \\
&\vdots \\
&= \lim_{n\to\infty} \boldsymbol{A}^{n-(r-i)}(c_{r-1}\delta_1 + \delta_r) \\
&= \underbrace{\lim_{n\to\infty} \boldsymbol{A}^{n-(r-i)} c_{r-1}\delta_1}_{\longrightarrow 0} + \underbrace{\lim_{n\to\infty} \boldsymbol{A}^{n-(r-i)}\delta_r}_{\longrightarrow 0} \\
&= 0.
\end{aligned}
$$

The last equality of course holds since we have already shown that (2.13) holds for $i = 1$ and $i = r$. Therefore (2.13) holds for $i = 1, \ldots, r$ and the "only if" part of the theorem is proven.
The only thing left to show is the uniqueness. Assume there is another strictly stationary solution $\{\boldsymbol{U}_t\}$ that satisfies Equation (2.4). Then, $\boldsymbol{U}_0$ will satisfy Equation (2.11) only that $\boldsymbol{Y}_{-n}$ is swapped with $\boldsymbol{U}_{-n}$. Hence,

$$
\boldsymbol{Y}_0 - \boldsymbol{U}_0 = \left(\prod_{j=0}^{n-1} \boldsymbol{A}_j \right) (\boldsymbol{Y}_{-n} - \boldsymbol{U}_{-n}). \tag{2.14}
$$

Notice if $\lim_{n\to\infty} \boldsymbol{A}_0\boldsymbol{A}_{-1}\cdots\boldsymbol{A}_{-n} = 0$ a.s., then (2.14) converges in probability and the uniqueness immediately follows.
Note that all the coefficients of $\boldsymbol{A}_t$, $\boldsymbol{Y}_t$, and $\boldsymbol{B}_t$ ($t \in \mathbb{Z}$) are nonnegative and therefore for any $n \in \mathbb{N}$

$$
\sum_{k=1}^{n-1} \boldsymbol{A}_0 \cdots \boldsymbol{A}_{-k}\boldsymbol{B}_{-k} \leq \boldsymbol{Y}_0
$$

$$
\begin{aligned}
&\Rightarrow \sum_{k=1}^{n} \boldsymbol{A}_0 \cdots \boldsymbol{A}_{-k}\boldsymbol{B}_{-k} < \infty \\
&\Rightarrow \lim_{n\to\infty} \boldsymbol{A}_0 \cdots \boldsymbol{A}_{-n}\boldsymbol{B}_{-n} = 0, \quad a.s.
\end{aligned}
$$

Now, $\lim_{n\to\infty} \boldsymbol{A}_0\boldsymbol{A}_{-1}\cdots\boldsymbol{A}_{-n} = 0$ follows completely analogous to the lines starting from Equation (2.13). This completes the proof. □

2.2 Properties of the Stationary GJR-GARCH Model

The following two lemmas state some important properties of the (stationary) GJR-GARCH model which will be used in Section 3.

Corollary 2.3.
Consider the stationary GJR-GARCH(p, q) model. Then the unconditional variance of ϵ_t is given by

$$\mathrm{Var}[\epsilon_t] = \mathbb{E}[\epsilon_t^2] = \frac{\alpha_0}{1 - \sum\limits_{i=1}^{q}(\alpha_i + \lambda_i/2) - \sum\limits_{j=1}^{p}\beta_j}.$$

Proof. For the proof I make use of Equation (2.3), since $\mathbb{E}[\epsilon_t^2] = \mathbb{E}[z_t^2\sigma_t^2] = \mathbb{E}[\sigma_t^2]$.
We have

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{r} c_k(z_{t-i})\sigma_{t-i}^2 \\
&= \alpha_0 + \sum_{j=1}^{r} c_j(z_{t-j}) \left(\alpha_0 + \sum_{i=1}^{r} c_i(z_{t-i-j})\sigma_{t-i-j}^2\right) \\
&= \alpha_0 + \sum_{j=1}^{r} c_j(z_{t-j}) \left(\alpha_0 + \sum_{i=1}^{r} c_i(z_{t-i-j}) \left(\alpha_0 + \sum_{l=1}^{r} c_k(z_{t-i-j-l})\sigma_{t-i-j-l}^2\right)\right) \\
&\vdots \\
&= \alpha_0 \sum_{k=0} M(t,k),
\end{aligned}$$

where

$$\begin{aligned}
M(t,0) &= 1 \\
M(t,1) &= \sum_{i=1}^{r} c_i(z_{t-i}) \\
M(t,2) &= \sum_{j=1}^{r} c_j(z_{t-j}) \left(\sum_{i=1}^{r} c_i(z_{t-i-j})\right) \\
&\vdots \\
M(t,k+1) &= \sum_{i=1}^{r} c_i(z_{t-i}) M(t-i,k) \qquad (2.15)
\end{aligned}$$

Since z_t^2 is i.i.d., the moments of $M(t,k)$ do not depend on t. In particular

$$\mathbb{E}[M(t,k)] = \mathbb{E}[M(s,k)] \ \ \forall k,t,s. \tag{2.16}$$

From (2.15) and (2.16) we get

$$\begin{aligned}
\mathbb{E}[M(t,k+1)] &= \sum_{i=1}^{r} \mathbb{E}[c_i(z_{t-i})]\mathbb{E}[M(t,k)] \\
&\ \vdots \\
&= \left(\sum_{i=1}^{r} \mathbb{E}[c_i(z_{t-i})]\right)^{k+1} \underbrace{\mathbb{E}[M(t,0)]}_{=1} \\
&= \left(\sum_{i=1}^{r} \mathbb{E}[c_i(z_{t-i})]\right)^{k+1}
\end{aligned}$$

As shown in (2.7), we have

$$\begin{aligned}
\mathbb{E}[c_i(z_{t-i})] &= \mathbb{E}[\beta_i + (\alpha_i + \lambda_i I(z_{t-i} < 0))z_{t-i}^2] \\
&= \beta_i + (\alpha_i + \lambda_i \mathbb{E}[I(z_{t-i} < 0)])\underbrace{\mathbb{E}[z_{t-i}^2]}_{=1} \\
&= \beta_i + \alpha_i + \lambda_i/2.
\end{aligned}$$

This gives us

$$\mathbb{E}[M(t,k+1)] = \left(\sum_{i=1}^{r} \beta_i + \alpha_i + \lambda_i/2\right)^{k+1}.$$

Finally, we get for our unconditional variance

$$\begin{aligned}
\mathbb{E}[\epsilon_t^2] &= \alpha_0 \mathbb{E}[\sum_{k=0}^{\infty} M(t,k)] \\
&= \alpha_0 \sum_{k=0}^{\infty} \mathbb{E}[M(t,k)] \\
&= \alpha_0 \sum_{k=0}^{\infty} \left(\sum_{i=1}^{r} \beta_i + \alpha_i + \lambda_i/2\right)^{k} \\
&= \alpha_0 \sum_{k=0}^{\infty} \left(\sum_{i=1}^{q} (\alpha_i + \lambda_i/2) + \sum_{j=1}^{p} \beta_j\right)^{k},
\end{aligned}$$

which is equal to

$$\frac{\alpha_0}{1 - \sum\limits_{i=1}^{q} (\alpha_i + \lambda_i/2) - \sum\limits_{j=1}^{p} \beta_j}$$

if and only if

$$\sum_{i=1}^{q} (\alpha_i + \lambda_i/2) + \sum_{j=1}^{p} \beta_j < 1. \tag{2.17}$$

(2.17) is of course satisfied, since the GJR-GARCH model is stationary (see Theorem 2.2). □

The following lemma states another important property of the stationary GJR-GARCH model. This property will be needed in Section 3 to prove asymptotic normality of a (quasi-)maximum-likelihood-estimator.

Lemma 2.4. *Consider the stationary GJR-GARCH (p, q) model. Then there exists an $\eta > 0$ such that $\mathbb{E}|\sigma_0|^{2\eta} < \infty$.*

Proof. Since the stationarity of the GJR-GARCH model implies that the SRE (2.4) is stationary, we will prove this lemma by showing that there exists an $\eta > 0$ such that

$$\mathbb{E}|\boldsymbol{Y}_0|^{\eta} < \infty.$$

To do so we will exploit the fact that if (2.4) admits a unique stationary solution, then this is equivalent to $\{\boldsymbol{A}_t\}$ having a negative top Lyapunov exponent, i.e.,

$$\rho = \inf_{t \in \mathbb{N}} \{ \frac{1}{t+1} \mathbb{E}[\log \|\boldsymbol{A}_0 \cdots \boldsymbol{A}_{-t}\|_{op}\} < 0.^7 \tag{2.18}$$

For details on this result, see for example Liu (2009) or Bougerol and Picard (1992). Now, Equation (2.18) entails that there is a $r \geq 1$ with

$$\mathbb{E}[\log \|\boldsymbol{A}_0 \cdots \boldsymbol{A}_{-r+1}\|_{op}] = \mathbb{E}[\log \|\boldsymbol{A}_0^{(r)}\|_{op}] < 0.$$

Since every matrix is equivalent to the Frobenius norm and $\mathbb{E}[z_0^2] = 1$ we get $\mathbb{E}\|\boldsymbol{A}_0\|_{op} < \infty$ and therefore also $\mathbb{E}\|\boldsymbol{A}_0\|_{op}^{\tilde{\eta}} < \infty$ and $\mathbb{E}\|\boldsymbol{A}_0^{(r)}\|_{op}^{\tilde{\eta}} < \infty$ for some $\tilde{\eta} \in (0, 1]$.[8] Notice that the first

[7] $\|.\|_{op}$ denotes the matrix operator norm with respect to the Euclidean norm on $\mathbb{R}^{p+q-1}$. See Section 3.3 for a definition.

[8] Let $A \in \mathbb{R}^{m \times n}$ be a real-valued matrix. The Frobenius norm $\|.\|_F$ of A is defined as $\|A\|_F := \sqrt{\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} |a_{ij}|^2}$.

derivative of the map $s \mapsto \mathbb{E}\|\boldsymbol{A}_0^{(r)}\|_{op}^s$ on $[0,\tilde{\eta}]$ at $s=0$ is equal to $\mathbb{E}[\log\|\boldsymbol{A}_0^{(r)}\|_{op}] < 0$. This proves the existence of $\eta \in (0,\tilde{\eta}]$ with $\lambda := \mathbb{E}[\|\boldsymbol{A}_0^{(r)}\|_{op}^{\eta} < 1$. Because the $\{\boldsymbol{A}_t\}$ are i.i.d. we get

$$\mathbb{E}\left[\left(\prod_{\ell=1}^{m}\|\boldsymbol{A}_{(1-\ell)r}^{(r)}\|_{op}\right)^{\eta}\right] = \mathbb{E}[\|\boldsymbol{A}_0^{(r)}\|_{op}^{\eta}]^m = \lambda^m, \quad m \geq 1. \tag{2.19}$$

(2.19) together with the observation that $\|.\|_{op}$ is submultiplicative yields

$$\mathbb{E}\|\boldsymbol{A}_0^{(k)}\|_{op}^{\eta} \leq \left(\mathbb{E}\|\boldsymbol{A}_0^{(k)}\|_{op}^{\eta}\right)^{(k/r)}\left(\mathbb{E}\|\boldsymbol{A}_0^{(k)}\|_{op}^{\eta}\right)^{r-r(k/r)} \leq c\lambda^{(k/r)}, \quad k \geq 0,$$

where $c = \max(1, (\mathbb{E}\|\boldsymbol{A}_0\|_{op}^{\eta})^r)$. Finally, we will apply the Minowski inequality to the stationary solution (2.6), and get

$$\mathbb{E}\|\boldsymbol{Y}_1\|_{op}^{\eta} \leq \sum_{k=0}^{\infty}\mathbb{E}\|\boldsymbol{A}_0^{(k)}\|_{op}^{\eta}|\boldsymbol{B}|^{\eta} \leq c\alpha_0^{\eta}\sum_{k=0}^{\infty}\lambda^{(k/r)} < \infty.$$

The latter equation of course implies $\mathbb{E}|\boldsymbol{Y}_0|^{\eta}$ and this entails

$$\mathbb{E}|\sigma_0|^{2\eta} < \infty.$$

This completes the proof. $\square$

3 Volatilty by Variance Decomposition

This section introduces the main model discussed in this thesis developed by Amado and Teräsvirta (2013) and is organized as follows. 3.1 introduces the TV-GJR-GARCH model, 3.2 deals with the estimation of the model and states theorems for consistency and asymptotic normality of the (quasi)-maximum-likelihood estimators. Section 3.4 then provides an extensive proof for one of the theorems whereas important preliminaries for the proof are stated in Section 3.3.

