GARCH–like Models with Dynamic Crash-Probabilities

Paul Koether

Vom Fachbereich Mathematik der
Technischen Universität Kaiserslautern
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.)
genehmigte Dissertation

Erster Gutachter: Prof. Dr. Jürgen Franke

Zweiter Gutachter : Prof. Dr. Jens–Peter Kreiß

Datum der Disputation: 30.09.2005

D 386

Abstract

We work in the setting of time series of financial returns. Our starting point are the GARCH models, which are very common in practice. We introduce the possibility of having crashes in such GARCH models. A crash will be modeled by drawing innovations from a distribution with much mass on extremely negative events, while in "normal" times the innovations will be drawn from a normal distribution. The probability of a crash is modeled to be time dependent, depending on the past of the observed time series and/or exogenous variables. The aim is a splitting of risk into "normal" risk coming mainly from the GARCH dynamic and extreme event risk coming from the modeled crashes.
We will present several incarnations of this modeling idea and give some basic properties like the conditional first and second moments. For the special case that we just have an ARCH dynamic we can establish geometric ergodicity and, thus, stationarity and mixing conditions. Also in the ARCH case we formulate (quasi) maximum likelihood estimators and can derive conditions for consistency and asymptotic normality of the parameter estimates.
In a special case of genuine GARCH dynamic we are able to establish L_1 -approximability and hence laws of large numbers for the processes itself. We can formulate a conditional maximum likelihood estimator in this case, but cannot completely establish consistency for them.
On the practical side we look for the outcome of estimating models with genuine GARCH dynamic and compare the result to classical GARCH models. We apply the models to Value at Risk estimation and see that in comparison to the classical models many of ours seem to work better although we chose the crash distributions quite heuristically.

Acknowlegement

Many thanks to my supervisor Prof. Jürgen Franke for providing the topic and much support. Thanks to Dr. Jean Pierre Stockis for critical comments, support and encouragement. I'm indebted to my brothers in arms Dr. Joseph Tadjuidje Kamgaing and Dr. Charles Andoh for many mathematical discussions and mutual help. Many thanks to Mrs. Beate Siegler the secretary of our working group for friendliness and support in non mathematical matters. The financial and logistic support of the Graduiertenkolleg Mathematik und Praxis is highly appreciated.

Contents

Introduction

Motivation and modeling idea

GARCH models, especially the GARCH(1,1) model, are widely used in practice though some shortcomings are known. Let's consider e.g. a GARCH(1,1) model with normal distribution of the innovations.

$$X_t = \sigma_t \epsilon_t \text{ with } \sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2$$

with $\mathcal{L}(\epsilon_t) = N(0,1)$. When fitting such a model to financial return time series frequently values for α around 0.2 and for β around 0.8 arise. The fact that the whole dynamic is captured via modeling the volatility together with the relatively high value of β means that the effects of shocks are quite persistent in such a model. If we now have an isolated extreme event in a period of otherwise relatively low volatility the modeled volatility will tend to go down too slowly after this event. Now if we fix the Value at Risk as a risk measure to investigate, it is in this model proportional to the volatility. This is one of the reason that while using GARCH(1,1) e.g. the 5% Value at Risk seems to be a little bit pessimistic. On the other hand financial returns seem to be more heavy tailed than the normal innovations are able to mirror: the 1% Value at Risk is often overly optimistic. The latter problem is handled by practitioners often by using innovations following a t-distribution with few degrees of freedom. But in many cases this makes the situation at 5% level even worse.

In order to improve GARCH models in a new way we had the idea of modeling:

$$X_t = \sigma_t[(1 - B_t)\epsilon_t + B_t D_t] \text{ with } \sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2$$

where ϵ_t shall be the innovations in normal times and D_t is drawn from a distribution suited to model innovations in case of extreme events like crashes and defaults. The modeling of the occurrence of a crash is via the variables B_t with $\mathcal{L}(B_t|\mathfrak{F}_{t-1}) = \mathfrak{B}(1, p_t)$. Here p_t is the crash probability at the time t. We will let p_t depend on the past in various ways. The aim is to split risk into "normal risk" arising more or less from higher or lower volatilities and an extreme risk coming from really extreme events. This yields a more detailed description of risk. Further, when going back to the Value at Risk, we can expect that the combination of these two effects will expose nonlinearities by which we get a more balanced modeling of the Value at Risk on different levels.

Outline

In Chapter 1, we first give an overview of different models in the GARCH context. We start with the classic ones and then give some examples for more sophisticated concepts, e.g. GARCH with Markov switching in the parameters. We do so, that firstly the framework we are working in is specified and secondly to make clear what kind of theory is desirable for the models we will introduce.
Moreover in section 1.6 we give some features of models where we apply GARCH equations to innovations which are neither centered or scaled. Some of the features presented there are essentially present in [Nel90], others seem to be new.
In Chapter 2, we present a first GARCH like model class with crash probabilities depending on the past. We give some basic properties, discuss the problem of really modeling crashes, give an example and discuss slight modifications of the model, e.g. dependence of the crash probabilities on exogenous variables.
In Chapter 3, we discuss models which while keeping the basic spirit deviate more strongly from the model we first introduced. Again we explore basic properties. The discussion also gives a different view of our first model.
In Chapter 4, we discuss asymptotic properties like stationarity and mixing. The results only hold if we restrict ourselves to a pure ARCH dynamic. The discussion is done for the model we first introduced and the new model which has the most differences among the models from Chapter 2.
In Chapter 5, we give a short introduction to the methodology from the book of Pötscher and Prucha (1997) with a view towards consistency theory, because we will make heavy use of this methodology.
Again restricting ourselves to the models with pure ARCH dynamic we develop in Chapter 6 consistency and asymptotic normality results for different Quasi Maximum Likelihood estimators for the original and the alternative model.
Chapter 7 is dedicated to what happens asymptotically if we use genuine GARCH dynamic. At least in the restricted case of slightly altered models we get laws of large numbers for the processes.
In Chapter 8, we want to gain asymptotic properties for the Quasi Maximum Likelihood Estimators of the processes from Chapter 7. We are able to do some steps in that direction, but do not succeed completely. We get formal problems with the estimators because we would have to define an appropriate metric on the tuples of observations from the infinite past in order to verify continuity conditions.
Chapters 9 to 11 are dedicated to the more practical sides of the problem.
In Chapter 9, we have a look on the outcome of estimating log-return time series of several of our models, which only depend on their own past, in hindsight of how the crash probability and the volatility in these models behave. We also discuss the problem in how far reality is mirrored when we simulate from models we estimated from real world data.

This discussion is continued in Chapter 10 for models which additionally or exclusively depend on exogenous variables.
In Chapter 11, we compare the models how they work as Value at Risk estimators.

Chapter 1

GARCH processes

1.1 Financial Data

We are interested in time-series of financial data. Rather than the observed prices we investigate the so called returns.
Let S_t be the Price at time t.

Definition 1.1.1.

1. *The Return at time t is defined as*
   $R_t = \frac{S_t - S_{t-1}}{S_{t-1}}$

2. *The Log–Return at time t is defined as*
   $r_t = \log(\frac{S_t}{S_{t-1}}) = \log(S_t) - \log(S_{t-1})$.

Now the observed returns of assets seem to have some properties in common: (So called "stylized facts"):

1. Returns seem to be uncorrelated.

2. But they are not independent, because the squared returns seem to be correlated.

3. There seem to be "quiet" and "nervous" periods.(Volatility clustering)

4. The distribution of returns have more mass around the mean and in the tails than the Gaussian–distribution.(Leptokurtic distribution, "heavy tails")

To capture these properties several time-series models were developed. We focus here on ARCH and GARCH like models, in which we want to implement a kind of possibility of a crash for reasons we will discuss later.

1.2 Models

First we introduce the most important models already in use. In the following, we will state their theoretical properties.

Notation We consider a time series X_t the sigma algebras in the corresponding filtration we denote as $\mathfrak{F}_t$. Later, when using models with additional exogenous variables, $\mathfrak{F}_t$ denotes the whole information up to time t.

1.2.1 ARCH–models

A time-series X_t is said to follow an ARCH(q) model if the following equations hold.

$$E(\epsilon_t|\mathfrak{F}_{t-1}) = 0 \text{ and } Var(\epsilon_t|\mathfrak{F}_{t-1}) = 1$$

$$X_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i X_{t-i}^2$$

where the $\alpha_i > 0$ and $\omega > 0$.

1.2.2 GARCH–models

A time-series X_t is said to follow a GARCH(p,q) model if the following equations hold.

$$E(\epsilon_t|\mathfrak{F}_{t-1}) = 0 \text{ and } Var(\epsilon_t|\mathfrak{F}_{t-1}) = 1$$

$$X_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i X_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

where the $\alpha_i > 0$,$\beta_i > 0$ and $\omega > 0$.
Very common in practice is the GARCH(1,1) model:

$$X_t = \sigma_t \epsilon_t$$

with

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2$$

where the ϵ_t are iid with mean 0 and variance 1. The distribution of ϵ_t is often chosen to be standard normal.

1.2.3 ARMA–GARCH and GARCH in mean

GARCH models are designed to model conditional heteroskedasticy in the variance. But for economic modeling scientists as well as appliers of the models in practice are also interested in the mean of the returns. So besides fitting a GARCH model to the centered observations, models have been developed to combine GARCH regimes in the variance with time series models for the mean. Two of the most important types of these model classes are ARMA–GARCH models and GARCH–M models.

ARMA–GARCH
The ARMA–GARCH model is an ARMA–model with GARCH errors. By that we mean that an ARMA(n,m)–GARCH(p,q) model is defined by

$$Y_t = \nu + \sum_{i=1}^{n} a_i Y_{t-i} + \sum_{j=1}^{m} b_i X_{t-j} + X_t$$

where X_t is a GARCH(p,q) process.

GARCH–M
Y_t is said to be be a GARCH in mean or GARCH–M process if

$$Y_t = \nu + \lambda g(\sigma_t^2) + X_t$$

holds, where X_t is a GARCH process and g is a known parametric function. The CAPM, e.g., corresponds to the choice $g = id$. The term $\lambda g(\sigma_t^2)$ is interpreted as a risk premium.

1.2.4 T–GARCH and E–GARCH

The following two models allow for asymmetric dependence of volatility on past returns. They start from the common ARCH- and GARCH-models, but use different specifications of σ_t.

Threshold ARCH Here the equation for σ_t is given by

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i |X_{t-i}|^\delta + \sum_{i=1}^{q} \overline{\alpha}_i |X_{t-i}^\delta| I(X_{t-i} < 0)$$

There are models which use $\delta = 1$ and models with $\delta = 2$.

Exponential GARCH Here the equation for σ_t is given by

$$\log(\sigma_t^2) = \omega_t + \sum_{i=1}^{\infty} \beta_k g(Z_{t-k})$$

and

$$g(Z_t) = \theta Z_t + \gamma(|Z_t| - E(|Z_t|))$$

where Z_t are iid random variables with mean zero and variance 1.

1.3 GARCH models with Markov–switching

Let ϵ_t be iid with zero mean and unit variance and Δ_t be a Markov chain with finite state space $\mathfrak{E} = \{1, 2, \ldots, d\}$ The model is given by:

$$X_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \omega(\Delta_t) + \sum_{i=1}^{q} \alpha_i(\Delta_t) X_{t-i}^2 + \sum_{i=1}^{p} \beta_i(\Delta_t) \sigma_{t-i}^2$$

where for all $k \in \mathfrak{E}$, $\alpha_i(k)$ and $\beta_i(k)$ are nonnegative and $\omega(k)$ is positive.

1.4 Estimation and theoretical properties

We have to investigate the theoretical properties of the models. The question of stationarity is here of utmost importance. The other important question is if there are consistent and asymptotically normal estimators. Usually the model parameters discussed here are estimated via conditional maximum likelihood estimators.