3.1 The TV-GJR-GARCH Model

As in Section 1 let X_t denote the log-returns of an asset represented by the sum of a predictable part and an unpredictable part

$$X_t = \mu_t + \epsilon_t, \tag{3.1}$$

with

$$\begin{aligned} \mathbb{E}[\epsilon_t \mid \mathcal{F}_{t-1}] &= 0, \\ \mathrm{Var}[X_t \mid \mathcal{F}_{t-1}] = \mathbb{E}[\epsilon_t^2 \mid \mathcal{F}_{t-1}] &= \sigma_t^2, \text{ and} \\ \mathbb{E}[y_t \mid \mathcal{F}_{t-1}] &= 0. \end{aligned}$$

The innovation sequence $(\epsilon_t : t \in \mathbb{Z})$ is defined as

$$\epsilon_t = z_t \sigma_t, \tag{3.2}$$

where $(z_t : t \in \mathbb{Z})$ is an i.i.d. sequence of normally distributed random variables with $\mathbb{E}[z_t] = 0$ and $\mathrm{Var}[z_t] = 1$, analogous to the previous described models. The difference is that we assume σ_t^2 to be time-varying and measurable with respect to $\mathcal{F}_{t-1}$ and splittable into a multiplicative decomposition [9]

$$\sigma_t^2 = h_t g_t. \tag{3.3}$$

The function h_t is responsible for describing the heteroscedastic effects in X_t, while g_t is the component that introduces nonstationarity. As mentioned in Section 2 asset returns usually respond differently to negative and positive shocks. To incorporate this, h_t will follow the

[9] As mentioned in Footnote 4 Amado and Teräsvirta (2013) also introduce an additive decomposition, i.e. $\sigma_t^2 = h_t + g_t$. However, this thesis will solely focus on the multiplicative model.

stationary GJR-GARCH(p, q) model described in Definition 2.1, i.e.:

$$h_t = \alpha_0 + \sum_{i=1}^{q} (\alpha_i + \lambda_i I(\epsilon_{t-i} < 0))\epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}.^{10} \tag{3.4}$$

Note that when (3.3) holds, ϵ_{t-i}^2 is replaced by $\phi_{t-i}^2 = \epsilon_{t-i}^2 / g_{t-i}$, $i = 1, \ldots, q$ in (3.4). Like in for example Kılıç (2011), Enders and Holt (2014), or Teterin et al. (2016), Amado and Teräsvirta (2013) believe that it is more reasonable to model the structural changes as being smooth rather than sharp. This is done by defining the nonstationary component g_t as follows:

$$g_t = 1 + \delta G(t/T; \gamma, \boldsymbol{c}), \tag{3.5}$$

where δ is a proportionality factor,[11] and $G(t/T; \gamma, \boldsymbol{c})$ is a continuous and nonnegative function bounded by zero and one, the so-called transition function. Note that the time t/T is the transition variable and is defined on the interval $[0, 1]$, where T is the number of observations. Amado and Teräsvirta (2013) find the general logistic function

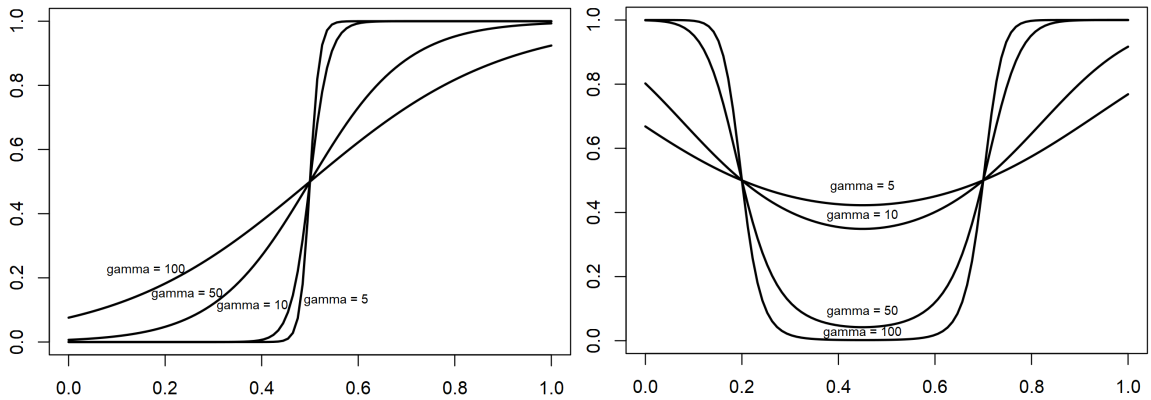
$$\left(1 + \exp\left(-\gamma \prod_{k=1}^{M} (t/T - c_k)\right)\right)^{-1}, \quad \gamma > 0, \ c_1 \leq \ldots \leq c_k \tag{3.6}$$

to be a suitable choice for $G(t/T; \gamma, \boldsymbol{c})$, $\boldsymbol{c} = c_1, \ldots, c_k$. This function lets the parameters of the GJR-GARCH model (3.2), (3.3), and (3.4) fluctuate smoothly over time between $(\alpha_i, \lambda_i, \beta_j)$ and $(\alpha_i + \delta\alpha_i, \gamma_i + \delta\gamma_i, \beta_j + \delta\beta_j)$, $i = 0, 1, \ldots, q$, $j = 1, \ldots, p$. The order $M \in \mathbb{Z}_+$ is responsible for the shape of the transition function. In practice, M is usually set to $M = 1$ or $M = 2$. The parameter γ controls the degree of smoothness of the transition function. As γ approaches infinity the process is characterized by abrupt structural breaks at $c_1, \ldots, c_M$, meaning the switch from one set of parameters to another is instantaneous. In other words, the larger γ the faster the transition of 0 to 1 as a function of t/T. This can be seen in Figure 2, where the transition functions for $M = 1$, $M = 2$ and a several values for γ, c_1, and c_2 are illustrated. In general, one can see that a TV-GJR-GARCH model with $M = 1$ should be used if the volatility dynamics are different before and after a smooth structural change. $M = 2$ should be used if the parameters change but eventually return towards their original values.

[10] Spot the notational predicament here. I opted to denote the (total) squared volatility of each model in question with σ_t^2 throughout this thesis. However, in the TV-GJR-GARCH model the squared volatility of the GJR-GARCH model only makes up one component of the total squared volatility. Therefore, I cannot denote the squared volatility of the GJR-GARCH model with σ_t^2 as in Section 2. Nevertheless, to be consistent with the literature I designate it with h_t.

[11] We will see later that δ has to suffice certain restrictions to ensure $g_t > 0$ for all t. Details are discussed below for the more general version (Equation (3.7))

Figure 2: `Transition Functions for` $M = 1, 2$



Notes: The left plot shows the general logistic transition function defined in (3.6) for $M = 1$, $c_1 = 0.5$, and $\gamma = 5$, 10, 50, and 100. The right plot shows the general logistic transition function defined in (3.6) for $M = 2$, $c_1 = 0.2$, $c_2 = 0.7$, and $\gamma = 5$, 10, 50, and 100.

A more general and extended version of the multiplicative TV-GJR-GARCH model can defined by permitting more than one transition function in g_t,

$$g_t = 1 + \sum_{l=1}^{r} \delta_l G_l(t/T; \gamma_l, \boldsymbol{c}_l). \tag{3.7}$$

Here, $G_l(t/T; \gamma_l, \boldsymbol{c}_l)$, $l = 1, \ldots, r$ are logistic functions as in (3.6) with smoothness parameter γ_l and threshold parameter vector $\boldsymbol{c}_l$. The parameter vectors in (3.4) and (3.7) satisfy the restrictions $\alpha_0 + \sum_{l=1}^{j} \alpha_0 \delta_l > 0$, $\alpha_i + \lambda_i/2 + \sum_{l=1}^{j} (\alpha_i + \lambda_i/2)\delta_l > 0$, $i = 1, \ldots, q$, and $\beta_i + \sum_{l=1}^{j} \beta_i \delta_l \geq 0$, $i = 1, \ldots, p$, all for any $j \in \{1, \ldots, r\}$. These conditions ensure $g_t > 0$ for all t.
(3.3) can now be rewritten as

$$\sigma_t^2 = h_t \left(1 + \sum_{l=1}^{r} \delta_l G_l(t/T; \gamma_l, \boldsymbol{c}_l) \right). \tag{3.8}$$

The model (3.2), (3.3), (3.4), (3.7) has a straightforward interpretation, as (3.2) and (3.3) can be written as

$$\phi_t = \epsilon_t / g_t^{1/2} = z h_t^{1/2}, \quad t = 1, \ldots, T \tag{3.9}$$

3.2 Estimation

For the estimation of the TV-GJR-GARCH model we estimate the parameters vectors

$$\boldsymbol{\theta}_1 = (\boldsymbol{\delta}^t, \boldsymbol{\gamma}^t, \boldsymbol{c}^t)^t \text{ and } \boldsymbol{\theta}_2 = (\alpha_0, \boldsymbol{\alpha}^t, \boldsymbol{\lambda}^t, \boldsymbol{\beta}^t)^t$$

using a quasi-maximum likelihood approach. In $\boldsymbol{\theta}_1$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_r)^t, \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)^t, \gamma_l > 0, l = 1, \ldots, r$, and $\boldsymbol{c} = (c_1, \ldots, c_r)^t$, $c_1 < \ldots < c_r$.[12] In $\boldsymbol{\theta}_2$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)^t$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)^t$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$. Then the functions h_t and g_t can be written as:

$$h_t = h_t(\boldsymbol{\theta_1}, \boldsymbol{\theta}_2) \text{ and } g_t = g(\boldsymbol{\theta}_1, t/T)$$

When we perform a maximum likelihood estimation, we are interested in the joint distribution $f(\epsilon_1, \ldots, \epsilon_T; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. At this point it is crucial to notice that since $z_t \sim \mathcal{N}(0,1)$, we have $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$ conditional on the information set at time t. We re-express $f(\epsilon_0, \ldots, \epsilon_T; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ as a likelihood function $\mathcal{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \epsilon_0, \ldots, \epsilon_T)$ defined as:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \epsilon_0, \ldots, \epsilon_T) &= f(\epsilon_0, \ldots, \epsilon_T; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&= f(\epsilon_0; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) f(\epsilon_1, \ldots, \epsilon_T \mid \epsilon_0; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&= f(\epsilon_0; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \prod_{t=1}^{T} f(\epsilon_t \mid \epsilon_{t-1}, \ldots, \epsilon_0; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&= f(\epsilon_0; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \prod_{t=1}^{T} f(\epsilon_t \mid \epsilon_{t-1}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&= f(\epsilon_0; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma_t^2}\right) \\
&= f(\epsilon_0; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) g(\boldsymbol{\theta}_1, t/T))}} \exp\left(-\frac{\epsilon_t^2}{2h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) g(\boldsymbol{\theta}_1, t/T)}\right)
\end{aligned}$$

For large sample sizes the impact of $f(\epsilon_0; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is usually assumed to be relatively small and therefore dropped. Defining $\boldsymbol{\epsilon} = \epsilon_1, \ldots, \epsilon_T$ and taking the log of $\mathcal{L}$, we obtain the conditional

[12] Note that for ease of notation we assume here that each transition function $G_l(t/T; \gamma_l, \boldsymbol{c}_l)$ in $g(\boldsymbol{\theta}_1, t/T)$ is a simple logistic function (meaning $M = 1$). This means that the c_l's are just scalars and the bold writing is not necessary.

(quasi-)log-likelihood function

$$\begin{aligned}
L_T(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \boldsymbol{\epsilon}) &= \sum_{t=1}^{T} \left(\log(1) - \log\left((2\pi h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) g(\boldsymbol{\theta}_1, t/T))^{1/2} \right) - \frac{1}{2} \frac{\epsilon_t^2}{h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) g(\boldsymbol{\theta}_1, t/T)} \right) \\
&= \sum_{t=1}^{T} \left(-\frac{1}{2} (\log(2\pi) + \log(h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) g(\boldsymbol{\theta}_1, t/T))) - \frac{1}{2} \frac{\epsilon_t^2}{h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) g(\boldsymbol{\theta}_1, t/T)} \right) \qquad (3.10) \\
&= -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \left(\log(h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)) + \log(g(\boldsymbol{\theta}_1, t/T)) + \frac{\epsilon_t^2}{h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) g(\boldsymbol{\theta}_1, t/T)} \right).
\end{aligned}$$

Usually the maximum likelihood estimator $(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)$ of $(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0)$ is calculated by maximizing $L_T(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \boldsymbol{\epsilon})$,

$$(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) = \arg \max_{(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)} L_T(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \boldsymbol{\epsilon})$$

However, as Amado and Teräsvirta (2013) point out, a straightforward maximization of L_T is numerically very difficult. Therefore the aforementioned authors use a method called *maximization by parts* to find a numerically tractable solution. The approach in the general case is extensively discussed in Song et al. (2005) or Fan et al. (2007).

In the case of (3.10), estimation proceeds as follows:

1. Estimate $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ by constraining $\boldsymbol{\theta}_2 = (\alpha_0, \mathbf{0}^t, \mathbf{0}^t, \mathbf{0}^t)^t$, obtain $\boldsymbol{\theta}_1^{(1)}$ and $\boldsymbol{\theta}_2^{(0)} = (\widehat{\alpha}_0^{(0)}, \mathbf{0}^t, \mathbf{0}^t, \mathbf{0}^t)^t$. Using the constraint on $\boldsymbol{\theta}_2$, one can make use of the following manipulations. Rewriting (3.8) yields

$$h_t \left(1 + \sum_{l=1}^{r} \delta_l G_l(t/T; \gamma_l, c_l) \right) = \alpha_0 \left(1 + \sum_{l=1}^{r} \delta_l G_l(t/T; \gamma_l, c_l) \right) \qquad (3.11)$$

$$= \alpha_0 + \sum_{l=1}^{r} \underbrace{\alpha_0 \delta_l}_{=:\delta_l^*} G_l(t/T; \gamma_l, c_l) \qquad (3.12)$$

where $\alpha_0 > 0$ in (3.10). Therefore, constraining $\boldsymbol{\theta}_2 = (\alpha_0, \mathbf{0}^t, \mathbf{0}^t, \mathbf{0}^t)^t$ makes it possible to set $h_t \equiv 1$ and define

$$g(\boldsymbol{\theta}_1, t/T) = \alpha_0 + \sum_{l=1}^{r} \delta_l^* G_l(t/T; \gamma_l, c_l). \qquad (3.13)$$

Then, we estimate $\alpha_0, \delta_1^*, \ldots, \delta_r^*, \boldsymbol{\gamma}$, and $\boldsymbol{c}$ to obtain the estimates $\widehat{\alpha}_0^{(0)}$, $\boldsymbol{\delta}^{*(1)}$, $\boldsymbol{\gamma}^{(1)}$, and

$\boldsymbol{c}^{(1)}$. Eventually, the estimate $\boldsymbol{\delta}^{(1)}$ is obtained as follows: $\delta_l^{(1)} = \delta_l^{*(1)}/\alpha_0^{(0)}$, $l = 1, \ldots, r$.

2. Estimate $\boldsymbol{\theta}_2$ including α_0, setting $g(\boldsymbol{\theta}_1, t/T) = g_t(\boldsymbol{\theta}_1^{(1)}, t/T)$.

3. Estimate $\boldsymbol{\theta}_1$ assuming

$$g(\boldsymbol{\theta}_1, t/T) = 1 + \sum_{l=1}^{r} \delta_l G_l(t/T; \gamma_l, c_l) \tag{3.14}$$

while setting $h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = h_t(\boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_2^{(1)})$. This entails $\boldsymbol{\theta}_1^{(2)}$. The imperative detail here is that the parameter estimates in $h_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are held unchanged when $\boldsymbol{\theta}_1$ is re-estimated.

Steps 2 and 3 are repeated until convergence.
Now, we have to show that the obtained estimates are consistent and asymptotically normal. To prove this, Amado and Teräsvirta (2013) make the following assumptions:

AG1 The parameter space $\Theta_1 = \{\alpha_0 \times \Delta \times \Gamma \times C\}$ is compact, where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_r)^t \in \Delta$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)^t \in \Gamma$, and $\boldsymbol{c} = (c_1, \ldots, c_r)^t \in C$.

AG2 The log-likelihood function L_T has a unique maximum at the true parameter $\boldsymbol{\theta}_1^0$ which is an interior point of Θ_1.

AG3 The elements of $\boldsymbol{\delta} \in \Delta$ are restricted such that $\max_{j=0,1,\ldots,q} |\delta_j| \leq M_\delta < \infty$, and $\inf_{\boldsymbol{\theta}_1 \in \boldsymbol{\theta}_1} g(\boldsymbol{\theta}_1, t/T) \geq g_{\min} > 0$ for all t. Furthermore $\alpha_0 > 0$.

AG4 The slope parameters γ_l, $l = 1, \ldots, r$, are strictly positive, i.e. $\gamma_l > 0$ for all l. The location parameters $c_1, \ldots, c_r$ follow an ascending pattern, i.e. $c_1 < \ldots < c_r$.

AG5 In (3.2), $\{z_t\}$ is a sequence of independent identically normally distributed random variables with mean zero and variance 1, in addition, $\mathbb{E}|z_t|^{2(2+\phi)} < \infty$ for some $\phi > 0$.

AG6 $\int \sup_{\boldsymbol{\theta}_1 \in \mathcal{N}} \|\frac{\partial}{\partial \boldsymbol{\theta}_1} f(z_t \mid \boldsymbol{\theta}_1)\| dz_t < \infty$, and $\int \sup_{\boldsymbol{\theta}_1 \in \mathcal{N}} \|\frac{\partial^2}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^t} f(z_t \mid \boldsymbol{\theta}_1)\| dz_t < \infty$, where $f(z_t \mid \boldsymbol{\theta}_1)$ is the density of z_t, $\|.\|$ is the Euclidean norm and $\mathcal{N}$ a neighborhood of $\boldsymbol{\theta}_1^0$.

The proof that maximization by parts leads to consistent and asymptotically normal estimates is split into two theorems (Theorem 3.1 and 3.2). The first theorem is according to Amado and Teräsvirta (2013, Theorem 1) and shows that given $h_t \equiv 1$, the estimator of $\boldsymbol{\theta}_1$ is consistent and asymptotically normal. The just mentioned authors provide an extensive proof for this theorem in their appendix. That is why a proof of this theorem is omitted in this thesis. The second theorem is according to Amado and Teräsvirta (2013, Theorem 2) and yields the result that given g_t is known, the estimator of $\boldsymbol{\theta}_2$ is consistent and asymptotically normal. For this a rigorous proof will be supplied (see Section 3.4).

In the following theorem, I will make use of the following notation:

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}_1} g(\boldsymbol{\theta}_1^0, t/T) &= \left. \frac{\partial}{\partial \boldsymbol{\theta}_1} g(\boldsymbol{\theta}_1, t/T) \right|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0} \quad \text{and} \\
\frac{\partial^2}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^t} g(\boldsymbol{\theta}_1^0, t/T) &= \left. \frac{\partial^2}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^t} g(\boldsymbol{\theta}_1, t/T) \right|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0}
\end{aligned}
$$

Theorem 3.1.
Let $\widehat{\boldsymbol{\theta}}_1$ *be the maximum likelihood estimator of* $\boldsymbol{\theta}_1^0$,

$$\widehat{\boldsymbol{\theta}}_1 = \arg\max L_T(\boldsymbol{\theta}_1, \boldsymbol{\epsilon})$$

where the quasi-log-likelihood function for the model (3.2), (3.3), (3.14) is

$$L_T(\boldsymbol{\theta}_1, \boldsymbol{\epsilon}) = \sum_{t=1}^{T} \ell(\boldsymbol{\theta}_1, \epsilon_t) \tag{3.15}$$

with

$$\ell(\boldsymbol{\theta}_1, \epsilon_t) = -\frac{T}{2} \log(2\pi) - (1/2) \left(\log g(\sigma_1, t/T) + \frac{\epsilon_t^2}{g(\sigma_1, t/T)} \right), \tag{3.16}$$

Assume that the assumptions AG1-6 hold with $z_t = \frac{\epsilon_t}{g(\boldsymbol{\theta}_1^0, t/T)^{1/2}}$ *in AG5, i.e.,* $h_t \equiv 1$ *in (3.3). Then*

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{A}^{-1}(\boldsymbol{\theta}_1^0) \boldsymbol{B}(\boldsymbol{\theta}_1^0) \boldsymbol{A}^{-1}(\boldsymbol{\theta}_1^0))$$

where

$$
\begin{aligned}
A(\boldsymbol{\theta}_1^0) &= -\frac{1}{2} \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \frac{1}{g(\boldsymbol{\theta}_1^0, t/T)^2} \frac{\partial}{\partial \boldsymbol{\theta}_1} g(\boldsymbol{\theta}_1^0, t/T) \left(\frac{\partial}{\partial \boldsymbol{\theta}_1} g(\boldsymbol{\theta}_1^0, t/T) \right)^t \\
&= -\frac{1}{2} \int_0^1 \frac{\partial}{\partial \boldsymbol{\theta}_1} \log g(\boldsymbol{\theta}_1^0, r) \left(\frac{\partial}{\partial \boldsymbol{\theta}_1} \log g(\boldsymbol{\theta}_1^0, r) \right)^t dr
\end{aligned}
$$

and

$$B(\boldsymbol{\theta}_1^0) = \mathbb{E} \left(\frac{z_t^2 - 1}{2} \right)^2 \int_0^1 \frac{\partial}{\partial \boldsymbol{\theta}_1} \log g(\boldsymbol{\theta}_1^0, r) \left(\frac{\partial}{\partial \boldsymbol{\theta}_1} \log g(\boldsymbol{\theta}_1^0, r) \right)^t dr$$

when $T \to \infty$.

Proof. See Appendix of Amado and Teräsvirta (2013). □

Assuming Theorem 3.1 holds true, to show that both $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_2$ are consistent and asymptotically normal, we still have to prove that the estimator $\widehat{\boldsymbol{\theta}}_2$ of the parameter vector $\boldsymbol{\theta}_2$ of the model

$$\phi_t = z_t h_t^{1/2}, \tag{3.17}$$

given $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$, is consistent and asymptotically normal. This will be the subject of Theorem 3.2 below. Since Amado and Teräsvirta (2013) did not provide a proof for this theorem and rather just referred to other papers, this thesis will bring the mentioned literature together and provide a rigorous proof. Straumann et al. (2006) show the consistency and asymptotic normality for the (quasi-)maximum-likelihood estimator of the AGARCH model (see Equation (4.1) in the appendix for a definition). Their work serves as the groundwork for the discussed proof of Theorem 3.2 in this thesis (see Section 3.4).
We make the following assumptions in order to build on the work of Straumann et al. (2006).

AH1 In (3.4), α_0, and $\sum_{i=1}^{q}(\alpha_i + \lambda_i/2) + \sum_{j=1}^{p} \beta_j < 1$, where $\lambda_i = \kappa\alpha_i$, $i = 1, \dots, q$.[13]

AH2 The polynomials $\sum_{i=1}^{q}(1 + \kappa/2)\alpha_i z^i$ and $1 - \sum_{j=1}^{p} \beta_j z^j$ do not have common roots.

AH3 The parameter space $\Theta_2 = \{\alpha_0 \times A \times \kappa \times B\}$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_1)^t \in A$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t \in B$, is compact and the true parameter value $\boldsymbol{\theta}_2^0$ belongs to the interior of $\boldsymbol{\theta}_2$.

AH4 The distribution of z_0 is not concentrated at two points.

It is important to mention that since $\lambda_i = \kappa\alpha_i$, we also have that $\boldsymbol{\theta}_2 = (\alpha_0, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p, \kappa)$ and the GJR-GARCH model takes the following representation

$$h_t = \alpha_0 + \sum_{i=1}^{q}(1 + \kappa I(\phi_{t-i} < 0))\alpha_i \phi_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3.18}$$

Theorem 3.2.
Consider the GJR-GARCH model (3.17) with (3.18) and assume that the assumptions AG5 and AH1-4 hold. Furthermore, assume that $\phi_t = \epsilon_t / g_t^{1/2}$, $t = 1, \dots, T$, where g_t is given in (3.13). Then the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_2$ of $\boldsymbol{\theta}_2^0$, given $\boldsymbol{\theta}_1^0$, is consistent and

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2^0\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{V}_2^0)$$

[13]The exact definition of κ is given in Theorem 4.3 in the appendix, where it is shown that the AGARCH model can also be written in the GJR-GARCH form.

as $T \to \infty$, and where $\boldsymbol{V}_2^0$ is the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}_2$ given $\boldsymbol{\theta}_1^0$.

Before I can begin with the actual proof, firstly some notation, definitions, propositions and lemmas need to be introduced that will be used in the proof of the theorem, i.e. in Section 3.4. These preliminaries will be given in the next subsection.

3.3 Preliminaries for the Proof of Theorem 3.2

For the remainder of this thesis, we will write the squared volatility h_t of the GJR-GARCH model as such:

$$h_t = g_{\boldsymbol{\theta}_2}(\boldsymbol{\phi}_{t-1}, \boldsymbol{h}_{t-1}), \quad t \in \mathbb{Z}, \boldsymbol{\theta}_2 \in \boldsymbol{\theta}_2,$$

where

$$\boldsymbol{\phi}_t = (\phi_t, \ldots, \phi_{t-q+1})^t \text{ and } \boldsymbol{h}_t = (h_t, \ldots, h_{t-p+1})^t.$$

We will assume that $h_t^{1/2}$ is nonnegative and $\mathcal{F}_{t-1}$-measurable, where

$$\mathcal{F}_t = \sigma(z_k; k \leq t), \quad t \in \mathbb{Z}$$

denotes the σ-field generated by the random variables $\{z_k; k \leq t\}$.

Definition 3.3. (r-fold transformations)
For a sequence $\{\varphi_t\}$ ($t \in \mathbb{Z}$) of transformations on a certain space E, we denote by $\{\varphi_t^{(r)}\}$ the sequence of the r-fold iterations of past and present transformations defined by

$$\varphi_t^{(r)} = \begin{cases} Id_E, & r = 0, \\ \varphi_t \circ \varphi_{t-1} \circ \cdots \circ \varphi_{t-r+1}, & r \geq 1, \end{cases}$$

where Id_E denotes the identity map in E.

In the following we will assume that $K \subset \mathbb{R}^d$ is a compact set and write $\mathbb{C}(K, \mathbb{R}^{d'})$ for the space of continuous $\mathbb{R}^{d'}$-valued functions equipped with the sup-norm $\|v\|_K = \sup_{\boldsymbol{s} \in K} |v(\boldsymbol{s})|$.

Definition 3.4. (operator norm)
Let $\boldsymbol{A} = (a_{ij}) \in \mathbb{R}^{d' \times d'}$ be a matrix. The operator norm of $\boldsymbol{A}$ with respect to the Euclidean norm is defined as

$$\|\boldsymbol{A}\|_{op} = \sup_{\boldsymbol{x} \neq 0} \frac{|\boldsymbol{A}\boldsymbol{x}|}{|\boldsymbol{x}|}$$

Definition 3.5.
Let $\boldsymbol{A} = (a_{ij}) \in \mathbb{R}^{d' \times d'}$ be a matrix. The norm of a continuous matrix-valued function $\boldsymbol{A}$ on a compact set $K \subset \mathbb{R}^d$, that is, $\boldsymbol{A} \in \mathbb{C}(K, \mathbb{R}^{d' \times d'})$, is given by

$$\|\boldsymbol{A}\|_K = \sup_{\boldsymbol{s} \in K} \|\boldsymbol{A}(\boldsymbol{s})\|$$

Finally, let (E, d) be a Polish space equipped with its Borel σ- field $\mathcal{B}(E)$.[14] For a map $\varphi : E \to E$, we define

$$\Lambda(\varphi) := \sup_{x,y \in E, x \neq y} \left(\frac{d(\varphi(x), \varphi(y))}{d(x,y)} \right).$$

Assume that $\boldsymbol{\theta}_2 \in K$ and that $\{(\phi_t, h_t^{1/2})\}$ is the unique stationary ergodic solution to the model (3.4) with unknown true parameter $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0$. For any initial value $\zeta_0^2 \in [0, \infty)^p$, we define the following random vector function $\widehat{\boldsymbol{f}}_t$ on K:

$$\widehat{\boldsymbol{f}}_t = \begin{cases} \zeta_0^2, & t = 0, \\ \varphi_{t-1}(\widehat{\boldsymbol{f}}_{t-1}), & t \geq 1, \end{cases}$$

where the random maps $\varphi_t : \mathbb{C}(K, [0,\infty)^p) \to \mathbb{C}(K, [0,\infty)^p)$ are given by

$$[\varphi_t(\boldsymbol{s})](\boldsymbol{\theta}_2) = (g_{\boldsymbol{\theta}_2}(\phi_t, \boldsymbol{s}(\boldsymbol{\theta}_2)), s_1(\boldsymbol{\theta}_2), \ldots, s_{p-1}(\boldsymbol{\theta}_2))^t, \quad t \in \mathbb{Z}, \boldsymbol{s} \in [0,\infty)^p.$$

We can regard $\widehat{\boldsymbol{f}}_t(\boldsymbol{\theta}_2) = \left(\widehat{f}_t(\boldsymbol{\theta}_2), \ldots, \widehat{f}_{t-p+1}(\boldsymbol{\theta}_2)\right)^t$ as an "estimate" of the squared volatility vector $\boldsymbol{h}_t$ under the parameter hypothesis $\boldsymbol{\theta}_2$, which is based on the data $\phi_{-q+1}, \ldots, \phi_t$.
We will see that in order to establish consistency of $\widehat{\boldsymbol{\theta}}_2$, it is fundamental that one can approximate $\{\widehat{f}_t\}$ by a stationary ergodic sequence $\{f_t\}$ such that the error $\widehat{f}_t - f_t$ converges to zero sufficiently fast as $t \to \infty$ and such that $\{f_t(\boldsymbol{\theta}_2)\} = \{h_t\}$ almost surely if and only if $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0$. Straumann et al. (2006) state the following proposition, which ensures the existence of such a sequence $\{f_t\}$ under certain conditions.