1.4.1 GARCH(p,q)

Stationarity, Representation

Theorem 1.4.1. *Let X_t be a GARCH(p,q) process with $E(X_t^4) = c < \infty$ then*

1. $\zeta_t := \sigma_t^2(\epsilon_t^2 - 1) = X_t^2 - \sigma_t^2$ *is white noise*
2. X_t^2 *is an ARMA(m,p) process with*

$$X_t^2 = \omega + \sum_{i=1}^{m} \gamma_i X_{t-i} - \sum_{j=1}^{p} \beta_j \zeta_{t-j} + \zeta_t$$

where $m = max(p, q)$, $\gamma_i = \alpha_i + \beta_i$ and $\alpha_i = 0$ if $i > q$, $\beta_i = 0$ if $i > p$

The condition $E(X_t^4) = c < \infty$ is just needed for ζ_t having a finite variance. The representation formula in Part 2 of the above theorem holds without this condition.
To investigate stationarity we must either consider a starting distribution in an equilibrium state, double infinite sequences or we can just talk about asymptotic stationarity. Unfortunately the diverse theorems are best stated in various of these contexts.

Theorem 1.4.2.

1. *A GARCH(p,q) process with nonnegative coefficients ω, α_i, β_j is asymptotically second order stationary if*

$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$$

[Gou97]

2. *On the other hand if a second order stationary process has an existing constant variance σ^2, then*

$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$$

holds and

$$\sigma^2 = \frac{\omega}{1 - (\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i)}$$

[FHH01]

3. *Consider a GARCH(1,1) process which is double infinite.Let the ϵ_t be iid. If $E(ln(\beta + \alpha\epsilon_t)) < 0$ then X_t is strictly stationary.[Nel90]*

4. *[BP92b] Let A_n be defined as the $(p+q-1) \times (p+q-1)$ matrix*

$$\begin{pmatrix} \tau_n & \beta_p & \alpha & \alpha_q \\ I_{p-1} & 0 & 0 & 0 \\ \zeta_n & 0 & 0 & 0 \\ 0 & 0 & I_{q-2} & 0 \end{pmatrix}$$

with

$$\tau_n = (\beta_1 + \alpha_1\epsilon_n^2, \beta_2, \dots, \beta_{p-1}) \in R^{p-1}$$
$$\eta_n = (\epsilon_n^2, 0, \dots, 0) \in R^{p-1}$$
$$\alpha = (\alpha_2, \dots, \alpha_{q-1}) \in R^{q-2}$$

If $\omega > 0$ the GARCH(p,q) equation has a strictly stationary solution if and only if the top Lyapounov exponent $\gamma = \inf\{E(\frac{1}{n}\log||A_nA_{n-1}\dots A_1||), n \in N\}$ associated with the matrices $\{A_n, n \in \mathbf{Z}\}$ is strictly negative. Moreover this strictly stationary solution is ergodic.For given ϵ_t's it is unique.

Consistency of the maximum likelihood estimator

Usually GARCH models are estimated via the maximum likelihood method. Because the initial distribution of the time series is unknown, usually the term belonging to it is skipped. The estimator arising in such a way is called conditional maximum likelihood estimator. When using conditionally normal distributions in the estimator while not requiring the data arising from such a distribution, the estimator is called a quasi maximum likelihood estimator.

Theorem 1.4.3. *[LH94] Consider a GARCH(1,1) model with true parameters ω,α,β and rescaled variable ϵ_t. If*

1. *ϵ_t is strictly stationary and ergodic*
2. *ϵ_t^2 is nondegenerate*
3. *for some $\delta > 0$ there exists $S_\delta < \infty$ such that $E(\epsilon_t^{2+\delta}|\mathfrak{F}_{t-1}) \leq S_\delta$ a.s.*
4. *$\sup_t E(\log(\beta + \alpha\epsilon_t^2)|\mathfrak{F}_{t-1}) < 0$*

then the quasi maximum likelihood estimator of a GARCH(1,1) model restricted to any compact parameter subspace is consistent. If moreover $\alpha + \beta < 1$ the quasi maximum likelihood estimator with no restricted parameter space is consistent.

Theorem 1.4.4. *[LH94] Under the assumptions of the previous theorem and additionally*

1. *$E(\epsilon_t^4|\mathfrak{F}_{t-1}) < \infty$*
2. *The true parameter θ_0 is in the interior of the parameter space.*

The QMLE estimator $\hat{\theta}_n$ is asymptotically normal i.e.

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_D N(0, V_0)$$

with $V_0 = B_0^{-1} A_0 B_0^{-1}$ where $A_0 = E(\nabla l_t(\theta_0) \nabla l_t(\theta_0)')$ and $B_0 = -E(\nabla^2 l_t(\theta_0))$. Here l_t denotes the logarithm of the quasi likelihood function.

In the case of an iid random source [BHK03] give us a theorem for the consistency of the QMLE in the GARCH(p,q) case for general p,q.

Theorem 1.4.5. *[BHK03] Consider a GARCH(p,q) model with iid random source ϵ_t Let $\mathfrak{A}(x) = \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_p x^p$ and $\mathfrak{B}(x) = 1 - \beta_1 x - \beta_2 x^2 - \cdots - \beta_q x^q$. Let $\theta_0 = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ be the true parameter. Let us parameterize the problem via* **u**$= (x, s_1, \ldots, s_p, t_1, \ldots, t_q)$. *If*

1. *ϵ_t^2 is nondegenerate*

2. $E(\epsilon_t^2) = 1$

3. $E|\epsilon_t^2|^{1+\delta} < \infty$ *for some* $\delta > 0$

4. $\lim_{s \to 0} s^{-\mu} P(\epsilon_t^2 \leq s) = 0$ *for some* $\mu > 0$

5. *The polynomials* $\mathfrak{A}(x)$ *and* $\mathfrak{B}(x)$ *are coprimes in the set of polynomials with real coefficients.*

6. θ_0 *is in the interior of*

$$U = \{\mathbf{u} | t_1 + t_2 + \cdots + t_q \leq \rho_0 \text{ and } \underline{u} \leq \min(x, s_1, s_2, \ldots, s_p, t_1, t_2, \ldots, t_q) \\ \leq \max(x, s_1, s_2, \ldots, s_p, t_1, t_2, \ldots, t_q) \leq \overline{u}\}$$

with $0 < \underline{u} < \overline{u}, 0 < \rho_0 < 1, q\underline{u} < \rho_0$

Then the quasi maximum likelihood estimator converges almost surely towards the true value θ_0.

The following is a generalization of the asymptotic normality theorem of GARCH(1,1) to GARCH(p,q) in the case of an iid random source.

Theorem 1.4.6. *Define* $A_0 = E(\nabla l_t(\theta_0) \nabla l_t(\theta_0)')$ *and* $B_0 = -E(\nabla^2 l_t(\theta_0))$. *Assume the conditions of the theorem above are satisfied and additionally:* $E(|\epsilon_t^2|^{2+\delta}) < \infty$ *for some* $\delta > 0$, *then the following holds:*

1. A_0 *and* B_0 *are nonsingular*

2. $\sqrt{n}(\hat{\theta_n} - \theta_0) \to_D N(0, B_0^{-1} A_0 B_0^{-1})$ *as* $n \to \infty$.

1.4.2 The Markov switching model

Strong stationarity

We now describe some properties of the model introduced in 1.3. In [FRZ01] conditions for existence of a stationary solution of the Markov switching model are given. For this purpose they rewrite the model in matrix form.

$$\underline{\sigma}_t^2 = \underline{\omega}_t + A_t \underline{\sigma}_{t-1}^2$$

with

$$\underline{\omega}_t = \underline{\omega}_t(\Delta_t) = \begin{pmatrix} \omega_t(\Delta_t) \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}$$

and

$$\underline{\sigma}_t^2 = \begin{pmatrix} \sigma_t^2 \\ \sigma_{t-1}^2 \\ . \\ . \\ . \\ \sigma_{t-r}^2 \end{pmatrix}$$

and

$$A_t = \begin{pmatrix} \alpha_1(\Delta_t)\epsilon_{t-1}^2 + \beta_1(\Delta_t) & \alpha_2(\Delta_t)\epsilon_{t-2}^2 + \beta_2(\Delta_t) & \dots & \alpha_r(\Delta_t)\epsilon_{t-r}^2 + \beta_r(\Delta_t) \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \dots & \ddots & \ddots & 0 \\ 0 & \dots & 1 & 0 \end{pmatrix}$$

where A_t is an $r \times r$ matrix and $\alpha_i(\Delta_t)$ and $\beta_j(\Delta_t)$ are equal to zero for $i > q$ and $j > 0$ and where $r = max(p, q)$.

Theorem 1.4.7. *Whenever the top Lyapounov exponent*

$$\gamma = \inf\{E(\frac{1}{n}\log ||A_n A_{n-1} \dots A_1||), n \in N\} < 0$$

there exists a unique strictly stationary solution of the above model which is given by

$$\underline{\sigma}_t^2 = \underline{\omega}_t + \sum_{i=1}^{\infty} A_t A_{t-1} \dots A_{t-i+1} \underline{\omega}_{t-i}$$

In [FRZ01] a condition for second order stationarity is also given including a formula for the unconditional variance:

Theorem 1.4.8. *Let $\overline{A}(\Delta_t)$ be the matrix obtained by replacing ϵ_{t-i}^2 by 1 in A_t. Let $p(i, j)$, $(i, j) \in \{1, \dots, d\}^2$ denote the probability of changing from state i into state j. Define the following $dr \times dr$ matrix.*

$$P = \begin{pmatrix} p(1,1)\overline{A}(1) & p(2,1)\overline{A}(1) & \dots & p(d,1)\overline{A}(1) \\ p(1,2)\overline{A}(2) & p(2,2)\overline{A}(2) & \dots & p(d,2)\overline{A}(2) \\ \vdots & & & \\ p(1,d)\overline{A}(d) & p(2,d)\overline{A}(d) & \dots & p(d,d)\overline{A}(d) \end{pmatrix}$$

If the spectral radius of P is strictly less than 1 then the Markov switching GARCH model has a unique stationary solution, belonging to L_2. The unconditional variance of this process is:

$$\frac{1}{r}\sum_{k=0}^{\infty} 1' P^k \underline{\omega}$$

with $1'$ *being the* $1 \times dr$ *unit vector and* $\underline{\omega} = (\pi(1)\underline{\omega}(1)', \ldots, \pi(d)\underline{\omega}(d)')' \in R^{dr\times 1}$

Consistency in the pure ARCH case

To achieve consistency for the maximum likelihood estimator [FRZ01] focuses on the pure ARCH case. That is due to the fact that they use Markov chain methodology, namely a backward forward algorithm, which restricts them to this case. They also use some identifiability condition.

1.5 Methods of the Theory

The proof technique used in [BP92a] and [BP92b] for stationarity surveys demand the random source ϵ_t has to be iid. So this method cannot be used to investigate the properties of the models introduced in the next chapter. Because there the role of the ϵ_t is taken by a process far from being iid in order to model "normal" and crash regimes. In certain cases the use of the right law of large numbers may transfer the methods used here to our problem.

1.6 Non normalized GARCH

The results in this section are auxiliary results for the models we really want to investigate. Although the results are pretty straightforward, they don't seem to be described in present literature. So we choose to present them here. In fact we want to consider a process $X_t = \sigma_t\epsilon_t$ where σ_t^2 follows a GARCH–equation and the ϵ_t are iid, but we don't consider any restriction on the mean or variance of the ϵ_t except them being finite.

Definition 1.6.1. *A process* X_t *is called a non normalized–GARCH(p,q)–process with respect to* ϵ_t *if*

1. $X_t = \sigma_t\epsilon_t$ *and* $\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{i=1}^p \beta_i\sigma_{t-i}^2$.
2. *The* $\{\epsilon_t\}$ *are iid random variables with* $0 < c = E\epsilon_t^2 < \infty$.

It is clear that for a non normalized GARCH process $E(X_t|\mathfrak{F}_{t-1}) = \sigma_t E(\epsilon_t)$ and $E(X_t^2|\mathfrak{F}_{t-1}) = \sigma_t^2 E(\epsilon_t^2)$ holds. So $Var(X_t|\mathfrak{F}_{t-1}) = \sigma_t^2(E(\epsilon_t^2) - E(\epsilon_t)^2) = \sigma_t^2 Var(\epsilon_t)$.