Proposition 3.6. *Assume that model (3.21) admits a unique stationary ergodic solution* $((\phi_t, h_t^{1/2}))$ *and that the map* $(\boldsymbol{\theta}_2, \boldsymbol{s}) \mapsto g_{\boldsymbol{\theta}_2}(\boldsymbol{x}, \boldsymbol{s})$ *is continuous for every* $\boldsymbol{x} \in \mathbb{R}^q$*, which implies that* $\{\varphi_t\}$ *is a stationary ergodic sequence of mappings* $\mathbb{C}(K, [0,\infty)^p) \to \mathbb{C}(K, [0,\infty)^p)$*. We suppose the following conditions hold:*

1. $\mathbb{E}[\log^+ \|\varphi_0(\zeta_0^2)\|_K] < \infty$

[14] A topological space E is completely metrizable if it admits a compatible metric d such that (X, d) is complete. A separable completely metrizable space is called Polish.

2. $\mathbb{E}[\log^+ \Lambda(\varphi_0)] < \infty$ *and there exists an integer* $r \geq 1$ *such that* $\mathbb{E}[\log \Lambda(\varphi_0^{(r)})] < 0$.

Then the stochastic recurrence equation $\boldsymbol{s}_{t+1} = \varphi_t(\boldsymbol{s}_t)$, $t \in \mathbb{Z}$, $\boldsymbol{s}_t = (s_t, \ldots, s_{t-p+1})^t$, *has a unique stationary solution* $\{\boldsymbol{f}_t\}$, *which is ergodic. For every* $t \in \mathbb{Z}$, *the random elements* $\boldsymbol{f}_t$ *are* $\mathcal{F}_{t-1}$*-measurable and* $\boldsymbol{f}_t(\boldsymbol{\theta}_2^0) = h_t$ *almost surely. Moreover*

$$\|\widehat{\boldsymbol{f}}_t - \boldsymbol{f}_t\|_K \xrightarrow{c.a.s.} 0, \quad t \to \infty.$$

For a proof, see Straumann et al. (2006).

In the proof of Theorem (3.2) we have to work with first- and second-order partial derivatives of a function, e.g. $\boldsymbol{v} = (v_1, \ldots, v_m)^t : V \subset \mathbb{R}^n \to \mathbb{R}^m$. These derivatives are written as vectors:

$$\begin{aligned} \boldsymbol{v}' &= (\partial_1^1 \boldsymbol{v}, \partial_1^2 \boldsymbol{v}, \ldots, \partial_1^k, \ldots, \ldots, \partial_m^1, \ldots, \partial_m^n \boldsymbol{v})^t, \\ \boldsymbol{v}'' &= ()\partial_1^{1,1} \boldsymbol{v}, \ldots, \partial_1^{1,n} \boldsymbol{v}, \ldots, \partial_1^{n,1} \boldsymbol{v}, \ldots, \partial_1^{n,n} \boldsymbol{v}, \partial_2^{1,1} \boldsymbol{v}, \ldots, \partial_2^{n,n} \boldsymbol{v}, \ldots, \partial_m^{1,1} \boldsymbol{v}, \ldots, \partial_m^{n,n} \boldsymbol{v})^t, \end{aligned} \tag{3.19}$$

where $\partial_j^k \boldsymbol{v} := (\partial v_j)/\partial x_k)$ and $\partial_j^{k_1,k_2} \boldsymbol{v} := (\partial^2 v_j)/(\partial x_{k_1} \partial x_{k_2})$.

Lemma 3.7. *The squared volatility of the GJR-GARCH model*

$$h_t = \alpha_0 + \sum_{i=1}^{q} (1 + \kappa I(\phi_{t-i} < 0) \alpha_i \phi_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3.20}$$

has the a.s. representation

$$h_t = \alpha_0 + \sum_{i=1}^{q} (1 + \kappa/2) \alpha_i \left(\frac{2(1+\kappa)}{2+\kappa} I(\phi_{t-i} < 0) \phi_{t-i}^2 + \frac{1}{1+\kappa/2} I(\phi_{t-i} \geq 0) \phi_{t-i}^2 \right) + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3.21}$$

Proof. It is sufficient to show that

$$(1 + \kappa I(\phi_{t-i} < 0) \alpha_i \phi_{t-i}^2 = (1 + \kappa/2) \alpha_i \left(\frac{2(1+\kappa)}{2+\kappa} I(\phi_{t-i} < 0) \phi_{t-i}^2 + \frac{1}{1+\kappa/2} I(\phi_{t-i} \geq 0) \phi_{t-i}^2 \right),$$

$i = 1, \ldots, q$.

It is

$$\begin{aligned} & (1 + \kappa/2) \alpha_i \left(\frac{2(1+\kappa)}{2+\kappa} I(\phi_{t-i} < 0) \phi_{t-i}^2 + \frac{1}{1+\kappa/2} I(\phi_{t-i} \geq 0) \phi_{t-i}^2 \right) \\ = \; & \left((1+\kappa) I(\phi_{t-i} < 0) + I(\phi_{t-i} \geq 0) \right) \alpha_i \phi_{t-i}^2 \\ = \; & (1 + \kappa I(\phi_{t-i} < 0)) \alpha_i \phi_{t-i}^2, \end{aligned}$$

since $1 = I(\phi_{t-i} < 0) + I(\phi_{t-i} \geq 0)$. This completes the proof. $\square$

We will see that it is important that the absolute values of the coefficients $\frac{2(1+\kappa)}{2+\kappa}$ and $\frac{1}{1+\kappa/2}$in (3.21) have the upper bound 2. This will be proven in the following lemma.

Lemma 3.8.
The coefficients $\frac{2(1+\kappa)}{2+\kappa}$ *and* $\frac{1}{1+\kappa/2}$*in (3.21) have the following properties:*

$$\left|\frac{2(1+\kappa)}{2+\kappa}\right| \leq 2 \quad \text{and}$$
$$\left|\frac{1}{1+\kappa/2}\right| \leq 2.$$

Proof. Recall that $\kappa = \frac{4\gamma^*}{(\gamma^*-1)^2}$ (see Appendix Equation (4.3)) and $|\gamma^*|< 1$ by definition of the AGARCH-model (see appendix Equation (4.1)). Using this and applying the triangle inequality we get

$$\begin{aligned} \left|\frac{2(1+\kappa)}{2+\kappa}\right| &= \left|\frac{2(4\gamma^* + (\gamma^*-1)^2}{(\gamma^*-1)^2} \frac{(\gamma^*-1)^2}{4\gamma^* + 2(\gamma^*-1)^2}\right| \\ &= \left|\frac{8\gamma^* + 2(\gamma^*-1)^2}{4\gamma^* + 2(\gamma^*-1)^2}\right| \\ &= \left|\frac{(\gamma^*+1)^2}{{\gamma^*}^2+1}\right| \leq \frac{1+2+1}{2} = 2. \end{aligned} \tag{3.22}$$

Similarly, we have

$$\left|\frac{1}{1+\kappa/2}\right| = \left|\frac{1}{\frac{2\gamma^*+(\gamma^*-1)^2}{(\gamma^*-1)^2}}\right| = \left|\frac{(\gamma^*-1)^2}{{\gamma^*}^2+1}\right| \leq 2 \tag{3.23}$$

This completes the proof. $\square$

3.4 Proof of Theorem 3.2

Let $\widehat{\boldsymbol{\theta}}_2$ be the maximum likelihood estimator if $\boldsymbol{\theta}_2^0$,

$$\widehat{\boldsymbol{\theta}}_2 = \arg\max \tilde{L}_T(\boldsymbol{\theta}_2, \boldsymbol{\phi}),$$

where the quasi log-likelihood function for the model (3.2), (3.3), (3.21), and $\phi_t = \epsilon_t/g_t^{1/2}$ is

$$\tilde{L}_T(\boldsymbol{\theta}_2, \boldsymbol{\phi}) = \sum_{t=1}^{T} \ell(\boldsymbol{\theta}_2, \phi_t)$$

with

$$\ell(\boldsymbol{\theta}_2, \epsilon_t) = k - \frac{1}{2}\left(\log h_t(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2) + \frac{\phi_t^2}{h_t(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2)}\right).^{15}$$

To prove the consistency and asymptotic normality, it is sufficient to show that the QMLE of a GJR-GARCH model (here denoted by $\widehat{\boldsymbol{\theta}}_2$) is consistent asymptotically normal. Straumann et al. (2006) show that if the conditions in the following Theorem are satisfied, then the QMLE is indeed consistent and asymptotically normal.

Theorem 3.9. *Consider*

$$\epsilon_t = z_t h_t^{1/2} \tag{3.24}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q}(1+\kappa/2)\alpha_i\left(\frac{2(\kappa+1)}{\kappa+2}I(\phi_{t-i}<0) + \frac{1}{1+\kappa/2}I(\phi_{t-i}\geq 0)\right)\phi_{t-i}^2 + \sum_{j=1}^{p}\beta_j h_{t-j} \tag{3.25}$$

with $\boldsymbol{\theta}_2 = (\alpha_0, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p, \kappa) = (\alpha_0^0, \alpha_1^0, \ldots, \alpha_q^0, \beta_1^0, \ldots, \beta_p^0, \kappa^0) = \boldsymbol{\theta}_2^0$. *Let* K *be a compact set of* Θ_2 *containing the true parameter* $\boldsymbol{\theta}_2^0$.
If

- *C.1 Model (3.24), (3.25) with* $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0$ *admits a unique stationary solution with* $\mathbb{E}[\log^+ h_0] < \infty$,
- *C.2 The conditions of Proposition (3.6) are satisfied for a compact set* $K \in \boldsymbol{\theta}_2$ *with* $\boldsymbol{\theta}_2^0 \in K$.
- *C.3 The class of functions* $\{g_{\boldsymbol{\theta}_2} \mid \boldsymbol{\theta}_2 \in K\}$ *is uniformly bounded from below, that is, there exists a constant* $\underline{g} > 0$, *such that* $g_{\boldsymbol{\theta}_2}(\boldsymbol{x}, \boldsymbol{s}) \geq \underline{g}$ *for all* $(\boldsymbol{x}, \boldsymbol{s}) \in \mathbb{R}^q \times [0,\infty)^p$, $\boldsymbol{\theta}_2 \in K$,
- *C.4 The following identifiability condition holds on* K*: For all* $\boldsymbol{\theta}_2 \in K$,

  $$f_0(\boldsymbol{\theta}_2) \equiv h_0 \ a.s.$$

  if and only if $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0$,

hold, then $\widehat{\boldsymbol{\theta}}_2 \xrightarrow{p} \boldsymbol{\theta}_2^0$. *If additionally the conditions*

- *N.1 For all* $j \in \{1, \ldots, p\}$ *and* $k \in \{1, \ldots, d+p\}$

  $$\mathbb{E}\left[\log^+\left(\sup_{\boldsymbol{\theta}_2 \in K}\left|\partial_j^k[\varphi_0^{(1)}(\boldsymbol{f}_0)](\boldsymbol{\theta}_2)\right|\right)\right] < \infty$$

[15] For ease of notation $\boldsymbol{\theta}_1^0$ will be dropped for the remainder of the proof.

Moreover, there exists a stationary sequence $\bar{C}_1(t)$) with $\mathbb{E}\left[\log^+ \bar{C}_1(0)\right] < \infty$ and $\nu \in (0,1]$ such that

$$\sup_{\boldsymbol{\theta}_2 \in K} \left| \partial_j^k [\varphi_t^{(1)}(\boldsymbol{u})](\boldsymbol{\theta}_2) - \partial_j^k [\varphi_t^{(1)}(\tilde{\boldsymbol{u}})](\boldsymbol{\theta}_2) \right| \leq \bar{C}_1(t) \left| \boldsymbol{u} - \tilde{\boldsymbol{u}} \right|^{\nu}, \quad \boldsymbol{u}, \tilde{\boldsymbol{u}} \in [0,\infty)^p,$$

for every $j \in \{1,\ldots,p\}$, $k \in \{1,\ldots,d+p\}$ and $t \in \mathbb{Z}$.