Theorem 1.6.2. *An unconditional non normalized–GARCH(1,1)–process with* $\beta > 0$ *is strictly stationary iff* $E(\log(\beta + \alpha\epsilon_t)) < 0$.

Proof. In fact $\{\epsilon_t\}$ is iid by definition, the assumptions on ϵ_t imply that ϵ_t^2 is nondegenerate and the probability of the absolute value of ϵ_t being ∞ is zero. The condition $\beta > 0$ implies the existence of $E(\log(\beta + \alpha\epsilon_t))$ by the remark of [Nel90] following condition (5) of this paper. So we are in the context of [Nel90] which drops the condition of ϵ_t being scaled and centered for just $E(\log(\beta+\alpha\epsilon_t))$ existing. (Condition (5) of [Nel90]). $\square$

Theorem 1.6.3. *Let X_t be a non normalized–GARCH(1,1)–process with $c = E\epsilon_t$.*

1. *If $c \geq 1$ then X_t is second order stationary iff $\alpha c + \beta < 1$.*
2. *If $c < 1$ then X_t is second order stationary iff $\alpha\sqrt{c} + \beta < 1$.*

Proof. By Jensen's inequality we have

$$E\log(\beta + \alpha\epsilon_t) \leq \log(\beta + \alpha E\epsilon_t) \leq \log(\beta + \alpha\sqrt{c})$$

as $|E\epsilon_t| \leq E|\epsilon_t| \leq \sqrt{E\epsilon_t^2} = \sqrt{c}$. In both cases 1 and 2, the condition on α, β, c guarantees, therefore, that $E\log(\beta + \alpha\epsilon_t) < 0$ and X_t is strictly stationary by Theorem 1.6.2. So, second order stationarity follows if the second moment exists. In that case, we have

$$\begin{aligned} EX_t^2 &= E\sigma_t^2 E\epsilon_t^2 = c(\omega + \alpha EX_{t-1}^2 + \beta E\sigma_{t-1}^2) \\ &= \omega c + (\alpha c + \beta)EX_{t-1}^2 \\ &= \omega c + (\alpha c + \beta)EX_t^2 \end{aligned}$$

by stationarity. We conclude

$$EX_t^2 = \frac{\omega c}{1 - (\alpha c + \beta)}$$

Therefore in case of strict stationarity a necessary and sufficient condition for second order stationarity is $\alpha c + \beta < 1$ which in case 2 follows from $\alpha\sqrt{c} + \beta < 1$. $\square$

Remark 1.6.4. For $c < 1$, we can construct examples where $\alpha c + \beta < 1$ but $E\log(\beta + \alpha\epsilon_t) > 0$. Consider, e.g., for some $0 < w < 1$, $0 < z < 1$ a two point distribution with $\epsilon_t \in \{-z, z\}$, $pr(\epsilon_t = z) = w$. Then, $E\epsilon_t^2 = c = z^2$ and

$$E\log(\beta + \alpha\epsilon_t) = w\log(\beta + \alpha z) + (1 - w)\log(\beta - \alpha z)$$

provided $\alpha z < \beta$. Now, choose α, β, z such that $\beta + \alpha z > 1$ but $\beta + \alpha z^2 = \beta + \alpha c < 1$ and then choose w close enough to 1 such that $E\log(\beta + \alpha\epsilon_t) > 0$.

We define $s := E(\sigma_t)$.
It is of interest how non normalized GARCH and normalized GARCH processes are related.

Lemma 1.6.5. *Let X_t be a non normalized GARCH(1,1) process with respect to ϵ_t with parameters ω, α, β . Then $Y_t := X_t - E(X_t|\mathfrak{F}_{t-1})$ is a non normalized GARCH(1,1) process with respect to $\epsilon_t - E(\epsilon_t|\mathfrak{F}_{t-1}) = \epsilon_t - E(\epsilon_t)$ with parameters $\omega, \alpha, \beta + \alpha e^2$, where $e := E(\epsilon_t)$.*

Proof. We have on one hand $Y_t = \sigma_t\epsilon_t - E(\sigma_t\epsilon_t|\mathfrak{F}_{t-1}) = \sigma_t\epsilon_t - \sigma_t E(\epsilon_t) = \sigma_t(\epsilon_t - e)$. On the other hand we have:

$$\begin{aligned}\sigma_t^2 = \omega + \alpha X_{t-1} + \beta\sigma_{t-1}^2 &= \omega + \alpha\sigma_{t-1}^2\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ &= \omega + \alpha\sigma_{t-1}^2(\epsilon_{t-1}^2 - e^2 + e^2) + \beta\sigma_{t-1}^2 \\ &= \omega + \alpha\sigma_{t-1}^2(\epsilon_{t-1}^2 - e^2) + \alpha\sigma_{t-1}^2 e^2 + \beta\sigma_{t-1}^2 \\ &= \omega + \alpha Y_{t-1}^2 + \alpha\sigma_{t-1}^2 e^2 + \beta\sigma_{t-1}^2 \\ &= \omega + \alpha Y_{t-1}^2 + (\beta + \alpha e^2)\sigma_{t-1}^2\end{aligned}$$

So Y_t is a non normalized GARCH(1,1) process with respect to $\epsilon_t - e$ with parameters $\omega, \alpha, \beta + \alpha e^2$. ☐

Theorem 1.6.6. *1. Let $E(X_t|\mathfrak{F}_{t-1}) = 0$, i.e. $E(\epsilon_t) = 0$ and define $c := E(\epsilon_t^2)$, then*

(a) X_t is a non normalized GARCH(1,1) process with respect to ϵ_t.

(b) X_t is a GARCH process with residuals $\frac{1}{\sqrt{c}}\epsilon_t$.

are equivalent. Moreover if X_t is a non normalized GARCH(1,1) process with parameters ω, α, β then it is a GARCH(1,1) process with parameters $c\omega, c\alpha, \beta$, and if X_t is an GARCH process with residuals $\frac{1}{\sqrt{c}}\epsilon_t$ and parameters ω, α, β it is a non normalized GARCH process with respect to ϵ_t with parameters $\frac{1}{c}\omega, \frac{1}{c}\alpha, \beta$.

2. Let X_t be a non normalized GARCH(1,1) process with parameters ω, α, β, then $\zeta_t := X_t - E(X_t|\mathfrak{F}_{t-1})$ is a GARCH process with parameters $c\omega, c\alpha, \beta + \alpha e^2$ where $c = Var(\epsilon_t)$ and $e = E\epsilon_t$. Moreover X_t is a special case of a GARCH–M model, namely $X_t = \sigma_t E\epsilon_t + Y_t$ for a GARCH(1,1) process Y_t.

Proof. 1. a $\to$ b Let $X_t = \sigma_t\epsilon_t$ be an non normalized GARCH process with $E(X_t|\mathfrak{F}_{t-1}) = 0$ let $E(\epsilon_t^2) = c$. Then $Var(X_t|\mathfrak{F}_{t-1}) = c\sigma_t^2$. Then

$$Var(X_t|\mathfrak{F}_{t-1}) = c\omega + c\alpha X_{t-1}^2 + c\beta\sigma_{t-1}^2$$

holds. But this equals

$$c\omega + c\alpha X_{t-1}^2 + \beta(c\sigma_{t-1}^2) = c\omega + c\alpha X_{t-1} + \beta Var(X_{t-1}|\mathfrak{F}_{t-2})$$

So X_t is a GARCH process with parameters $c\omega, c\alpha, \beta$.

b $\to$ a holds when we set $\sigma_t^2 := \frac{1}{c}Var(X_t|\mathfrak{F}_{t-1})$ and applying the same argumentation as above the GARCH equation for $Var(X_t|\mathfrak{F}_{t-1})$ taking the role of the non normalized GARCH equation above with $\frac{1}{c}$ taking the role of c above.

2. (a) If X_t is a non normalized GARCH(1,1) process the previous lemma shows that $Y_t := X_t - E(X_t|\mathfrak{F}_{t-1})$ is a non normalized GARCH(1,1)process with $E(Y_t|\mathfrak{F}_{t-1}) = 0$. Its parameters are $\omega, \alpha, \beta + \alpha e^2$. Part one of this theorem yields that Y_t is in fact a GARCH(1,1) process with parameters $c\omega, c\alpha, \beta + \alpha e^2$.

   (b) This follows from the previous part realizing that $E(X_t|\mathfrak{F}_{t-1}) = \sigma_t E\epsilon_t$. □

The first part of the theorem shows that in the centered case non normalized GARCH is up to scale nothing really new. The second part of the theorem shows also that a non normalized GARCH process is in fact a special case of a GARCH–M process, with the constant term for the conditional mean being zero and the factor for σ_t being $E(\epsilon_t)$. But clearly the way from non normalized GARCH to the GARCH regime of the centered variable cannot be reversed. An appropriate generalization towards GARCH(p,q) is also true.

So speaking more formally the argument is that if we have $Var(\epsilon_t^o) = 1$ and $\epsilon_t = \gamma\epsilon_t^o$ with $\gamma^2 = c$ and further define $\overline{\sigma_t} = \gamma\sigma_t$ where σ_t satisfies a GARCH equation with parameters ω, α, β then $X_t := \sigma_t\epsilon_t = \overline{\sigma_t}\epsilon_t^o$ and also $\overline{\sigma_t}^2 = c{\sigma_t}^2 = c\omega + \sum c\alpha_i X_{t-i} + \sum c\beta_j \frac{\overline{\sigma_t}^2}{c}$.
Define $\epsilon_t^o = \frac{\epsilon_t - e}{\gamma}$ where $e = E(\epsilon_t)$ and γ is the square root of its variance. So here we have $\epsilon_t = e + \gamma\epsilon_t^o$ with ϵ_t^o being a $(0,1)$ variable. If we define again $\overline{\sigma_t} = \gamma\sigma_t$ we get:

$$X_t = \sigma_t\epsilon_t = e\sigma_t + \gamma\sigma_t\epsilon_t^o = \frac{e}{\gamma}\overline{\sigma_t} + \overline{\sigma_t}\epsilon_t^o$$

Now we want to gain more generality:

Definition 1.6.7. *A generalized GARCH process is a process* $X_t = \sigma_t\epsilon_t$ *and* $\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$ *where the innovation process* ϵ_t *is strictly stationary and ergodic with and* $Var(\epsilon_t) < \infty$ *.*

Clearly the theorems above hold also for GARCH–transformed processes, as the ϵ_t are square integrable.
Now we want to switch to the models we want to discuss. In fact some of them will turn out to be generalized GARCH allowing us to use the theorems stated here.

Chapter 2

A first model with probability of a crash

We want to introduce a possibility of a crash into models of the GARCH world. Why do we want to do this ? In spite of being very popular with practitioners, there are some weak spots in standard GARCH, which usually arise from extreme events. Firstly the residuals after fitting a GARCH model to time series of real world financial returns are still heavy tailed. But the most popular assumed distribution of these residuals is still the standard normal. So if interested in extreme risk, meaning in the world of Value at Risk extreme quantiles, there will be a systematic underestimation of the risk.