N.2 The function $(\boldsymbol{\theta}_2, \boldsymbol{s}) \mapsto g_{\boldsymbol{\theta}_2}(\boldsymbol{x}, \boldsymbol{s})$ on $K \times [0,\infty)^p$ is twice continuously differentiable for every fixed $\boldsymbol{x} \in \mathbb{R}^q$. For all $j \in \{1,\ldots,p\}$ and $k_1, k_2 \in \{1,\ldots,d+p\}$,

$$\mathbb{E}\left[\log^+ \left(\sup_{\boldsymbol{\theta}_2 \in K} \left| \partial_j^{k_1,k_2} [\varphi_0^{(1)}(\boldsymbol{f}_0)](\boldsymbol{\theta}_2) \right| \right) \right] < \infty$$

The sequence of first derivatives $(\boldsymbol{f}_t')$ satisfies $\mathbb{E}(\log^+ \|\boldsymbol{f}_0'\|_K) < \infty$. Moreover, there exists a stationary sequence $(\bar{C}_2(t))$ with $\mathbb{E}[\log^+ \bar{C}_2(0)] < \infty$ and $\tilde{\nu} \in (0,1]$ such that

$$\sup_{\boldsymbol{\theta}_2 \in K} \left| \partial_j^{k_1,k_2} [\varphi_t^{(1)}(\boldsymbol{u})](\boldsymbol{\theta}_2) - \partial_j^{k_1,k_2} [\varphi_t^{(1)}(\tilde{\boldsymbol{u}})](\boldsymbol{\theta}_2) \right| \leq \bar{C}_2(t) \left| \boldsymbol{u} - \tilde{\boldsymbol{u}} \right|^{\tilde{\nu}}, \quad \boldsymbol{u}, \tilde{\boldsymbol{u}} \in [0,\infty)^p$$

for every $j \in \{1,\ldots,p\}$, $k_1, k_2 \in \{1,\ldots,d+p\}$ and $t \in \mathbb{Z}$.

N.3 The following moment conditions hold:

(i) $\mathbb{E} z_0^4 < \infty$,

(ii) $\mathbb{E}\left[\frac{|f_0'(\boldsymbol{\theta}_2^0)|^2}{h_0^2}\right] < \infty$,

(iii) $\mathbb{E}\|\ell_0'\|_K < \infty$,

(iv) $\mathbb{E}\|\ell_0''\|_K < \infty$, *where*

$$\ell_t'(\boldsymbol{\theta}_2) = -\frac{1}{2} \frac{f_t'(\boldsymbol{\theta}_2)}{f_t(\boldsymbol{\theta}_2)} \left(1 - \frac{\phi_t^2}{f_t(\boldsymbol{\theta}_2)}\right) \tag{3.26}$$

$$\ell_t''(\boldsymbol{\theta}_2) = -\frac{1}{2} \frac{1}{f_t(\boldsymbol{\theta}_2)^2} \left((f_t'(\boldsymbol{\theta}_2))^t f_t'(\boldsymbol{\theta}_2) \left(2 \frac{\phi_t^2}{f_t(\boldsymbol{\theta}_2)} - 1\right) + f_t''(\boldsymbol{\theta}_2)(f_t(\boldsymbol{\theta}_2) - \phi_t^2) \right) \tag{3.27}$$

N.4 The components of the vector $\frac{\partial g_{\boldsymbol{\theta}_2}}{\partial \boldsymbol{\theta}_2}(\phi_0, h_0)|_{\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0}$ are linearly independent random variables,

hold, then $\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2^0\right) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_2^0)$.

Consistency:

We verify the conditions C.1-C.4. Stationarity holds because of AH1 (see Theorem 2.2), and according to Jensen's inequality as well as Lemma 2.4 we have $\mathbb{E}[\log^+ h_0] \leq \log^+ \mathbb{E}[h_0] < \infty$. Thus, condition C.1 is valid. We turn to Condition C.2. Note that

$$[\varphi_t(\boldsymbol{s})](\boldsymbol{\theta_2}) = \left(\alpha_0 + \sum_{i=1}^{q}(1 + \kappa I(\phi_{t+1-i} < 0))\alpha_i \phi_{t+1-i}^2\right) \boldsymbol{e}_1 + \boldsymbol{C}(\boldsymbol{\theta}_2)\boldsymbol{s}(\boldsymbol{\theta}_2),$$

for $\boldsymbol{s} \in \mathbb{C}(K, [0,\infty)^p)$, where

$$\boldsymbol{e}_1 = (1, 0, \ldots, 0)^t \in \mathbb{R}^p, \tag{3.28}$$

$$\boldsymbol{C}(\boldsymbol{\theta}_2) = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \cdots & \beta_{p-1} & \beta_p \\ 1 & 0\cdots & \cdots & 0 & 0 & \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \cdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 & 0 \end{pmatrix} \in \mathbb{R}^{p \times p} \tag{3.29}$$

Since $\mathbb{E}|h_0|^{2\eta} < \infty$ for $\eta \in (0,1]$ (see Lemma 2.4), $\mathbb{E}|\phi_0|^{2\eta} < \infty$ also holds. Together with Jensen's inequality and AH1 we obtain

$$\mathbb{E}[\log^+ \|\varphi_0(\zeta_0^2)\|_K] \overset{\text{Jensen}}{\leq} \log^+ \mathbb{E}[\|\varphi_0(\zeta_0^2)\|_K] = \log^+ \mathbb{E}[\sup_{\boldsymbol{\theta}_2 \in K} [\varphi_0(\zeta_0^2)](\boldsymbol{\theta}_2)] < \infty$$

Furthermore, because $\varphi_t^{(r)} = \varphi_t \circ \varphi_{t-1} \circ \varphi_{t-r+1}$, $\varphi_0^{(r)}(\boldsymbol{s})$ can also be written as

$$\varphi_0^{(r)}(\boldsymbol{s}) = \boldsymbol{C}^r \boldsymbol{s} + \boldsymbol{R},$$

where $\boldsymbol{R}$ is a term that does not depend on $\boldsymbol{s}$. Therefore, by the definition of Λ, we have $\Lambda(\varphi_0^{(r)}) = \Lambda(\boldsymbol{C}^r)$, $r \in \mathbb{N}$, where $\boldsymbol{C}^r$ has to be understood as the map $\boldsymbol{s} \mapsto \boldsymbol{C}^r \boldsymbol{s}$ on $\mathbb{C}(K, [0,\infty)^p)$. Because of the stationarity, $\beta_\Sigma := \sum_{j=1}^{p} \beta_j < 1$. Using the exact same manipulations as in (2.10) we get that the characteristic polynomial of $\boldsymbol{C}$ is given by:

$$p_{\boldsymbol{C}}(\mu) = \det(\mu \boldsymbol{I} - \boldsymbol{C}) = \mu^p - \sum_{j=1}^{p} \beta_j \mu^{p-j} = \mu^p \left(1 - \sum_{j=1}^{p} \beta_j \mu^{-j}\right).$$

We show now that the spectral radius of $\boldsymbol{C}$, $\rho(\boldsymbol{C})$, is strictly smaller than $\beta_\Sigma^{1/p}$. Assuming that $|\mu| > \beta_\Sigma^{1/p}$, then $0 < \beta_\Sigma < 1$ together with repeated application of the triangle inequality gives us

$$|\det(\mu \boldsymbol{I} - \boldsymbol{C})| \geq 1 - \sum_{j=1}^{p} \beta_j |\mu|^{-j} > 1 - \sum_{j=1}^{p} \beta_j \beta_{\Sigma}^{-j/q} \geq 1 - \beta_{\Sigma}^{-1} \sum_{j=1}^{p} \beta_j = 0.$$

Consequently, we have indeed $\rho(\boldsymbol{C}) < 1$, and by the Jordan normal decomposition, this gives us

$$\|\boldsymbol{C}^r\|_{op} \leq c\beta_{\Sigma}^{r/p}, \quad r \geq 0, \tag{3.30}$$

where c is a constant not depending on p or β_{Σ}.
Now, since K is compact, $\bar{\beta} := \sup_{\boldsymbol{\theta}_2 \in K} (\sum_{j=1}^{p} \beta_j) < 1$ and a pointwise application of (3.30) for each $\boldsymbol{\theta}_2 \in K$ entails

$$|(\boldsymbol{C}^r \boldsymbol{s})(\boldsymbol{\theta}_2)| \leq c' \left(\sum_{j=1}^{p} \beta_j \right)^{r/p} |\boldsymbol{s}(\boldsymbol{\theta}_2)|, \quad \boldsymbol{\theta}_2 \in K,$$

for any $\boldsymbol{s} \in \mathbb{C}(K, [0, \infty)^p)$, where c' does not depend on $\boldsymbol{\theta}_2$ or $\boldsymbol{s}$. Taking the supremum on both sides, gives us

$$\|\boldsymbol{C}^r \boldsymbol{s}\|_K \leq c' \bar{\beta}^{r/p} \|\boldsymbol{s}\|_K, \quad r \geq 0.$$

This shows that $\mu(\boldsymbol{C}^r) \geq c' \bar{\beta}^{r/p} \longrightarrow 0$ as $r \to \infty$. Since $\varphi_0^{(1)} = \varphi_0$, condition C.2 is satisfied and with this that $\boldsymbol{f}_t = (f_t, \ldots, f_{t-p+1})^t$ is properly defined (see Proposition 3.6).
Since condition C.3 is obviously satisfied (by definition of $g_{\boldsymbol{\theta}_2}$), we directly move on to condition C.4. The "if"-part is trivially satisfied by the virtue of Proposition 3.6. To show that if $f_0(\boldsymbol{\theta}_2) = h_0$ a.s., then $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0$, we will use two lemmas. The first one derives an almost sure presentation of f_t, whereas the second is concerned with the identifiability of κ.

Lemma 3.10.
The following almost sure representation for $f_t(\boldsymbol{\theta}_2)$ is valid:

$$f_t(\boldsymbol{\theta}_2) = \xi_0(\boldsymbol{\theta}_2) + \sum_{\ell=1}^{\infty} \xi_\ell(\boldsymbol{\theta}_2) D_{t-\ell}, \quad \boldsymbol{\theta}_2 \in K, \tag{3.31}$$

where

$$D_{t-\ell} = \left(\frac{2(\kappa+1)}{(\kappa+2)} I(\phi_{t-\ell} < 0)\phi_{t-\ell}^2 + \frac{1}{1+\kappa/2} I(\phi_{t-\ell} \geq 0)\phi_{t-\ell}^2 \right),$$

and the sequence $\{\xi_\ell(\boldsymbol{\theta}_2)\}$ *(*$\ell \in \mathbb{N}$*) is given by*

$$\xi_0(\boldsymbol{\theta}_2) = \frac{\alpha_0}{b_{\boldsymbol{\theta}}(1)} = \frac{\alpha_0}{1-\sum\limits_{j=1}^{p}\beta_j} \quad \textit{and} \tag{3.32}$$

$$\sum_{\ell=1}^{\infty}\xi_\ell(\boldsymbol{\theta}_2)z^\ell = \frac{a_{\boldsymbol{\theta}_2}(z)}{b_{\boldsymbol{\theta}_2}(z)}, \quad |z|< 1, \tag{3.33}$$

where $a_{\boldsymbol{\theta}_2}(z) = \sum\limits_{i=1}^{q}(1+\kappa/2)\alpha_i z^i$ *and* $b_{\boldsymbol{\theta}_2}(z) = 1 - \sum\limits_{j=1}^{p}\beta_j z^j$*.*

Proof. Before beginning with the actual proof of the lemma, recall that in Lemma 3.7 it was proven that the GJR-GARCH model with ϕ_t (Equation (3.20)) obeys the following ARMA(p,q) equation:[16]

$$h_t = \alpha_0 + \sum_{i=1}^{q}(1+\kappa/2)\alpha_i D_{t-i} + \sum_{j=1}^{p}\beta_j h_{t-j}, \quad t \in \mathbb{Z}. \tag{3.34}$$

For the proof of this lemma I will use the results of Brockwell et al. (1991, sections 3.1-3.2). The proof builds on the fact that (3.31) obeys an ARMA (p,q) process, i.e. the right side of Equation (3.34) where the h_{t-j} are replaced with $f_{t-j}(\boldsymbol{\theta}_2)$. Because of (3.22), (3.23) and Corollary 2.3, we have

$$\mathbb{E}\left[\sup_t\left|\frac{2(\kappa+1)}{(\kappa+2)}I(\phi_t<0)\phi_t^2 + \frac{1}{1+\kappa/2}I(\phi_t \geq 0)\phi_t^2\right|\right] < \infty$$

This together with $\sum\limits_{\ell=1}^{\infty}|\xi_\ell(\boldsymbol{\theta}_2)|< \infty$ for all $\boldsymbol{\theta}_2 \in K$ (due to AH1 and $|z|< 1$) gives us, according to Brockwell et al. (1991, Proposition 3.1.1), that

$$\sum_{\ell=1}^{\infty}\xi_\ell(\boldsymbol{\theta}_2)\left(\frac{2(\kappa+1)}{(\kappa+2)}I(\phi_{t-\ell}<0)\phi_{t-\ell}^2 + \frac{1}{1+\kappa/2}I(\phi_{t-\ell}\geq 0)\phi_{t-\ell}^2\right)$$

converges absolutely a.s.