2.1 Definition of the model

Our first, rather naive idea is to model crashes with a “crash distribution” and let the probability of a crash depend on the past history.

$$X_t = \begin{cases} \sigma_t \epsilon_t \text{ with probability } 1 - p_t \\ \sigma_t D_t \text{ with probability } p_t \end{cases} \tag{2.1}$$

where D_t is a random variable modeling the crash behavior. We further assume that the $\{\epsilon_t\}$ are iid. and the $\{D_t\}$ are iid. both independent from each other. p_t is the probability of a crash. We choose to model it as a function of the process and the volatility of the preceeding time–step: $p_t = f(X_{t-1}, \sigma_t)$, e.g. as a logistic function:

$$f(x, s) = \frac{1}{1 + \exp(-(\alpha_0 + \alpha_1 x + \alpha_3 s))}$$

Later on we assume a GARCH(1,1) recursion

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2$$

but the first results hold in general. The model with GARCH recursion on σ_t^2 we will call the CGARCH-S model where C stands for "crash" and S symbolizes that we model σ_t^2 via the GARCH equation in order to distinguish it from some models we will introduce in Chapter 2.
It might be convenient to write this model in a closed form:

$$X_t = \sigma_t((1 - B_t)\epsilon_t + B_t D_t)$$

where the B_t are $\mathfrak{B}(1, p_t)$ distributed.
In the case that ϵ_t and D_t have zero mean and have up to scale the same distribution, and p_t is constant, this would fall into the framework of the model with Markov switching GARCH coefficients. Namely in such a model with two states where all parameters of the second state are multiples of the ones of the first state and the probability to switch between the states are equal no matter which state we are in. So while there is an nonempty intersection between these model classes, we see that the concepts are rather different, such that only very special cases of both classes belong to the intersection.
We remark that in model (2.1) and all subsequent crash models we have in contrast to the common GARCH methodology, two random variables representing risk: σ_t corresponds to the usual notion of volatility though not necessarily coinciding with the conditional standard deviation (compare 2.2.2). It represents the change between more or less volatile market phases and the market risk in "normal" times. p_t however represents the risk of extreme price changes, i.e. market risk arising from extraordinary situations. This allows for a more detailed description of risk, separating normal risk from extreme risk. In this sense, our approach differs from the standard methods of capturing the extreme risk behavior in the bounds of the GARCH world, which just replaces the normal distribution of innovations ϵ_t by a heavy tailed one.

2.2 Expected Value, Variance, Covariances

2.2.1 Expected value

Propositition 2.2.1. *1.*

$$\begin{aligned} E(X_t|\mathfrak{F}_{t-1}) &= (1 - p_t)\sigma_t E(\epsilon_t) + p_t \sigma_t E(D_t) \\ &= \sigma_t[(1 - p_t)E\epsilon_t + p_t E D_t] \end{aligned} \tag{2.2}$$

2.

$$EX_t = (E(\sigma_t) - E(\sigma_t p_t))E\epsilon_t + E(p_t \sigma_t) E D_t$$

Proof. 1.

$$E(X_t|\mathfrak{F}_{t-1}) = (1-p_t)\sigma_t E(\epsilon_t|\mathfrak{F}_{t-1}) + p_t\sigma_t E(D_t|\mathfrak{F}_{t-1})$$

If ϵ_t and D_t are iid this is:

$$(1-p_t)\sigma_t E\epsilon_t + p_t\sigma_t E(D_t)$$

2.

$$\begin{aligned} EX_t = E(E(X_t|\mathfrak{F}_{t-1})) &= E((1-p_t)\sigma_t E\epsilon_t + p_t\sigma_t ED_t) \\ &= E(\sigma_t - \sigma_t p_t)E\epsilon_t + E(p_t\sigma_t)ED_t \\ &= (E(\sigma_t) - E(\sigma_t p_t))E\epsilon_t + E(p_t\sigma_t)ED_t \end{aligned}$$

□

For sake of simplicity in this first introduction we assume in some cases ϵ_t to be iid $N(0,1)$ distributed. If we follow the assumption that ϵ_t has mean zero the conditional mean is:

$$E(X_t|\mathfrak{F}_{t-1}) = (1-p_t)0 + p_t\sigma_t E(D_t|\mathfrak{F}_{t-1}) = p_t\sigma_t ED_t$$

So the following equation holds.

$$EX_t = E(E(X_t|\mathfrak{F}_{t-1})) = E(p_t\sigma_t E(D_t|\mathfrak{F}_{t-1})) = E(p_t\sigma_t)ED_t$$

This illustrates a first problem with this model: Due to the fact that we modeled p_t and σ_t in a dependent way, this formula doesn't factorize. And even if it would do so, we have to calculate Ep_t and $E\sigma_t$, at least the first problem being not analytically tractable. And we cannot see from this formula, if the expected value is constant over time.

If we ask in the general context, when the expected value is zero, we can come up with the following calculation. If both, $E\epsilon_t$ and ED_t, are zero then the mean of X_t also will be. So we assume that at least one of these means is nonzero.

$$EX_t = E(\sigma_t(1-p_t)E\epsilon_t + \sigma_t p_t ED_t) = E(\sigma_t)E\epsilon_t - E(\sigma_t p_t)E\epsilon_t + E(\sigma_t p_t)ED_t \stackrel{!}{=} 0$$

Clearly when ϵ_t has zero mean this can only be satisfied if $E(D_t) = 0$ or $E(\sigma_t p_t) = 0$. Where the latter is not possible except in degenerate cases as $\sigma_t > 0$ a.s. and $p_t > 0$ with positive probability. So let us assume that $E\epsilon_t \neq 0$.

$$\Rightarrow \frac{ED_t}{E\epsilon_t} = \frac{E(\sigma_t p_t) - E(\sigma_t)}{E(\sigma_t p_t)}$$

So if the condition

$$\frac{ED_t}{E\epsilon_t} = 1 - \frac{E(\sigma_t)}{E(\sigma_t p_t)} \tag{2.3}$$

holds, $E(X_t)$ is zero. So if we want to choose the D_t and ϵ_t to be iid, the left side is constant. If the denominator of the right side factorizes, this boils down to Ep_t being the right constant namely

$$E(p_t) = \frac{-1}{\frac{E(D_t)}{E(\epsilon_t)} - 1}.$$

In general we won't be able to check this condition analytically. We can only check it by e.g. Monte Carlo methods. Anyhow, we would not expect factorization of $E(\sigma_t p_t)$ in general, as p_t is assumed to be a function of σ_t.

2.2.2 Variance

Propositition 2.2.2.

$$E(X_t^2|\mathfrak{F}_{t-1}) = \sigma_t^2[(1-p_t)E\epsilon_t^2 + p_t ED_t^2]$$

Proof.

$$\begin{aligned} E(X_t^2|\mathfrak{F}_{t-1}) &= (1-p_t)\sigma_t^2 E(\epsilon_t^2|\mathfrak{F}_{t-1}) + p_t\sigma_t^2 E(D_t^2|\mathfrak{F}_{t-1}) \\ &= (1-p_t)\sigma_t^2 E\epsilon_t^2 + p_t\sigma_t^2 ED_t^2 \\ &= \sigma_t^2((1-p_t)E\epsilon_t^2 + p_t ED_t^2) \end{aligned}$$

□

Propositition 2.2.3. *Assume that ϵ_t has zero mean and unit variance then the following holds:*

1. $$Var(X_t|\mathfrak{F}_{t-1})) = \sigma_t^2((1-p_t) + p_t(Var(D_t) + (1-p_t)ED_t^2))$$

2. $$Var(X_t) = E\sigma_t^2 + E(p_t\sigma_t^2)(ED_t^2 - 1) - (E(p_t\sigma_t))^2(ED_t)^2$$

Proof. Now we consider the assumption that ϵ_t has zero mean and variance one, so the following holds:

$$E(X_t^2|\mathfrak{F}_{t-1}) = (1-p_t)\sigma_t^2 + p_t\sigma_t^2 ED_t^2$$

and

$$E(X_t|\mathfrak{F}_{t-1}) = p_t\sigma_t ED_t$$

To proof 1. we calculate:

$$\begin{aligned} Var(X_t|\mathfrak{F}_{t-1})) &= E(X_t^2|\mathfrak{F}_{t-1}) - (E(X_t|\mathfrak{F}_{t-1}))^2 \\ &= (1-p_t)\sigma_t^2 + p_t\sigma_t^2 ED_t^2 - p_t^2\sigma_t^2(ED_t)^2 \\ &= \sigma_t^2((1-p_t) + p_t ED_t^2 - p_t^2(ED_t)^2) \\ &= \sigma_t^2((1-p_t) + p_t(Var(D_t) + (1-p_t)ED_t^2)) \end{aligned}$$

For point 2. we realize:

$$
\begin{aligned}
Var(X_t) &= EX_t^2 - (EX_t)^2 = E(E(X_t^2|\mathfrak{F}_{t-1})) - (E(E(X_t||\mathfrak{F}_{t-1})))^2 \\
&= E(\sigma_t^2) - E(p_t\sigma_t^2) + E(p_t\sigma_t^2)ED_t^2 - (E(p_t\sigma_t)ED_t)^2 \\
&= E\sigma_t^2 + E(p_t\sigma_t^2)(ED_t^2 - 1) - (E(p_t\sigma_t))^2(ED_t)^2
\end{aligned}
$$

□

In the more general case the formula for the conditional variance gets more complicated:

Propositition 2.2.4.

$$
\begin{aligned}
&Var(X_t|\mathfrak{F}_{t-1}) \\
&= \sigma_t^2[(1-p_t)E\epsilon_t^2 + p_tED_t^2 \\
&\quad -(1-p_t)^2(E\epsilon_t)^2 - (1-p_t)p_tE\epsilon_tED_t - p_t^2(ED_t)^2]
\end{aligned}
$$

Proof. Just using Propositions 2.2.1 and 2.2.2 we get this result simply calculating. □

We can write this result alternatively:

Propositition 2.2.5.

$$
\begin{aligned}
&Var(X_t|\mathfrak{F}_{t-1}) \\
&= \sigma_t^2[(E\epsilon_t^2 + p_t(ED_t^2 - E\epsilon_t^2) \\
&\quad -(E\epsilon_t + p_t(ED_t - E\epsilon_t))^2]
\end{aligned}
$$

Proof. We just apply

$$(1-p_t)x + p_ty = x + p_t(y-x)$$

for $x = E\epsilon_t$ and $y = ED_t$ in the statement of Proposition 2.2.1 and to $x = E\epsilon_t^2$ and $y = ED_t^2$ in the statement of Proposition 2.2.2. Then we use

$$Var(X_t|\mathfrak{F}_{t-1}) = E(X_t^2|\mathfrak{F}_{t-1}) - E(X_t|\mathfrak{F}_{t-1})^2$$

□

Remark 2.2.6. We consider the definition

$$\eta_t = (1-B_t)\epsilon_t + B_tD_t$$

Then in the sense of Chapter 1 X_t is a generalized GARCH with innovations η_t provided B_t is stationary. In fact consider that

$$E(\eta_t|\mathfrak{F}_{t-1}) = (1-p_t)E\epsilon_t + p_tED_t$$

and

$$E(\eta_t^2|\mathfrak{F}_{t-1}) = (1-p_t)E\epsilon_t^2 + p_t E(D_t^2)$$

hold. So

$$Var(\eta_t|\mathfrak{F}_{t-1}) = [(E\epsilon_t^2 + p_t(ED_t^2 - E\epsilon_t^2) - (E\epsilon_t + p_t(ED_t - E\epsilon_t))^2]$$

holds. So

$$E(X_t|\mathfrak{F}_{t-1}) = \sigma_t E(\eta_t|\mathfrak{F}_{t-1})$$

holds, which shows that we are formally in a GARCH–M setting. Further we see that

$$Var(X_t|\mathfrak{F}_{t-1}) = \sigma_t^2 Var(\eta_t|\mathfrak{F}_{t-1})$$

holds. So the variance of the process X_t is dependent on the GARCH dynamic we impose and on the conditional variance of the crash–non-crash mixture innovations.

Closely related to the question of the variance is, when does $E\sigma_t^2$ exist and is constant over time. We take a first look at this problem under the assumption that the ϵ_t are iid $N(0,1)$ distributed and that

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta\sigma_{t-1}^2$$

$$\begin{aligned}
\Rightarrow E(\sigma_t^2) &= \omega + \alpha E(X_{t-1}^2) + \beta E(\sigma_{t-1}^2) \\
&= \omega + \alpha E(E(X_{t-1}^2|\mathfrak{F}_{t-2})) + \beta E(\sigma_{t-1}^2) \\
&= \omega + \alpha E(\sigma_{t-1}^2((1-p_{t-1}) + p_{t-1}E(D_{t-1}^2))) + \beta E(\sigma_{t-1}^2)
\end{aligned}$$

Assuming p_t and σ_t^2 are uncorrelated this boils down to:

$$\omega + \alpha E(\sigma_{t-1}^2)E((1-p_{t-1}) + p_t E(D_{t-1}^2)) + \beta E(\sigma_{t-1}^2)$$

So if $E(\sigma_t^2) = E(\sigma_{t-1}^2)$:

$$E(\sigma_t^2) = \frac{\omega}{1 - \alpha(E(1-p_{t-1} + p_{t-1}ED_{t-1}^2)) - \beta}$$

So for $E(\sigma_t^2)$ to exist $\alpha(E(1-p_{t-1} + p_{t-1}ED_{t-1}^2)) + \beta$ has to be smaller than 1. To be constant over time moreover $(E(1-p_{t-1} + p_{t-1}ED_{t-1}^2))$ has to be also constant over time. Switching back to the model with arbitrary distribution of the ϵ_t we get by analogous reasoning the following proposition:

Propositition 2.2.7. *Let σ_t^2 and p_t be uncorrelated. If $\alpha(\max(E\epsilon_t^2, ED_t^2)) + \beta < 1$ and $(1-Ep_t)E\epsilon_t^2 + Ep_t ED_t^2$ is constant, $E(\sigma_t^2)$ exists and is constant over time.*

The condition in this corollary is rather strong, and might lead to unnecessary strong constraints. That is one of the reasons why we consider different approaches in later chapters.
At least we can get a relation between the parameters and EX_t^2, $E\sigma_t^2$ if we assume that these moments are constant over time and some other reasonable assumptions: If we assume $EX_t^2 = EX_{t-1}^2$ and $E\sigma_t^2 = E\sigma_{t-1}^2$ then we get via the GARCH equation:

$$E\sigma_t^2 = \omega + \alpha EX_{t-1}^2 + \beta E\sigma_{t-1}^2$$

$$\Rightarrow EX_t^2 = \frac{(1-\beta)E\sigma_{t-1}^2 - \omega}{\alpha} = (*)$$

From Proposition 2.2.2 we get:

$$EX_t^2 = E\epsilon_t^2 E\sigma_t^2 + E(p_t\sigma_t^2)(ED_t^2 - E\sigma_t^2)$$

If we combine that with $(*)$ we get:

$$(1 - \beta - \alpha E\epsilon_t^2)E\sigma_t^2 = \omega + \alpha E(p_t\sigma_t^2)(ED_t^2 - E\sigma_t^2)$$

If we assume now $E\epsilon_t^2 = 1$, which is just scaling, and $ED_t^2 \geq E\epsilon_t^2$ which makes sense if we want to model crashes we get:

$$(1 - \alpha - \beta))E\sigma_t^2 \geq \omega \tag{2.4}$$

and

$$EX_t^2 = \frac{(1-\alpha-\beta)E\sigma_{t-1}^2 - \omega + \alpha E\sigma_{t-1}^2}{\alpha} \geq E\sigma_{t-1}^2 \tag{2.5}$$

2.2.3 Covariances

Even if we assume the ϵ_t to be iid $N(0,1)$ distributed and the to D_t be iid, both sequences independent of each other, we have problems to calculate the autocovariances.

$$\begin{aligned}
&Cov(X_t, X_{t+k}) = E(X_tX_{t+k}) - E(X_t)E(X_{t+k}) \\
&\quad= E(X_tX_{t+k}) - E(p_t\sigma_t)E(p_{t+k}\sigma_{t+k})E(D_t)E(D_{t+k}) \\
&\quad= E(E(E(X_tX_{t+k}|\mathfrak{F}_{t+k-1})|\mathfrak{F}_{t-1})) - E(p_t\sigma_t)E(p_{t+k}\sigma_{t+k})E(D_t)E(D_{t+k}) \\
&\quad= E(E(0 + 0 + (1-B_t)\sigma_t\epsilon_t p_{t+k}\sigma_{t+k}E(D_{t+k}) + B_t\sigma_tD_tp_{t+k}\sigma_{t+k}E(D_{t+k})|\mathfrak{F}_{t-1})) \\
&\qquad -E(p_t\sigma_t)E(p_{t+k}\sigma_{t+k})E(D_t)E(D_{t+k}) \\
&\quad= 0 + E(E(B_tp_{t+k}\sigma_{t+k}|\mathfrak{F}_{t-1})\sigma_t)E(D_t)E(D_{t+k}) - E(p_t\sigma_t)E(p_{t+k}\sigma_{t+k})E(D_t)E(D_{t+k}) \\
&\quad= E(D_t)E(D_{t+k})(E(B_tp_{t+k}\sigma_{t+k}\sigma_t) - E(p_t\sigma_t)E(p_{t+k}\sigma_{t+k}))
\end{aligned}$$

If $E(D_t) \neq 0$, then this expression doesn't simplify without further assumptions, because the correlation between the p_ts is unknown as well as the correlation between them and the sigmas.

2.3 The choice of the crash distribution and the problem of defining a crash

By the choice of the crash distribution D_t we implicitly define the meaning of crash in our model. So we have to know what we mean by a crash prior to choose such a distribution. If we mean by crash that we get large losses, we should use a distribution which lives only on the negative numbers. If we also take mean reversion effects being part of the crash regime, we also can allow for positive values of the crash distribution, but with overall negative skewness. Even having chosen a parametric family of distribution, it is the question, which distribution we choose. Adding the parameters concerning the distribution to be estimated will bring a certain muddiness into the model: Properties like mean and variance depend on the choice of these parameters and of the parameters ω, α, β. Moreover the estimates of the parameters concerning p_t clearly rely on the choice what we mean by a crash. So there is little wonder that trying to do this ended in complete failure.
As we have seen there are difficulties in the CGARCH-S model to calculate the unconditional mean of the process. If we take, like we have done in the examples, the non-crash distribution to have zero mean and the crash distribution to be purely negative, it is however clear, that that means that the mean is overall negative. If we interpret the process arising as asset returns, this means eventual bankruptcy. So to have a more realistic model, or at least a model not implying bankruptcy per se, we would have to choose the non-crash distribution to have a positive mean. But to do this in a not purely erratic way, we don't have enough information, not knowing what unconditional mean a choice of pairs of distributions leads to.

2.4 Higher Moments

For sake of simplicity we assume here $E(\epsilon_t|\mathfrak{F}_{t-1}) = 0$. Further we assume ϵ_t and D_t to be iid independent from each other.

2.4.1 Skewness

If we choose for ϵ_t or D_t distributions, which are non symmetric, X_t can be skewed. This isn't the case in the standard GARCH model. This even holds if

$\mathcal{L}(\epsilon_t|\mathfrak{F}_{t-1})$ is symmetric around zero, as can be seen by the formula:

$$
\begin{aligned}
E(X_t - EX_t)^3 &= E(E[(X_t - EX_t)^3|\mathfrak{F}_{t-1}]) \\
&= E((1-p_t)\sigma_t^3 E\epsilon_t^3 + p_t\sigma_t^3 ED_t^3) \\
&\quad -3E[(p_t\sigma_t ED_t)((1-p_t)\sigma_t^2 E\epsilon_t^2 + p_t\sigma_t^2 ED_t^2)] \\
&\quad +3E[(p_t\sigma_t ED_t)^2((1-p_t)\sigma_t E\epsilon_t + p_t\sigma_t ED_t)] \\
&\quad +E(p_t\sigma_t ED_t)^3
\end{aligned}
$$

2.4.2 Kurtosis

We give here a formula for the centered fourth moment:

$$
\begin{aligned}
E(X_t - EX_t)^4 &= E(E[(X_t - EX_t)^4|\mathfrak{F}_{t-1}]) \\
&= E((1-p_t)\sigma_t^4 E\epsilon_t^4 + p_t\sigma_t^4 ED_t^4) \\
&\quad -4E(p_t\sigma_t ED_t)((1-p_t)\sigma_t^3 E\epsilon_t^3 + p_t\sigma_t^3 ED_t^3) \\
&\quad +2E[(p_t\sigma_t ED_t)^2((1-p_t)\sigma_t^2 E\epsilon_t^2 + p_t\sigma_t^2 ED_t^2)] \\
&\quad -4E[(p_t\sigma_t ED_t)^3((1-p_t)\sigma_t E\epsilon_t + p_t\sigma_t ED_t)] \\
&\quad +E(p_t\sigma_t ED_t)^4
\end{aligned}
$$

Pitifully this formula doesn't simplify. But we can see that the higher order moment structure is quite complex even if $\mathcal{L}(\epsilon_t|\mathfrak{F}_{t-1})$ is symmetric around 0.

2.5 An example

We now give an example of a process following the dynamic described in the model above. We chose a standard normal distribution for ϵ_t and a lognormal distribution with parameters (0,1) as a crash-distribution. We modeled The parameters are $\omega = 10^{-5}$, $\alpha = 0.01218$, $\beta = 0.9$, further $a = -10$, $b = -490$ and $c = 100$. That means we modeled

$$p_t = \frac{1}{1 - \exp(-(-10 - 490X_{t-1} + 100\sigma_t^2))}$$

In 2.1 the process itself is shown. In Figures 2.3 and 2.4 we see that the volatility, i.e. the conditional standard deviation, of a process following the model can be extremely different from σ_t namely when as in this example p_t has a big range. Further the irregular nature of p_t is shown in 2.2.

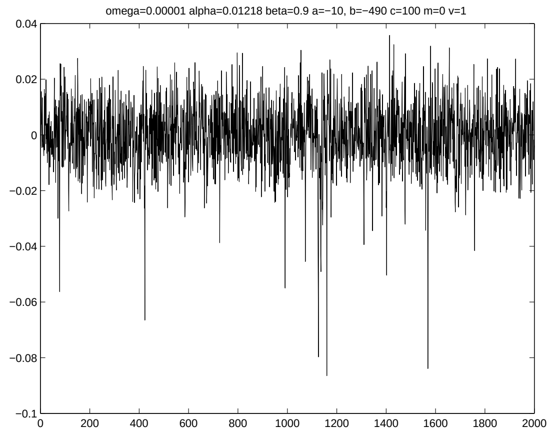


Figure 2.1: An example of a process following the model

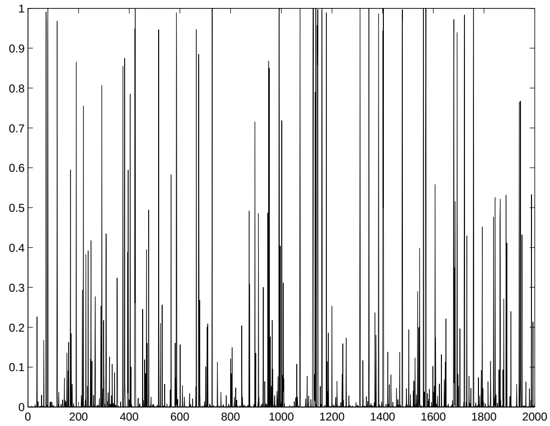


Figure 2.2: The crash-probability of this process

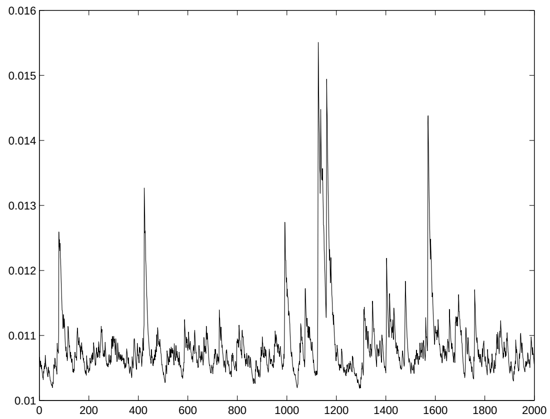


Figure 2.3: The sigma of this process

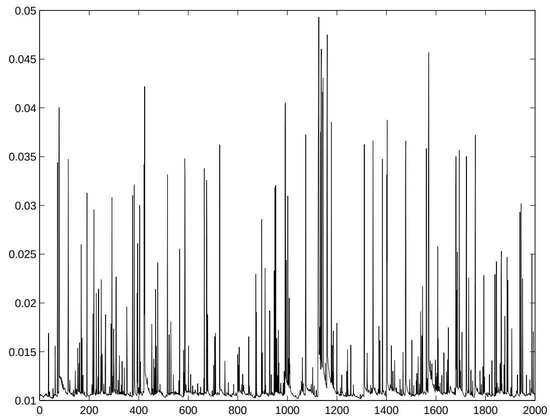


Figure 2.4: The volatility of the process

In Figure 2.5 we show the shape of the price process if 2.1 were returns of an asset. The stars actually show, when a (pseudo–)crash occurred in the simulation, i.e. $B_t = 1$. In my opinion an intuitive idea of what a crash is isn't mirrored by the model's immanent (pseudo–)crashes. Only the cluster of such events near time instant 1100 displays a price path one might expect from a crash.