Moreover, we observe that the zeros of the $b_{\boldsymbol{\theta}_2}(z)$ lie outside the unit disc. This is fulfilled by the virtue of $\bar{\beta} < 1$, because if $|z|< \bar{\beta}^{-1/p}$, then

$$|b_{\boldsymbol{\theta}_2}(z)|\geq 1 - \sum_{j=1}^{p}\beta_j|z|^j > 1 - \bar{\beta}(\bar{\beta}^{-1/p})^p = 0.$$

[16] See Appendix for a definition of a ARMA(p,q)-process

Additionally, since $a_{\boldsymbol{\theta}_2}(z)$ and $b_{\boldsymbol{\theta}_2}(z)$ do not have any common zeroes (AH2) and $|z|< 1$, Brockwell et al. (1991, Theorem 3.1.1) confirms that $\{\xi_\ell(\boldsymbol{\theta}_2)\}$ is given by (3.32) and (3.33).
It remains to be checked that the a.s. representation for $\{f_t\}$ indeed obeys (3.34) (again replacing the h_{t-j} with $f_{t-j}(\boldsymbol{\theta}_2)$).
Notice that (3.31) can be rewritten as

$$
\begin{aligned}
h_t(\boldsymbol{\theta}_2) &= \frac{\alpha_0}{1-\sum_{j=1}^{p}\beta_j} + \sum_{\ell=1}^{\infty}\xi_\ell(\boldsymbol{\theta}_2)L^\ell D_t^2 \\
&= \frac{\alpha_0}{b_{\boldsymbol{\theta}_2}(1)} + \frac{a_{\boldsymbol{\theta}_2}(L)}{b_{\boldsymbol{\theta}_2}(L)}D_t, \qquad (3.35)
\end{aligned}
$$

where L is the lag-operator, i.e. $L^k X_t = X_{t-k}$ for a time series $\{X_t\}$. (3.35) can now be rewritten as

$$b_{\boldsymbol{\theta}_2}(L)f_t(\boldsymbol{\theta}_2) = \alpha_0 + a_{\boldsymbol{\theta}_2}(L)D_t,$$

which is just

$$f_t(\boldsymbol{\theta}_2) = \alpha_0 + \sum_{i=1}^{q}(1+\kappa/2)\alpha_i D_{t-i} + \sum_{j=1}^{p}\beta_j h_{t-j}(\boldsymbol{\theta}_2).$$

This completes the proof. □

We now turn to the second Lemma.

Lemma 3.11.
Suppose that the distribution of z_0 is not concentrated at two points. Then for any $\boldsymbol{\theta}_2 \in K$, the relation $f_0(\boldsymbol{\theta}_2) = f_0(\boldsymbol{\theta}_2^0)$ a.s. implies $\kappa = \kappa^0$.

Proof. Firstly, observe that $f_0(\boldsymbol{\theta}_2) = f_0(\boldsymbol{\theta}_2^0)$ is equivalent to $f_r(\boldsymbol{\theta}_2) = f_r(\boldsymbol{\theta}_2^0)$ for any r, in particular for $r := \max\{p, q\}$.
Similarly to Section 2, we define

$$
\begin{aligned}
c_k(z_{t-k}) &= \beta_k + (1+\kappa I(z_{t-k}<0))\alpha_k z_{t-k}^2, \quad \text{and} \\
c_k^0(z_{t-k}) &= \beta_k^0 + (1+\kappa^0 I(z_{t-k}<0))\alpha_k^0 z_{t-k}^2
\end{aligned}
$$

$k = 1, \ldots, r$ and $\alpha_i = 0$ for $i > q$ and $\beta_j = 0$ for $j > p$. Now, $f_r(\boldsymbol{\theta}_2) = f_r(\boldsymbol{\theta}_2^0)$ a.s. entails

$$\begin{aligned} & \alpha_0^0 + \sum_{i=1}^{r} c_i^0(z_{r-i})h_{r-i} - \left(\alpha_0 + \sum_{i=1}^{r} c_i(z_{r-i})h_{r-i}\right) \\ = \quad & (\alpha_0 - \alpha_0^0) + \sum_{i=1}^{r} \underbrace{c_i^0(z_{r-i}) - c_i(z_{r-i})}_{=:Q_{r-i}} h_{r-i} = 0 \end{aligned} \tag{3.36}$$

Define $r^* := \min(i \in \{1, \ldots, q\} : \alpha_i^0 > 0)$. Through repeatedly re-expressing each term h_{t-j}, $j = 1, \ldots, (k^* - 1)$, in the equation

$$h_t = \alpha_0^0 + \sum_{i=r^*}^{q} (1 + \kappa^0 I(\phi_{t-i} < 0))\alpha_i^0 \phi_{t-i}^2 + \sum_{j=1}^{p} \beta_j^0 h_{t-j}$$

by past observations and past standardized squared volatilities, one sees that h_t can be rewritten as a function of $\{\phi_{t-r^*}^2, \phi_{t-r^*-1}, \ldots; h_{t-r^*}, h_{t-r^*-1}, \ldots\}$. This means h_t is $\mathcal{F}_{t-r^*}$-measurable. Observe that $h_{r-1} \geq \alpha_0^0 > 0$. This implies, together with Equation (3.36), that Q_{r-1} is a function of $h_0, \ldots, h_{r-1}$ and consequently (with the above observations) $\mathcal{F}_{r-r^*-1}$-measurable. Since Q_{r-1} is also independent of $\mathcal{F}_{r-r^*-1}$, it must be a constant (see Lemma 4.4 in the appendix). It follows by the exact same arguments, that $Q_{r-2}, \ldots, Q_{r-r^*}$ are constants.
Notice that the degeneracy of Q_{r-r^*} means that

$$(1 + \kappa^0 I(z_{r-r^*} < 0))\alpha_{r^*}^0 z_{r-r^*}^2 - (1 + \kappa I(z_{r-r^*} < 0))\alpha_{r^*} z_{r-r^*}^2 = \tilde{c}$$

for a certain constant $\tilde{c}$. Now, if $z_{r-r^*} \geq 0$, we obtain

$$(\alpha_{r^*}^0 - \alpha_{r^*})z_{r-r^*}^2 = \tilde{c}, \tag{3.37}$$

and conversely, if $z_{r-r^*} < 0$, we get

$$\left((1 + \kappa^0)\alpha_{r^*}^0 - (1 + \kappa)\alpha_{r^*}\right) z_{r-r^*}^2 = \tilde{c}. \tag{3.38}$$

Equations (3.37) and (3.38) can only be jointly satisfied if $\tilde{c} = 0$ since z_{r-r^*} is, per assumption, not concentrated at two points. Moreover, since z_{r-r^*} is centered at zero with variance one, the distribution of z_{r-r^*} has positive mass on either side of the real line (assuming zero is the center of the real line). This implies

$$\alpha_{r^*}^0 = \alpha_{r^*},$$

and at the same time also

$$(1+\kappa^0)\alpha^0_{r^*} = (1+\kappa)\alpha_{r^*}.$$

Together this implies $\kappa^0 = \kappa$, which was what we wanted to show. $\square$

Finally, the following lemma will establish the "only if"-part of condition C.4.

Lemma 3.12. *Suppose that AH2 and AH4 hold. Then, if for any $\boldsymbol{\theta}_2 \in K$,*

$$f_0(\boldsymbol{\theta}_2) = h_0,$$

then $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0$.

Proof. In the preceding lemma, it was shown that $f_0(\boldsymbol{\theta}_2) = f_0(\boldsymbol{\theta}_2^0)$ implies $\kappa = \kappa^0$,[17] which implies $D_t = D_t^0$ (see Equation (3.31)).[18] Now using representation (3.31), $f_0(\boldsymbol{\theta}_2) = f_0(\boldsymbol{\theta}_2^0)$ can be rewritten as:

$$\xi_0(\boldsymbol{\theta}_2^0) + \sum_{\ell=1}^{\infty} \xi_\ell(\boldsymbol{\theta}_2^0) D^0_{0-\ell} - \left(\xi_0(\boldsymbol{\theta}_2) + \sum_{\ell=1}^{\infty} (\xi_\ell(\boldsymbol{\theta}_2) D_{0-\ell} \right) \tag{3.39}$$

$$= \left(\xi_0(\boldsymbol{\theta}_2^0) - \xi_0(\boldsymbol{\theta}_2)\right) + \sum_{\ell=1}^{\infty} \left(\xi_\ell(\boldsymbol{\theta}_2) - \xi_\ell(\boldsymbol{\theta}_2^0)\right) D^0_{0-\ell} \equiv 0. \tag{3.40}$$

The proof proceeds now as follows. I will show that $\xi_\ell(\boldsymbol{\theta}_2) = \xi_\ell(\boldsymbol{\theta}_2^0)$. And then that this implies $a_{\boldsymbol{\theta}_2} = a_{\boldsymbol{\theta}_2^0}$ as well as $b_{\boldsymbol{\theta}_2} = b_{\boldsymbol{\theta}_2^0}$. The former will be shown by contradiction.
Define $\ell \geq 1$ as the smallest integer $\ell \geq 1$ for which $w_\ell = \xi_\ell(\boldsymbol{\theta}_2) - \xi_\ell(\boldsymbol{\theta}_2^0) \neq 0$. Since $h_{-\ell^*} \geq \alpha_0^0 > 0$, we have that

$$\frac{2(\kappa+1)}{\kappa+2} I(z_{-\ell^*} < 0) + \frac{1}{1+\kappa/2} I(z_{-\ell^*} \geq 0)) z^2_{-\ell^*} \tag{3.41}$$

$$= \frac{\left(\xi_0(\boldsymbol{\theta}_2^0) - \xi_0(\boldsymbol{\theta}_2)\right) + \sum_{\ell=\ell^*+1}^{\infty} \left(\xi_\ell(\boldsymbol{\theta}_2) - \xi_\ell(\boldsymbol{\theta}_2^0)\right) D^0_{0-\ell}}{w_{\ell^*} h_{-\ell^*}} \tag{3.42}$$

$\mathcal{F}_{-\ell^*-1}$-measurable and at the same time independent of $\mathcal{F}_{-\ell^*-1}$. According to Lemma 4.4, this is only possible if $D^0_{-\ell^*}$ is a constant. However, $D^0_{-\ell^*}$ cannot be a constant, because it was assumed that the distribution of z_0 is not concentrated at two points. This gives us the desired

---

[17] Recall that $f_0(\boldsymbol{\theta}_2^0) = h_0$, see Proposition 3.6

[18] Consistent with the notation that was used before, D_t^0 denotes the term D_t, while κ is replaced by κ^0.

contradiction and $\xi_\ell(\boldsymbol{\theta}_2) = \xi_\ell(\boldsymbol{\theta}_2^0)$ is indeed true. Using this equality in (3.33) implies

$$\frac{a_{\boldsymbol{\theta}_2}(z)}{b_{\boldsymbol{\theta}_2}(z)} = \frac{a_{\boldsymbol{\theta}_2^0}(z)}{b_{\boldsymbol{\theta}_2^0}(z)}.$$

This yields

$$a_{\boldsymbol{\theta}_2} = q(z)a_{\boldsymbol{\theta}_2^0}(z) \text{ and } b_{\boldsymbol{\theta}_2} = q(z)b_{\boldsymbol{\theta}_2^0}$$

with a rational function $q(z)$. Notice that $q(z)$ cannot have any pole because of AH2.[19] Thus, $q(z)$ is a polynomial. Since $(\alpha_p^0, \beta_q^0) \neq (0,0)$ (see Definition 2.1), the degree of q is zero (otherwise either $a_{\boldsymbol{\theta}_2}(z)$ or $b_{\boldsymbol{\theta}_2}(z)$ would have a degree strictly greater than p or q). Eventually, we conclude that $q \equiv 1$ because the constants in the polynomials $b_{\boldsymbol{\theta}_2}$ and $b_{\boldsymbol{\theta}_2^0}$ are 1. Therefore, $a_{\boldsymbol{\theta}_2} = a_{\boldsymbol{\theta}_2^0}$ and $b_{\boldsymbol{\theta}_2} = b_{\boldsymbol{\theta}_2^0}$, which directly gives us $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0$. This completes the proof. □

Asymptotic Normality:

We will now move on to prove the asymptotic normality and verify conditions N.1 -N.4. Regarding N.1 and N.2, recall that

$$[\varphi_t(\boldsymbol{s})](\boldsymbol{\theta}_2) = \left(\alpha_0 + \sum_{i=1}^{q}(1+\kappa I(\phi_{t+1-i} < 0))\alpha_i \phi_{t+1-i}^2\right)\boldsymbol{e}_1 + \boldsymbol{C}(\boldsymbol{\theta}_2)\boldsymbol{s}(\boldsymbol{\theta}_2).$$

I will show that $\mathbb{E}[\log^+ h_0] < \infty$, $\mathbb{E}[\log^+ \|f_0\|_K] < \infty$, $\mathbb{E}[\|h_0\|_K^\eta] < \infty$, $\mathbb{E}[\|h_0'\|_K^\eta] < \infty$, and $\mathbb{E}[\log^+ \|h_0'\|_K^\eta] < \infty$, for $\eta > 0$. The rest of the steps to verification of N.1 and N.2 then follow straightforwardly.

In Lemma (2.4) it was shown that there exists an $\eta \in (0,1]$ such that $\mathbb{E}[|h_0|^\eta] < \infty$. This of course implies $\mathbb{E}[|\phi_0|^{2\eta}] < \infty$.[20] The Minowski inequality applied to the left-hand side of (3.31) entails $\mathbb{E}\|f_0\|_K^\eta < \infty$. Hence, the just made observations demonstrate

$$\mathbb{E}[\log^+ h_0] < \infty \text{ and } \mathbb{E}[\log^+ \|f_0\|_K] < \infty,$$

and we have established the moments of f_t. For the moments of f_t', we exploit the fact that in (3.31), the sum and the differentiation (with respect to $\boldsymbol{\theta}_2$) can be interchanged. This is shown by observing that

$$\xi_\ell'(\boldsymbol{\theta}_2) = \frac{\partial}{\partial \boldsymbol{\theta}_2}\left(\frac{1}{\ell!}\frac{\partial^\ell}{\partial z^\ell}\left(\frac{\alpha_{\boldsymbol{\theta}_2}(z)}{\beta_{\boldsymbol{\theta}_2}(z)}\right)\right)\Bigg|_{z=0} = \frac{1}{\ell!}\frac{\partial^\ell}{\partial z^\ell}\left(\frac{\partial}{\partial \boldsymbol{\theta}_2}\left(\frac{\alpha_{\boldsymbol{\theta}_2}(z)}{\beta_{\boldsymbol{\theta}_2}(z)}\right)\right)\Bigg|_{z=0}. \tag{3.43}$$

[19] $q(z)$ having a pole would imply that $a_{\boldsymbol{\theta}_2^0}$ and $b_{\boldsymbol{\theta}_2^0}$ have a common zero.