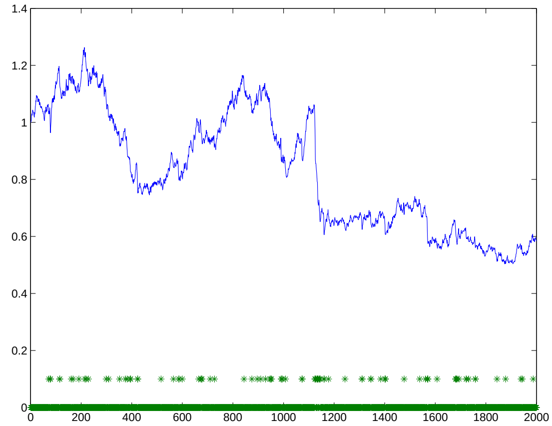


Figure 2.5: Pseudo-price and times of crashes

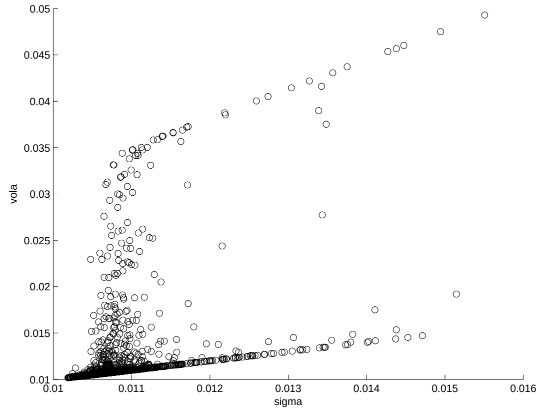


Figure 2.6: The volatility depending on sigma

Figure 2.6 shows the dependence of the volatility on σ_t and implicitly on p_t. In times when p_t is low and not very noisy the volatility depends almost linear on σ_t. In an intermediate phase the crash-probability is dominant on the volatility, shifting it up to a phase where another almost linear dependence on σ_t holds when p_t is near 1.

2.6 Discussion

The overall problem, whether discussing moments, stationarity, mixing or ergodicity question, is so to speak to get a fixed point to start from. Beside questions of correlation between the sigmas and the ps, we get stuck in vicious circles like: "The mean of X_t is constant, if that of p_t is, which is the case, when the mean of X_t is constant and some other assumptions are satisfied." Therefore we modify our first attempt on a crash model in the following chapters.

2.7 Transformations of the signum–function as Crash–probabilities

One idea to establish stability properties in the model is to use a function of X_t,which as such can be observed, for generating the crash–probabilities, but let the function of such nature that it is also a function of $\eta_t = (1 - B_t)\epsilon_t + B_t D_t$. This avoids vicious circles in the argumentation. A generalization of this idea, p_t being just a function of η_t we will use later in a more general context. One function with this property in our general model setup is the signum function.

We stick to the general assumption that $\{\epsilon_t\}$ are iid and the $\{D_t\}$ are iid both independent from each other.

We assume that the distributions of ϵ_t and the D_t both are absolutely continuous with respect to the Lebesgue measure. Because then p_t will only attain two values with positive probability we will call this model a CGARCH-SB model, where B stands for binary. In this setting we can do the following calculation:
Let $p_t = f(\text{sign}(X_t))$, where f is a transformation into the open interval $(0,1)$. Let $f(-1) = a$, $f(1) = b$ hold. Define $p_c := P(D_t \geq 0)$ and $p_{nc} := P(\epsilon_t \geq 0)$
Then

$$\begin{aligned} E(p_t|\mathfrak{F}_{t-1}) &= (p_{t-1}p_c + (1-p_{t-1})p_{nc})b + (p_{t-1}(1-p_c) + (1-p_{t-1})(1-p_{nc}))a \\ &= (p_{t-1}(p_c b + (1-p_c)a) + (1-p_{t-1})(p_{nc}b + (1-p_{nc}))a) \end{aligned}$$

We set $c = (p_c b + (1-p_c)a)$ and $d = (p_{nc}b + (1-p_{nc})a)$.

$$E(p_t|\mathfrak{F}_{t-1}) = p_{t-1}c + (1-p_{t-1})d = d + p_{t-1}(c-d)$$

Lemma 2.7.1. $E(p_t|\mathfrak{F}_{t-n}) = d(\sum_{i=1}^{n}(c-d)^{i-1}) + p_{t-n}(c-d)^n$

Proof. The previous discussion proves the fact for $n = 1$.

$n \to n+1$

$$\begin{aligned} E(p_t|\mathfrak{F}_{t-n+1}) &= E(E(p_t|\mathfrak{F}_{t-n})|\mathfrak{F}_{t-(n+1)}) \\ &= E(d(\sum_{i=1}^{n}(c-d)^{i-1}) + p_{t-n}(c-d)^n|\mathfrak{F}_{t-(n+1)}) \\ &= d(\sum_{i=1}^{n}(c-d)^{i-1}) + E(p_{t-n}|\mathfrak{F}_{t-(n+1)})(c-d)^n \\ &= d(\sum_{i=1}^{n}(c-d)^{i-1}) + (p_{t-(n+1)}c + (1-p_{t-(n+1)})d)(c-d)^n \\ &= d(\sum_{i=1}^{n}(c-d)^{i-1}) + (d + p_{t-(n+1)}(c-d))(c-d)^n \\ &= d(\sum_{i=1}^{n}(c-d)^{i-1}) + d(c-d)^n + p_{t-(n+1)}(c-d)^{n+1} \\ &= d(\sum_{i=1}^{n+1}(c-d)^{i-1}) + p_{t-(n+1)}(c-d)^{n+1} \end{aligned}$$

□

Lemma 2.7.2. *If we have a process which has an infinite past we have $E(p_t) = d\frac{1}{1-c+d}$ independent from t.*

Proof. Because a and b are in $(0,1)$ and the p_c and $(1-p_c)$ add up to one $0 < min(a,b) \leq c \leq \max(a,b) < 1$ The same holds for d. So

$$\begin{aligned} E(p_t) &= \lim_{n\to\infty} E(p_t|\mathfrak{F}_{t-n}) \\ &= \lim_{n\to\infty} d(\sum_{i=1}^{n}(c-d)^{i-1} + p_{t-n}(c-d)^n \\ &= \lim_{n\to\infty} d(\sum_{i=1}^{n}(c-d)^{i-1}) + \lim_{n\to\infty} p_{t-n}(c-d)^n \end{aligned}$$

Now p_{t-n} is bound to be between 0 and 1 and so $\lim_{n\to\infty} p_{t-n}(c-d)^n = 0$. On the other hand $\lim_{n\to\infty} d(\sum_{i=1}^{n}(c-d)^{i-1}) = d\frac{1}{1-c+d}$. ☐

A completely analogous discussion yields :

$$E(p_t^2) = d'^2\frac{1}{1-c'+d'}$$

Where $c' := (p_c b^2 + (1-p_c)a^2)$ and $d' := (p_{nc}b^2 + (1-p_{nc}))a^2)$ And so we can calculate the variance:

$$Var(p_t) = d'^2\frac{1}{1-c'+d'} - d^2\frac{1}{(1-c+d)^2}$$

which is obviously a constant.

$$\begin{aligned} E(p_t p_{t+\tau}) &= E(E(p_t p_{t+\tau}|\mathfrak{F}_t)) \\ &= E(p_t d(\sum_{i=1}^{\tau}(c-d)^{i-1}) + p_t^2(a-b)^\tau) \\ &= E(p_t d(\sum_{i=1}^{\tau}(c-d)^{i-1}) + E(p_t^2(a-b)^\tau) \\ &= d(\sum_{i=1}^{\tau}(c-d))^{i-1})E(p_t) + (a-b)^\tau E(p_t^2) \\ &= d\frac{(c-d)^\tau}{1-c+d} + \frac{(a-b)^\tau}{1-c^2+d^2} \end{aligned}$$

This clearly just depends on τ only.
This discussion yields the following Theorem:

Theorem 2.7.3. *If p_t is a transformation of the signum function into the open interval $(0,1)$, then p_t is second order stationary.*

Corollary 2.7.4. *Let ϵ_t and D_t be both square integrable. If $\alpha d' \frac{1}{1-c'+b'} E(D_t)^2 + \beta < 1$, then X_t is second order stationary.*

Proof. If $\alpha d' \frac{1}{1-c'+b'} + \beta < 1$ then $E(\sigma_t^2)$ exists and so does $E(\sigma_t)$.

$$E(X_t) = E(\sigma_t) \left(\left(1 - d \frac{1}{1-c+b} \right) E(\epsilon_t) + d \frac{1}{1-c+b} E(D_t) \right)$$

holds not depending on t and

$$E(X_t^2) = E(\sigma_t^2) \left(\left(1 - d' \frac{1}{1-c'+d'} \right) E(\epsilon_t^2) + d \frac{1}{1-c'+d'} E(D_t^2) \right) < \infty$$

The covariances are zero because of the independence of all components of the "residuals". □

2.8 Models with crash probabilities depending on external variables

p_t may partially or completely depend on external variables additionally to or instead of X_t and σ_t. These variables might be multivariate. They may be derived from an index or other outside data influencing the asset. If the crash probabilitiy is only a function of exogeneous variables then the discussion becomes much simpler. We consider this case in this section.

2.8.1 Mean and Variance

We return to model (2.1) and assume that $p_t = f(W_t)$ with external variables W_t chosen such that p_t is stationary with mean p and not correlated with σ_t. We will call this model CGARCH-SP model, where P symbolizes that p_t is a stochastic process in its own right. We get the following expected value.

$$E(X_t) = E[\sigma_t(E((1-p_t)E\epsilon_t) + (E(p_t ED_t)))] = E(\sigma_t)((1-p)E\epsilon_t + pED_t)$$

A nice calculation concerning the variance can also be done:

$$\begin{aligned}
Var(X_t) &= E(\sigma_t^2((1-p_t)E\epsilon_t^2) + p_t ED_t^2) - (EX_t)^2 \\
&= E(\sigma_t^2)((1-p)E\epsilon_t^2) + pED_t^2) \\
&\quad -E(\sigma_t)^2(p^2(ED_t)^2 - (1-p)^2(E\epsilon_t)^2 - p(1-p)E\epsilon_t ED_t) \\
&= \sigma^2(1-p)E\epsilon_t^2 + pED_t^2 - (p^2((ED_t)^2 - (1-p)^2(E\epsilon_t)^2 - p(1-p)E\epsilon_t ED_t) \\
&= p(Var(D_t) + (1-p)(ED_t)^2 + (1-p)Var(\epsilon_t) + p(E\epsilon_t)^2 - Var(B_t)E\epsilon_t ED_t)
\end{aligned}$$

If we have in the general model that $E(p_t) = p$ is constant the formulae above also hold. Surprisingly $Var(X_t)$ doesn't depend on $Var(p_t)$.This is evidence how restrictive the assumption of σ_t and p_t being not correlated is.

2.8.2 Stationarity

Assumption 2.8.1. *Let W_t be a possibly vector valued time series of exogenous variables. We assume that W_t is strictly stationary and ergodic. Let $p_t = f(W_t)$ where f is a measurable function taking values between 0 and 1.*

Because under stationarity α–mixing implies ergodicity this assumption is implied by the following alternative assumption.

Assumption 2.8.2. *Let W_t be a possibly vector valued time series of exogenous variables. We assume that W_t is strictly stationary and α–mixing. Let $p_t = f(W_t)$ where f is a measurable function taking values between 0 and 1.*

Lemma 2.8.3.

1. *Under Assumption 2.8.1 $\eta_t = (1 - B_t)\epsilon_t + B_t D_t$ is strictly stationary and ergodic.*
2. *Under Assumption 2.8.2 $\eta_t = (1 - B_t)\epsilon_t + B_t D_t$ is strictly stationary and α–mixing. If moreover ϵ_t and D_t are of finite mean and variance η_t is also second order stationary.*

Proof. If W_t is strictly stationary and mixing, so is p_t([PP97]). Because p_t is between 0 and 1, all its moments must exist. B_t is just an $\mathfrak{B}(1, p_t)$ distributed random variable. It's only dependence on time is via p_t so time related concepts like mixing and stationarity carry over from p_t. So if the ϵ_t are iid and the D_t are iid, each sequence independent from the other, dependency in η_t can only derive from p_t, which shows mixing. The strong stationarity of B_t and the distributions drawn from under its rule yields the strong stationarity of η_t. For the existence of the moments we remark that we have seen that $E(p_t)$ exists, so we have: $E(\eta_t) = (1 - E(p_t))E\epsilon_t + E(p_t)ED_t < \infty$ and $E(\eta_t^2) = (1 - E(p_t))E\epsilon_t^2 + E(p_t)ED_t^2 < \infty$ The mixed terms are zero because either B_t or $1 - B_t$ are zero. □

In fact the Assumptions 2.8.1 and 2.8.2 are special cases of the following assumptions:

Assumption 2.8.4.

We assume that p_t is strictly stationary and ergodic.

Assumption 2.8.5.

We assume that p_t is strictly stationary and α–mixing.

Lemma 2.8.6. *Suppose 2.8.4 holds. Suppose a GARCH(1,1) regime on $\overline{\eta}_t$ the centered and normalized version of η_t,that means we consider:*

$$\overline{\eta}_t = \frac{\eta_t - E(\eta_t|\mathfrak{F}_{t-1})}{\sqrt{Var(\eta_t|\mathfrak{F}_{t-1})}}$$

and

$$\sigma_t^2 = \omega + \alpha\overline{\eta}_t^2 + \beta\sigma_{t-1}^2.$$

This process is second order stationary iff $\alpha + \beta < 1$

Proof. The expected value of this process is clearly constant zero.
For the variance: In our situation the representation theorem for GARCH processes holds, with $\zeta_t = \sigma_t^2(\overline{\eta}_t^2 - 1)$ having zero mean, which is sufficent to use Property 3.19 in [Gou97]. So X_t doesn't have to have a finite fourth moment for this discussion. So the argument of [Gou97] goes through.
For the covariances: Because the ϵ_t and D_t are iid $E(\epsilon_t\epsilon_{t+\tau}) = E(D_t\epsilon_{t+\tau}) = E(\epsilon_t D_{t+\tau}) = E(D_t D_{t+\tau}) = 0$ holds. Consequently the covariances of η_t are zero. But $\overline{\eta}_t$ is just a centered and normalized version of η_t and hence is also uncorrelated. So $Cov(\overline{\eta}_t, \overline{\eta}_{t+\tau}) = 0$ for $\tau \neq 0$ constantly neither depending on t or on τ. □

Corollary 2.8.7. *Suppose 2.8.4 holds. If we get the double infinite process $X_t = \sigma_t\eta_t$ by imposing a GARCH(1,1) regime on σ_t, this process is second order stationary iff $\alpha((1-p)E\epsilon_t^2 + pED_t^2) + \beta < 1$ with $p = Ep_t$. In this case*

$$\sigma^2 := E(\sigma_t^2) = \frac{\omega}{1 - \alpha((1-p)E\epsilon_t^2 + pED_t^2) - \beta}$$

with $p = Ep_t$.

Proof. The discussion of the covariances in the preceeding proof holds here, too.
"$\Rightarrow$"
Suppose the variance of X_t is constant. By assumption $p = E(p_t)$, and the first two moments of D_t and ϵ_t are not depending on t either. Therefore $\sigma^2 := E(\sigma_t^2)$ being the only other term in the formula for the variance of X_t has to be also constant. If σ^2 exists and is constant we have

$$\begin{aligned}\sigma^2 = \omega + \alpha E(X_t^2) + \beta\sigma^2 = \omega + \alpha\sigma^2((1-p)E\epsilon_t^2 + pED_t^2) + \beta\sigma^2 \\ \Rightarrow \sigma^2(1 - \alpha((1-p)E\epsilon_t + pED_t) - \beta) = \omega \\ \Rightarrow \sigma^2 := E(\sigma_t^2) = \frac{\omega}{1 - \alpha((1-p)E\epsilon_t^2 + pED_t^2) - \beta}\end{aligned}$$

"$\Leftarrow$"
For sake of notation we define $c := (1-p)E\epsilon_t^2 + pED_t^2$. Due to the GARCH regime $E(\sigma_t^2) = \omega + \alpha c E(\sigma_{t-1}^2) + \beta E(\sigma_{t-1}^2)$ holds. If the roots of the characteristic polynomial $1 - (\alpha c + \beta)L$, are strictly outside the unit circle, the sequence EX_t^2 converges. But then X_t is asymptotically second order stationary, because $p = E(p_t)$ and the first two moments of D_t and ϵ_t are constant. The proof that roots of $1 - (\alpha c + \beta)L$ are strictly outside the unit circle iff $(\alpha c + \beta) < 1$ are completely analogous to [Gou97] page 37, we will give it in the more general form in the next proof. □

Corollary 2.8.8. *Suppose 2.8.4 holds. Suppose we get the double infinite process $X_t = \sigma_t\eta_t$ by imposing a GARCH(p,q) regime on σ_t. If this process is second order stationary then $((1-p)E\epsilon_t^2 + pED_t^2)\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$ In this case*

$$\sigma^2 := E(\sigma_t^2) = \frac{\omega}{1 - ((1-p)E\epsilon_t^2 + pED_t^2)\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j}$$

Proof. The same argumentation as above using the general GARCH(p,q) equations: If the roots of the characteristic polynomial $1 - \sum_{i=1}^{max(p,q)}(c\alpha_i + \beta_i)L^i$ are strictly outside the unit circle, the process σ_t^2 is asymptotically second order stationary. This is the case iff $c\alpha(1) + \beta(1) < 1$. If $c\alpha(1) + \beta(1)$ is bigger or equal than 1 $c\alpha(1)+\beta(1) \leq 0$ and $c\alpha(0)+\beta(0) > 0$ holds and the characteristic polynomial would have a real root inside the unit circle. If we assume on the other hand $c\alpha(1) + \beta(1) < 1$ and the existence of z, a root of the characteristic polynomial with modulus strictly smaller than 1 the following contradiction arises:

$$\begin{aligned} 1 = & \sum_{i=1}^{max(p,q)} (c\alpha_i + \beta_i)z^i = |\sum_{i=1}^{max(p,q)} (c\alpha_i + \beta_i)z^i| \\ \leq & \sum_{i=1}^{max(p,q)} (c\alpha_i + \beta_i)|z^i| \leq c\alpha(1) + \beta(1) < 1 \end{aligned}$$

□

Because we are also interested in strong stationarity we want to transfer some methods used in [Nel90] to our problem. The crucial point here is to use an adequate Law of Large Numbers, namely a strong one.

Theorem 2.8.9. *Let $\omega \neq 0$ and let Assumption 2.8.4 hold. Let $E|\log(\beta + \alpha\eta_t^2)| < \infty$. Then the double infinite sequence σ_t is strictly stationary iff $E(\log(\beta + \alpha\eta_t^2)) < 0$. The same holds for X_t.*

Proof. Assumption 2.8.4 implies that η_t and so $\log(\beta + \alpha\eta_t^2)$ are stationary and ergodic. The condition $E|\log(\beta + \alpha\eta_t^2)| < \infty$ allows us to a make use of the strong law of large numbers arising from the ergodic theorem. But with a strong law of large numbers the discussion of Nelson goes through: Then we have $\frac{1}{n}\sum_{i=1}^n \log(\beta + \alpha\eta_i^2) \to E(\log(\beta + \alpha\eta_1^2))$. But that means $\forall\epsilon > 0$ $\exists M \in \mathbb{N}$ such that $\forall n > M$ $|\frac{1}{n}\sum_{i=1}^n \log(\beta+\alpha\eta_{t-i}^2) - E(\log(\beta+\alpha\eta_i)^2)| < \epsilon$. This holds because of the definition of almost sure convergence does not depend on indexing. If we take $\epsilon = \frac{|E(\log(\beta+\alpha\eta_t^2))|}{2}$ we get as in Nelson: $|\frac{1}{n}\sum_{i=1}^n \log(\beta+\alpha\eta_{t-i}^2) - E(\log(\beta+\alpha\eta_i^2))| < \frac{|E(\log(\beta+\alpha\eta_t^2))|}{2}$ for n big enough. The rest of Nelson's proof clearly goes through: From the previous equation we can derive the following equations holding almost surely:

$$\sum_{i=1}^n \log(\beta + \alpha\eta_i^2) < \frac{n}{2}E\log(\beta + \alpha\eta_i^2)$$

because of our assumption and we get:

$$\prod_{i=1}^n (\beta + \alpha\eta_i^2) < \exp\left(\frac{n}{2}E\log(\beta + \alpha\eta_i^2)\right)$$

hence we have $\prod_{i=1}^{n}(\beta+\alpha\eta_{t-i}^2) = O(e^{-\lambda n})$ with $\lambda = \frac{|E(\log(\beta+\alpha\eta_t^2))|}{2} > 0$.
Therefore, $\sum_{n=1}^{\infty}\prod_{i=1}^{n}(\beta+\alpha\eta_{t-i}^2)$ converges. The measurability argument of Nelson also holds here. $\square$

In this section we have seen that stationarity and ergodicity of η_t will carry over to X_t provided some moment type conditions hold. That is not true for mixing if X_t depends on the infinite past. The next section will discuss how far the methods above can proceed in that context and why they ultimately fail.

2.8.3 Discussion of SLLN for mixing processes

Consider a sequence X_t with $E(X_t) = E(X_{t'})$ for all t, t'. We define $\tilde{X}_t := X_t - E(X_t)$. We want to find conditions for a strong law for X_t to hold. Clearly if $n^{-1}\sum_{t=1}^{n}\tilde{X}_t \to 0$ a.s. then $n^{-1}\sum_{t=1}^{n}X_t \to E(X_t)$.
We now want to use the theorems of [McL75]. To this extend we assume

1. $\tilde{X}_t$ is strongly mixing with α_n being of size $\frac{-p}{p-2}$ for some $p \geq 2$.
2. There exits a r, $\frac{p}{2} < r \leq p < \infty$ with
$$\sum_{t=1}^{\infty} E^{\frac{2}{p}}\frac{|\tilde{X}_t|^r}{t^r} < \infty$$

According to Theorem 2.10 of [McL75] $n^{-1}\sum_{t=1}^{n}\tilde{X}_t \to 0$ a.s holds under this assumption.
If $r = 2$ holds then the second point in the assumption can be substituted by
$$\int_0^{\infty}\sup_t P(\tilde{X}_t > x)dx < \infty$$

We now look at conditions from [Han91b]. Let $1 < q < r$ $1 \leq p < r < 2p$

1. $$\sum_{m=1}^{\infty}\alpha_m^{\frac{1}{q}-\frac{1}{2p}} < \infty$$
   and
   $$\sum_{t=1}^{\infty} t^{-\frac{r}{p}}||\tilde{X}_t||_r^{\frac{r}{p}} < \infty$$
2. $$\sum_{m=1}^{\infty}\alpha_m^{1-\frac{1}{2p}} < \infty$$
   and
   $$\sup_t E(|\tilde{X}_t|^r) < \infty$$

3.
$$\sup_{i\geq 1} E\left(|\frac{\tilde{X}_t}{t^\alpha}|^r\right) < \infty$$

for an $\alpha > 0$ and either

(a) $1 \leq r \leq 2$,
$$\sum_{m=1}^{\infty} \alpha_m^{\frac{1}{2}} < \infty$$
and $\alpha < \frac{r-1}{r}$ or

(b) $r \geq 2$,
$$\sum_{m=1}^{\infty} \alpha_m^{1-\frac{1}{r}} < \infty$$
and $\alpha < \frac{1}{2} - \delta$ for some $\delta > 0$.