[20] Recall the notational changes alluded in Footnote 10.

The latter identity shows that $\xi'_\ell(\boldsymbol{\theta}_2)$ can be computed from $a_{\boldsymbol{\theta}_2}(z)/b_{\boldsymbol{\theta}_2}(z)$. Recall that $b_{\boldsymbol{\theta}_2}(z) \neq 0$ for $|z| < \bar{\beta}^{-1/p}$ (see proof of Lemma 3.10). Applying the Cauchy-Schwartz inequality to (3.43) and applying Cauchy inequalities to $\partial/(\partial\boldsymbol{\theta}_2)(a_{\boldsymbol{\theta}_2}(z)/b_{\boldsymbol{\theta}_2}(z))$ (see Rudin (1987), mainly Theorem 10.26) yields the existence of a $\tilde{\lambda} \in (0,1)$ and a constant $\tilde{c} > 0$ such that

$$\|\xi'_\ell\|_K \leq \tilde{c}\tilde{\lambda}^\ell.$$

This together with Straumann et al. (2006, Lemma 2.1) directly implies that the sequence of first derivatives of

$$y_m(\boldsymbol{\theta}_2) = \xi_0(\boldsymbol{\theta}_2) + \sum_{\ell=1}^{m} \xi_\ell(\boldsymbol{\theta}_2)(I(\phi_{t-\ell} < 0)\phi_{t-\ell}^2 + I(\phi_{t-\ell} \geq 0)\phi_{t-\ell}^2)$$

converges a.s. on $\mathbb{C}(K, \mathbb{R}^{p+q+2})$ as $m \to \infty$.
Since $y_m \xrightarrow{a.s.} f_t$ in $\mathbb{C}(K)$ and $\{y'_m\}$ converges uniformly on K, we get

$$f'_t(\boldsymbol{\theta}_2) = \xi'_0(\boldsymbol{\theta}_2) + \left(\sum_{\ell=1}^{\infty} \xi'_\ell(\boldsymbol{\theta}_2) D_{t-\ell}\right)\boldsymbol{q} + \left(\sum_{\ell=1}^{\infty} \left(\xi'_\ell(\boldsymbol{\theta}_2)D_{t-\ell} + \xi_\ell(\boldsymbol{\theta}_2)D'_{t-\ell}\right)\right)\boldsymbol{e}_{p+q+2} \;\; a.s. \tag{3.44}$$

where $\boldsymbol{q} = (1, \ldots, 1, 0) \in \mathbb{R}^{p+q+2}$, $\boldsymbol{e}_{p+q+2} = (0, \ldots, 0, 1)^t \in \mathbb{R}^{p+q+2}$, and

$$D'_{t-\ell} = \left(\frac{2}{(\kappa+2)^2} I(\phi_{t-\ell} < 0)\phi_{t-\ell}^2 - \frac{2}{(\kappa+2)^2} I(\phi_{t-\ell} \geq 0)\phi_{t-\ell}^2\right).$$

This demonstrates that summation and differentiation can be interchanged in (3.31). Using the identity in (3.44) and identical arguments on $\mathbb{E}[\|f'_0\|_K^\eta]$ as for $\mathbb{E}[\|f_0\|_K^\eta] < \infty$ above entails $\mathbb{E}[\|f'_0\|_K^\eta] < \infty$. Thus also $\mathbb{E}[\log^+\|f'_0\|_K] < \infty$. In total, N.2 has been established.
Regarding N.3, the N.3 (i) is of course satisfied due to AG5. To verify N.3 (iii) and N.3 (iv) is a little more tricky, N.3 (ii) follows as a by-product of the verification of N.3 (iii).
I will show in the following that the random variables

$$\|f'_0/f_0\|_K \text{ and } \|f''_0/f_0\|_K \tag{3.45}$$

have finite moments of any order and that

$$\mathbb{E}\|\phi_0^2/f_0\|_K^\nu < \infty, \tag{3.46}$$

for any $\nu < 2$. The inequality (3.46) directly follows from Horv et al. (2003, Lemma 5.1) if one replaces $\{\epsilon_t\}$ (the innovations) by $\left(\frac{2(\kappa+1)}{(\kappa+2)} I(z_t < 0)z_t^2 + \frac{1}{1+\kappa/2} I(z_t \geq 0)z_t^2\right)$, $\{y_t\}$ (the observations) by D_t, as well as $\{\sigma_t\}$ by $h_t^{1/2}$. To follow the steps of in the proof of Horv et al. (2003, Lemma 5.1) it is also necessary to make the observations

$$c_1 z_t^2 \leq \left(\frac{2(\kappa+1)}{(\kappa+2)} I(z_t<0) z_t^2 + \frac{1}{1+\kappa/2} I(z_t \geq 0) z_t^2\right) \leq 2z_t^2 \quad \text{and} \tag{3.47}$$

$$c_1 \phi_t^2 \leq \left(\frac{2(\kappa+1)}{(\kappa+2)} I(\phi_t<0) \phi_t^2 + \frac{1}{1+\kappa/2} I(\phi_t \geq 0) \phi_t^2\right) \leq 2\phi_t^2, \tag{3.48}$$

for all $\boldsymbol{\theta}_2 \in K$, where $c_1 > 0$. This of course follows from the fact that $\left|\frac{2(\kappa+1)}{(\kappa+2)}\right| \leq 2$ and $\left|\frac{1}{1+\kappa/2}\right| \leq 2$ on K (see Lemma 3.8). For the boundedness of (3.45), we observe that following the steps of the proof of Horv et al. (2003, Lemma 3.2), there exists $c_2, c_3 > 0$ such that for all $\boldsymbol{\theta}_2 \in K$,

$$|\xi_\ell'(\boldsymbol{\theta}_2)| \leq c_2 \ell \xi_\ell(\boldsymbol{\theta}_2) \text{ and } \|\xi_\ell''(\boldsymbol{\theta}_2)\| \leq c_3 \ell^2 \xi_\ell(\boldsymbol{\theta}_2), \quad \ell \leq 1. \tag{3.49}$$

Differentiating both sides of (3.44) with respect to $\boldsymbol{\theta}_2$ gives us

$$\begin{aligned} f_t''(\boldsymbol{\theta}_2) = \xi_0''(\boldsymbol{\theta}_2) &+ \left(\sum_{\ell=1}^{\infty} \xi_\ell''(\boldsymbol{\theta}_2) D_{t-\ell}\right) \boldsymbol{q}\boldsymbol{q}^t \\ &+ \left(\sum_{\ell=1}^{\infty} \left(\xi_\ell''(\boldsymbol{\theta}_2) D_{t-\ell} + \xi_\ell'(\boldsymbol{\theta}_2) D_{t-\ell}'\right)\right) \boldsymbol{e}_{p+q+2}\boldsymbol{q}^t \\ &+ \left(\sum_{\ell=1}^{\infty} \left(\xi_\ell''(\boldsymbol{\theta}_2) D_{t-\ell} + \xi_\ell'(\boldsymbol{\theta}_2) D_{t-\ell}'\right)\right) \boldsymbol{q}\boldsymbol{e}_{p+q+2}^t \\ &+ \left(\sum_{\ell=1}^{\infty} \left(\xi_\ell''(\boldsymbol{\theta}_2) D_{t-\ell} + 2\xi_\ell'(\boldsymbol{\theta}_2) D_{t-\ell}' + \xi_\ell(\boldsymbol{\theta}_2) D_{t-\ell}''\right)\right) \boldsymbol{e}_{p+q+2}\boldsymbol{e}_{p+q+2}^t. \end{aligned}$$

This together with (3.48 and (3.49) shows that there is a constant $C > 0$ such that

$$|f_t'(\boldsymbol{\theta}_2)|, \|f_t''(\boldsymbol{\theta}_2)\| \leq C\left(1 + \sum_{\ell=1}^{\infty} \ell^2 \xi_\ell(\boldsymbol{\theta}_2) \phi_{t-\ell}^2\right).$$

Together with (3.48) and following the lines of the proof of Horv et al. (2003, Lemma 5.2) we obtain

$$\mathbb{E}\left[\left\|\frac{f_0'}{f_0}\right\|_K^\nu\right] < \infty \text{ and } \mathbb{E}\left[\left\|\frac{f_0''}{f_0}\right\|_K^\nu\right] < \infty,$$

for any $\nu > 0$. Now an application of the Hölder-inequality to the norm of $\ell_t'(\boldsymbol{\theta}_2)$ (see Equation

(3.26)) directly yields $\mathbb{E}[\|\ell_0'\|_K] < \infty$. For $\mathbb{E}[\|\ell_0''\|_K]$, notice that

$$\|xy^t\|_K \leq \|x\|_K \|y\|_K$$

for any two elements $x, y \in \mathbb{C}(K, \mathbb{R}^d)$ together with the triangle inequality applied to taking the norm of ℓ_0'' (see Equation 3.27) implies

$$\|\ell_0''\|_K \leq \frac{1}{2}\left(\left\|\frac{f_0'}{f_0}\right\|_K^2 \left(2\left\|\frac{\phi_0^2}{f_0}\right\|_K + 1\right) + \left\|\frac{f_0''}{f_0}\right\|_K \left(1 + \left\|\frac{\phi_0^2}{f_0}\right\|_K\right)\right).$$

Applying the Hölder-inequality to the latter equation directly yields $\mathbb{E}[\|\ell_0''\|_K] < \infty$, and N.3 is satisfied.

The only condition that has yet to be verified is N.4. To show that

$$\frac{\partial g_{\boldsymbol{\theta}_2}}{\partial \boldsymbol{\theta}_2}(\phi_0, h_0)|_{\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0}$$

is linearly independent, we have to show that for $\boldsymbol{\tau} = (\gamma_0, \ldots, \gamma_q, \mu_1, \ldots, \mu_p, \omega)^t \in \mathbb{R}^{p+q+2}$

$$\frac{\partial g_{\boldsymbol{\theta}_2}}{\partial \boldsymbol{\theta}_2}(\phi_0, h_0)|_{\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0} \boldsymbol{\tau} = 0,$$

only if $\boldsymbol{\tau} = 0$ (meaning all the components of $\boldsymbol{\tau}$ are zero). Using the a.s. representation (3.18) of $g_{\boldsymbol{\theta}_2}$ gives us

$$\begin{aligned} &\frac{\partial g_{\boldsymbol{\theta}_2}}{\partial \boldsymbol{\theta}_2}(\phi_0, h_0)|_{\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0} = \\ &\Big(1, \big((\kappa^0 + 1)I(\phi_0 < 0) + I(\phi \geq 0)\big)\, \phi_0^2, \ldots, \big((\kappa^0 + 1)I(\phi_{-q+1} < 0) + I(\phi_{-q+1} \geq 0)\big)\, \phi_{-q+1}^2, \\ &\quad h_0, \ldots, h_{-p+1}, \big(\alpha_1^0 I(\phi_0 < 0)\phi_0^2 + \ldots + \alpha_q^0 I(\phi_{-q+1} < 0)\phi_{-q+1}^2\big)\Big) \end{aligned}$$

Assume now that

$$\frac{\partial g_{\boldsymbol{\theta}_2}}{\partial \boldsymbol{\theta}_2}(\phi_0, h_0)|_{\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2^0} \boldsymbol{\tau} = 0. \tag{3.50}$$

Writing out this equation gives us

$$\begin{aligned} &\gamma_0 + \gamma_1 \big((\kappa^0 + 1)I(\phi_0 < 0) + I(\phi_0 \geq 0)\big)\, \phi_0^2 + \mu_1 h_0 + \alpha_1^0 I(\phi_0 < 0)\phi_0^2 \omega + \boldsymbol{Y}_{-1} \\ &= \gamma_0 + h_0 \left(\gamma_1 z_0^2 \big((\kappa^0 + 1)I(z_0 < 0) + I(z_0 \geq 0)\big) + \mu_1 + \omega \alpha_1^0 I(z_0 < 0) z_0^2\right) + \boldsymbol{Y}_{-1} = 0, \end{aligned}$$

where $\boldsymbol{Y}_{-1}$ is a $\mathcal{F}_{-1}$-measurable random variable. Consequently, by virtue of Lemma 4.4 (see appendix) the multiplier of h_0 in the latter identity must be a constant, i.e.

$$\left(z_0^2\left((\kappa^0+1)I(z_0<0)+I(z_0\geq 0)\right)+\mu_1+\omega\alpha_1^0 I(z_0<0)z_0^2\right)=c \quad \text{a.s.} \tag{3.51}$$

for a constant $c\in\mathbb{R}$. The linear independence of $\{1, z_0^2 I(z_0<0), z_0^2\}$ follows because z_0 is, per assumption, not concentrated in two points. In combination with (3.51) and $\alpha_1^0>0$, we get $\gamma_1=\omega=\mu_1-c=0$. This means we have to show the linear independence of $1, \left((\kappa^0+1)I(\phi_0<0)+I(\phi\geq 0)\right)\phi_0^2,\ldots,\left((\kappa^0+1)I(\phi_{-q+1}<0)+I(\phi_{-q+1}\geq 0)\right)\phi_{-q+1}^2$, $h_0,\ldots,h_{-p+1}$.
Notice that $\omega=0$ with Equation (3.50) and the stationarity of the GJR-GARCH model imply

$$\gamma_0+\sum_{i=1}^{q}\gamma_i(1+\kappa^0/2)D_{t-i}^0+\sum_{j=1}^{p}\mu_j h_{t-j}=0 \quad a.s.$$

or equivalently with the lag-operator notation

$$\gamma_0+\gamma(L)(1+\kappa^0)D_t+\mu(L)h_t=0 \quad a.s. \tag{3.52}$$

where $\gamma(z)=\sum\limits_{i=1}^{q}\gamma_i z^i$ and $\mu(z)=\sum\limits_{j=1}^{p}\mu_j z^j$. Recall that $h_0=f_0(\boldsymbol{\theta}_2)$ if and only if $\boldsymbol{\theta}_2=\boldsymbol{\theta}_2^0$. Therefore, using the a.s. representation (3.35) we derived in Lemma 3.10, we can write h_t as follows:

$$h_t=\frac{\alpha_0^0}{b_{\boldsymbol{\theta}_2}^0(1)}+\frac{a_{\boldsymbol{\theta}_2}^0(L)}{b_{\boldsymbol{\theta}_2}^0(L)}D_t^0, \tag{3.53}$$

with $a_{\boldsymbol{\theta}_2}^0(z)=\sum\limits_{i=1}^{q}\alpha_i^0(1+\kappa^0/2)z^i$ and $b_{\boldsymbol{\theta}_2}^0=1-\sum\limits_{j=1}^{p}\beta_j^0 z^j$.
Plugging (3.53) into (3.52) yields

$$\begin{aligned}
&\gamma_0+\gamma(L)\left(1+\kappa^0/2\right)D_t^0+\mu(L)\left(\frac{\alpha_0^0}{b_{\boldsymbol{\theta}_2}(1)}+\frac{a_{\boldsymbol{\theta}_2}^0(L)}{b_{\boldsymbol{\theta}_2}^0(L)}D_t^0\right)\\
&=\gamma_0+\frac{\mu(1)\alpha_0^0}{b_{\boldsymbol{\theta}_2}^0(1)}\left(\gamma(L)\left(1+\kappa^0/2\right)+\mu(L)\frac{a_{\boldsymbol{\theta}_2}^0(L)}{b_{\boldsymbol{\theta}_2}^0(L)}\right)D_t^0=0.
\end{aligned}$$

Using the identical arguments provided in Lemma 3.10, we arrive at

$$\gamma_0+\frac{\mu(1)\alpha_0^0}{b_{\boldsymbol{\theta}_2}^0(1)}=0 \text{ and } \gamma(z)\left(1+\kappa^0/2\right)+\mu(L)\frac{a_{\boldsymbol{\theta}_2}^0(L)}{b_{\boldsymbol{\theta}_2}^0(L)}=0. \tag{3.54}$$

In (3.54), it is $\mu = 0$, because if $\mu \neq 0$, then the rational function $\mu(z)a^0(z)(b^0(z))^{-1}$ would have at least $q+1$ zeros, whereas $\gamma(z)$ has at most q zeros, which is a contradiction. Thus, $\gamma(z) = \mu(z) = 0$ and therefore $\boldsymbol{\tau} = 0$, and N.4 holds.
This completes the proof of Theorem 3.2.

Theorems 3.1 and 3.2 together with the result of Song et al. (2005, Theorem 3) give us that after the kth iteration, the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\theta}}_1^{(k)^t}, \widehat{\boldsymbol{\theta}}_2^{(k)^t})^t$ is consistent and asymptotically normal, meaning

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}^0 \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{V}^{(k)})\right), \text{ as } T \to \infty,$$

where $\boldsymbol{V}^{(k)}$ (the asymptotic covariance matrix) is given in Song et al. (2005). Let $\widehat{\boldsymbol{\theta}}$ be the (final) maximum likelihood estimator of $\boldsymbol{\theta}^0 = (\boldsymbol{\theta}_1^{0^t}, \boldsymbol{\theta}_2^{0^t})^t$. According to Song et al. (2005), when k approaches infinity, then $\boldsymbol{V}^{(k)}$ converges to the covariance matrix of $\widehat{\boldsymbol{\theta}}$.

4 Concluding Remarks

The first part of this thesis dealt with the introduction of the standard GARCH and GJR-GARCH models as well as discussing their weakness with regards to accurate volatility modeling beyond the short term. This is mainly due to the parameter constancy of the standard GARCH models as they are unable to react to changes of the basic structure of the volatility (which may change due to political and economic factors). The main part of the thesis then discussed an extension of the GJR-GARCH model (the TV-GJR-GARCH model) such that time-variation in the conditional and unconditional variance of the model is incorporated. Because the parameter estimation of the model can be numerically very difficult, the consistency and asymptotic normality of the quasi-maximum-likelihood estimator of the TV-GJR-GARCH model was proven in two steps according to the "maximation by parts"-approach of Song et al. (2005).

Although the application of the TV-GJR-GARCH model is beyond the scope of this thesis, the validity and applicability of the theory covered in this thesis can be verified on real data. Amado and Teräsvirta (2013) show through a series of misspecification tests that their time-varying model is notably superior (at least in some contexts) to the mentioned standard models. In particular, their examination on a 2531 observations large data set of daily returns of the S&P 500 points to the presence of long memory in volatility. Consequently, a time-varying model yields much better results with regards to explaining the nonstationary behavior of volatility than models with constant parameters.

Appendix

In this appendix one can find some important definitions and theorems I used in this work, but which were not discussed any further.

Lemma 4.1.
Let $A \in \mathbb{C}^{n \times n}$ be a $n \times n$ matrix with spectral radius $\rho(A)$. If $\lim_{n\to\infty} A^n = 0$, then

$$\rho(A) < 1.$$

Proof. Let μ be an eigenvalue of A with eigenvector $\boldsymbol{v}$, meaning $A^n\boldsymbol{v} = \lambda^n\boldsymbol{v}$ holds. Assume $\lim_{n\to\infty} A^n = 0$, then

$$\begin{aligned}\left(\lim_{n\to\infty} A^n\right)\boldsymbol{v} &= \lim_{n\to\infty}\left(A^n\boldsymbol{v}\right)\\ &= \lim_{n\to\infty}\mu^n\boldsymbol{v}\\ &= \boldsymbol{v}\lim_{n\to\infty}\mu^n\end{aligned}$$

is also equal to zero. Since $\boldsymbol{v}$ is an eigenvector and therefore by definition unequal to zero, $\lim_{n\to\infty}\mu^n = 0$ must hold. This of course implies $|\mu| < 1$, and therefore $\rho(A) < 1$. ☐

Definition 4.2. (ARMA-process)
The process $X = (X_t : t \in \mathbb{Z})$ is called an ARMA(p, q)-process if $\{X_t\}$ is stationary and if for every t,

$$X_t = c + z_t + \sum_{i=1}^{q} a_i X_{t-i} + \sum_{i=1}^{p} b_i z_{t-i},$$

where c is a constant; $a_i, b_i \in \mathbb{C}$; $p, q \in \mathbb{N}$; $a_q \neq 0$, $b_p \neq 0$, and $\{z_t\} \sim WN(0, \sigma_z^2)$.

Theorem 4.3.
The asymmetric GARCH model (AGARCH)

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i^*(|\phi_{t-i}| - \gamma^*\phi_{t-i})^2 + \sum_{j=1}^{p} \beta_j^* h_{t-j}, \tag{4.1}$$

$\alpha_0 > 0, \alpha_i^*, \beta_j \geq 0$ and $|\gamma^*| \leq 1$ can be rewritten to the GJR-GARCH form

$$h_t = \alpha_0 + \sum_{i=1}^{q}(\alpha_i + \lambda_i I(\phi_{t-i} < 0))\phi_{t-i} + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{4.2}$$

if $\alpha_i = \alpha_i^*(\gamma^* - 1)^2$ *and* $\lambda_i = \alpha_i^* 4\gamma^*$.

Proof. It suffices to show that

$$\sum_{i=1}^{q} \alpha_i^*(|\phi_{t-i}| - \gamma^*\phi_{t-i})^2 = \sum_{i=1}^{q} (\alpha_i + \lambda_i I(\phi_{t-i} < 0))\phi_{t-i}^2$$

We will do this in a straightforward manner.

$$\begin{aligned}
\sum_{i=1}^{q} \alpha_i^*(|\phi_{t-i}| - \gamma^*\phi_{t-i})^2 &= \sum_{i=1}^{q} \alpha_i^* \Big((\phi_{t-i}^2 - 2\gamma^*\phi_{t-i}^2 + {\gamma^*}^2\phi_{t-i}^2) I(\phi_{t-i} \geq 0) \\
&\qquad + (\phi_{t-i}^2 + 2\gamma^*\phi_{t-i}^2 + {\gamma^*}^2\phi_{t-i}^2) I(\phi_{t-i} < 0) \Big) \\
&= \sum_{i=1}^{q} \alpha_i^*\phi_{t-i}^2 \Big((1 - 2\gamma^* + {\gamma^*}^2) I(\phi t - i \geq 0) \\
&\qquad + (1 + 2\gamma^* + {\gamma^*}^2) I(\phi_{t-i} < 0) \Big) \\
&= \sum_{i=1}^{q} \alpha_i^*\phi_{t-i}^2 \Big((1 - 2\gamma^* + {\gamma^*}^2)(1 - I(\phi_{t-i} < 0)) \\
&\qquad + (1 + 2\gamma^* + {\gamma^*}^2) I(\phi_{t-i} < 0) \Big) \\
&= \sum_{i=1}^{q} \alpha_i^*\phi_{t-i}^2 \Big((1 - 2\gamma^* + {\gamma^*}^2) + 4\gamma^* I(\phi_{t-i} < 0) \Big) \\
&= \sum_{i=1}^{q} \Big(\alpha_i^*({\gamma^*}^2 - 1)^2 + \alpha_i^* 4\gamma^* I(\phi_{t-i} < 0) \Big) \phi_{t-i}^2
\end{aligned}$$

It is obvious that this is just equal to (4.2) if $\alpha_i = \alpha_i^*(\gamma^* - 1)^2$ and $\lambda_i = \alpha_i^* 4\gamma^*$. This completes the proof. □

Notice that this means that $\lambda_i = \kappa\alpha_i$, where

$$\kappa = \frac{4\gamma^*}{(\gamma^* - 1)^2} \tag{4.3}$$

Lemma 4.4.
Let X be a random variable and $\mathcal{A}$ a σ-algebra. If X is $\mathcal{A}$-measurable and at the same time independent of $\mathcal{A}$, then X is a constant.

Proof. Assume X is not a constant. Then there must exist a $x \in \mathbb{R}$ such that $\mathbb{P}(X \leq x) \in (0, 1)$.

However, it is

$$\mathbb{P}(X \leq x) = \mathbb{P}(\{X \leq x\}, \{X \leq\}) = (\mathbb{P}(X \leq x))^2$$

since $\{X \leq x\} \in \mathcal{A}$ is independent of itself. This is of course a contradiction to $\mathbb{P}(X \leq x) \in (0, 1)$, and X is therefore, indeed, a constant. $\square$

References

AMADO, C. AND T. TERÄSVIRTA (2013): "Modelling volatility by variance decomposition," *Journal of Econometrics*, 175, 142–153.

BLACK, F. (1976): "The pricing of commodity contracts," *Journal of financial economics*, 3, 167–179.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, 31, 307–327.

BOUGEROL, P. AND N. PICARD (1992): "Stationarity of GARCH processes and of some nonnegative time series," *Journal of econometrics*, 52, 115–127.

BROCKWELL, P. J., R. A. DAVIS, AND S. E. FIENBERG (1991): *Time Series: Theory and Methods: Theory and Methods*, Springer Science & Business Media, New York, 2 ed.

CHEN, M. AND H. Z. AN (1998): "A note on the stationarity and the existence of moments of the GARCH model," *Statistica Sinica*, 8, 505–510.

ENDERS, W. AND M. T. HOLT (2014): "The evolving relationships between agricultural and energy commodity prices: a shifting-mean vector autoregressive analysis," in *The Economics of Food Price Volatility*, University of Chicago Press, 135–187.

ENGLE, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica: Journal of the Econometric Society*, 987–1007.

FAN, Y., S. PASTORELLO, AND E. RENAULT (2007): "Maximization by parts in extremum estimation," Working paper, University of North Carolina, Chapel Hill, availalable at `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.636.5451&rep=rep1&type=pdf`.

GLOSTEN, L. R., R. JAGANNATHAN, AND D. E. RUNKLE (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks," *The journal of finance*, 48, 1779–1801.

HORV, L., P. KOKOSZKA, ET AL. (2003): "GARCH processes: structure and estimation," *Bernoulli*, 9, 201–227.

JOHNSON, C. R. AND R. A. HORN (1985): *Matrix analysis*, Cambridge University Press, Cambridge.

KILIÇ, R. (2011): "Long memory and nonlinearity in conditional variances: A smooth transition FIGARCH model," *Journal of Empirical Finance*, 18, 368–378.

LIU, J.-C. (2009): "Stationarity of a family of GARCH processes," *The Econometrics Journal*, 12, 436–446.

MANDELBROT, B. (1963): "The variation of certain speculative prices," *XXXVI: Journal of Business*, 392–417.

MIKOSCH, T. AND C. STĂRICĂ (2004): "Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects," *Review of Economics and Statistics*, 86, 378–390.

NELSON, D. B. (1991): "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica: Journal of the Econometric Society*, 347–370.

NEWEY, W. K. AND D. MCFADDEN (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

RUDIN, W. (1987): *Complex and real analysis*, McGraw-Hill, Inc, 3 ed.

SENTANA, E. (1995): "Quadratic ARCH models," *The Review of Economic Studies*, 62, 639–661.

SONG, P. X.-K., Y. FAN, AND J. D. KALBFLEISCH (2005): "Maximization by parts in likelihood inference," *Journal of the American Statistical Association*, 100, 1145–1158.

STRAUMANN, D. (2005): *Estimation in Conditionally Heteroscedastic Time Series Models*, Springer, Heidelberg.

STRAUMANN, D., T. MIKOSCH, ET AL. (2006): "Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach," *The Annals of Statistics*, 34, 2449–2495.

TETERIN, P., R. BROOKS, AND W. ENDERS (2016): "Smooth volatility shifts and spillovers in US crude oil and corn futures markets," *Journal of Empirical Finance*, 38, 22–36.

TSAY, R. S. (2005): *Financial Time Series*, Wiley Online Library, Hoboken, New Jersey, 2 ed.