According to page 217 of [Han91b] in either of the cases of above assumption also $n^{-1}\sum_{t=1}^{n} \tilde{X}_t \to 0$ a.s. holds.
So in any of these cases we have a strong law for X_t. It would be nice if we could weaken the assumptions above by just assuming an mixing rate decreasing fast enough on η_t and that the mean is constant over time and come up with a mixing result for the whole process. Now if we assume that a strong law holds for $\log(\beta + \alpha\eta_t)$ nearly all arguments of the discussion above go through. In fact σ_t will be a measurable function. The stumbling stone is that a measurable function of a whole mixing process does not have to be mixing. Therefore, to gain anything at all we have to turn to approximation concepts like near epoch dependence or L_p approximability.

Chapter 3

Modifications of the model

3.1 Crash-models with GARCH-volatilities

For sake of simplicity we assume $E\epsilon_t = 0$. If we choose D_t non-symmetric or with a mean differing from zero, in the model class above σ_t won't be the conditional volatility like in the GARCH–model. So we can alternatively develop the following model.While keeping the equation

$$X_t = \begin{cases} \sigma_t \epsilon_t \text{ with probability } 1 - p_t \\ \sigma_t D_t \text{ with probability } p_t \end{cases} \tag{3.1}$$

we define the volatility

$$v_t := \sqrt{Var(X_t | \mathfrak{F}_{t-1})}$$

and we model further:

$$\begin{aligned} v_t^2 &= Var(X_t|\mathfrak{F}_{t-1}) = \omega + \alpha(X_{t-1} - E(X_{t-1}|\mathfrak{F}_{t-2}))^2 + \beta v_{t-1}^2 & (3.2) \\ &= \omega + \alpha(X_{t-1} - \sigma_{t-1} p_{t-1} E(D_{t-1}))^2 + \beta v_{t-1}^2 & (3.3) \end{aligned}$$

Then we let p_t depend on X_{t-1} and v_t.

$$p_t = f(X_{t-1}, v_t)$$

Finally we have to get the dynamic into the factor σ_t, using (3.1):

$$\sigma_t^2 = \frac{v_t^2}{(1 - p_t)E\epsilon_t^2 + p_t E D_t^2 - p_t^2 (ED_t)^2}$$

So we model the volatility rather than the multiplicator σ_t directly via the GARCH–equation. For that reason we will call models of this type "volatility models" or in concordance with the previous notation CGARCH-V models.We remark that in analogy to the CGARCH-S models we also can alter this model to a CGARCH-VP model, if we let p_t depend exclusively on external variables or

a CGARCH-VB model using a transformation of the signum function to model p_t.
Because the philosophy of variances is to deal with centered moments we put the centered version $(X_{t-1} - E(X_{t-1}|\mathfrak{F}_{t-2}))^2$ into the equation. This yields also a practical result: Assuming that the unconditional variance v^2 exists, taking expectation on both sides of the equation yields (by the definition of the conditional variance) the same result as in the ordinary GARCH(1,1) model:

$$v^2 = \frac{\omega}{1 - \alpha - \beta}$$

and $\alpha + \beta < 1$.
The converse is also true due to the same argumentation as in the GARCH case. So we have a simple condition for the variance not becoming explosive. But we traded in a rather nasty formula for the conditional mean:

$$\begin{aligned} E(X_t|\mathfrak{F}_{t-1}) &= p_t E(D_t) \sqrt{\frac{v_t^2}{(1-p_t)E\epsilon_t^2 + p_t ED_t^2 - p_t^2 (ED_t)^2}} \\ &= \frac{v_t p_t ED_t}{\sqrt{(1-p_t)E\epsilon_t^2 + p_t ED_t^2 - p_t^2 (ED_t)^2}} \end{aligned}$$

Moreover, the actual model of the process $\{X_t\}$ is due to the complicated form of σ_t too different from GARCH processes, that the methodology used to prove more complicated properties of GARCH processes is not applicable to this model.

3.2 An Example

Here we give an example of a process following the volatility model. Due to the fact that the modeling avoids explosivity in the mean by construction and that we want to model immanent crashes displaying at least a bit of the economic idea of a crash, we chose the crash-distribution to be lognormal with parameters $(1, 1)$. As the non-crash-distribution we chose $N(0, 1)$ so the formula for getting σ_t^2 out of the volatility and p_t reduces to

$$\sigma_t^2 = \frac{v_t^2}{(1-p_t) + p_t(Var(D_t) + (1-p_t)(E(D_t))^2)}$$

We chose

$$p_t = \frac{1}{1 - \exp(-(-10 - 490X_{t-1} + 100v_t^2))}$$

and $\omega = 10^{-5}$,$\alpha = 0.01218$ and $\beta = 0.9$.
We see in 3.2 that the more extreme distribution yields a behavior of the price looking more like a real crash. But this behavior could be too extreme

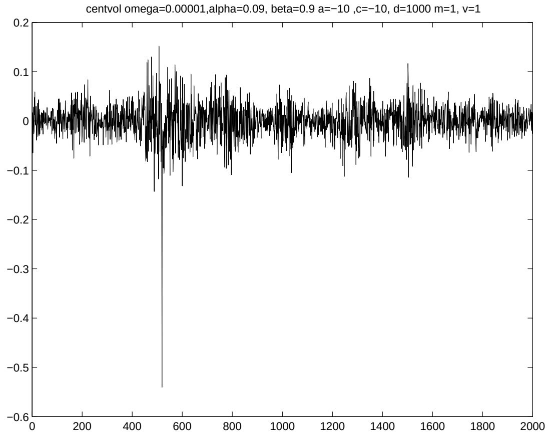


Figure 3.1: An example of a process of the volatility model

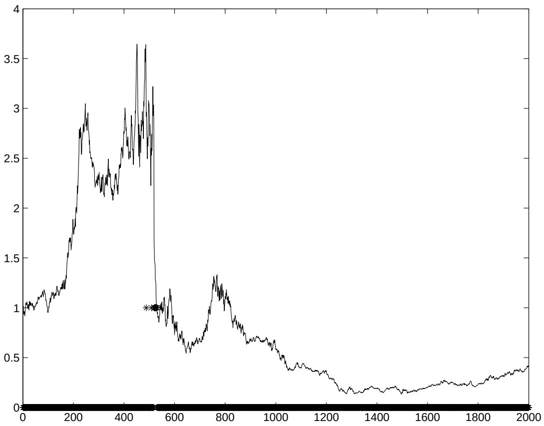


Figure 3.2: Pseudo-prices arising from the process and times of crashes

In Figure 3.3 we see the crash-probability, due to the fact that it gets very high at one point and is quite low elsewhere 3.4 zooms in a bit. We see that p_t is more structured than in the CGARCH-S model.

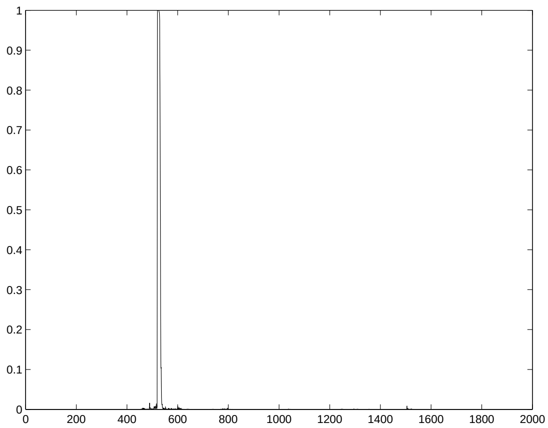


Figure 3.3: The crash-probability of the process

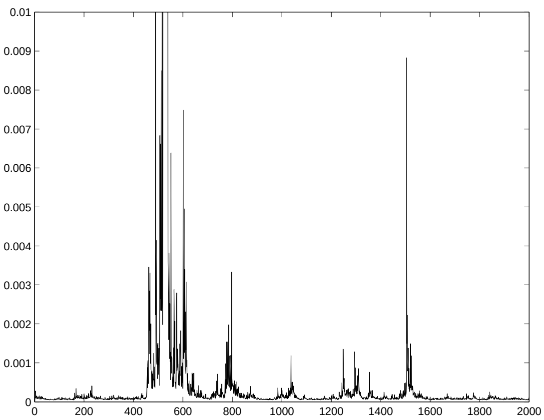


Figure 3.4: Detail of the crash-probability

We nicely see in 3.8 in connection with 3.7 that σ_t grows almost linearly with the volatility, until the volatility is so big that the steep part of the crash-probability-function is reached, then being down scaled and showing a slight growth again when the volatility is so big that p_t is almost constant again. So we see in fact a result which is rather counterintuitive: When the probability of a crash is big, σ_t becomes relatively small. This means in this model: If the probability of a

crash is big, the crash when it occurs is rather small (via down scaling). So this seems to be a major practical disadvantage of this model. These thoughts will be confirmed in the chapter concerning practical properties.
In 3.5 we see the multiplicator σ_t which is calculated via the modeled volatility and crash-probability. Here also we see that σ_t becomes quite small when p_t is close to 1.

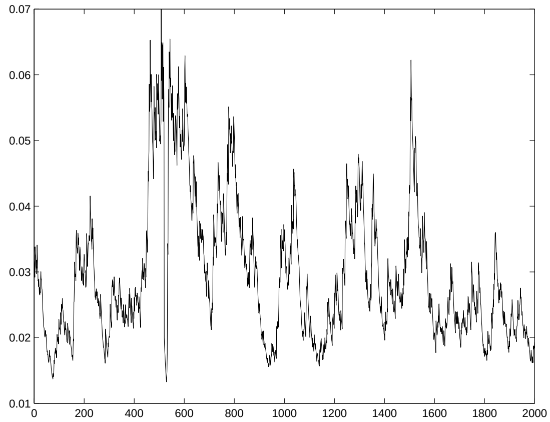


Figure 3.5: The sigma of the process

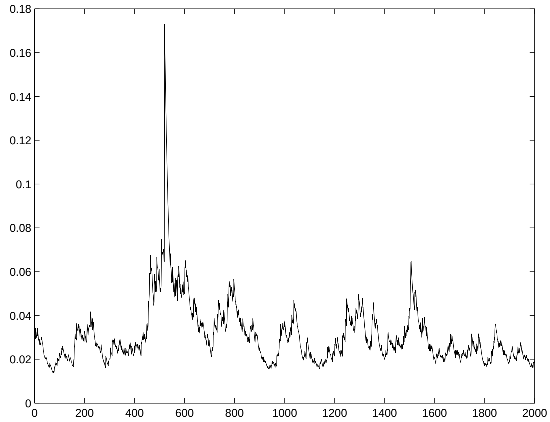


Figure 3.6: The volatility of the process

.

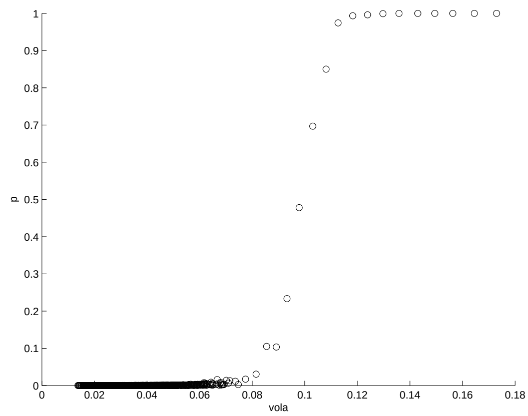


Figure 3.7: The crash-probability in dependence of the volatility

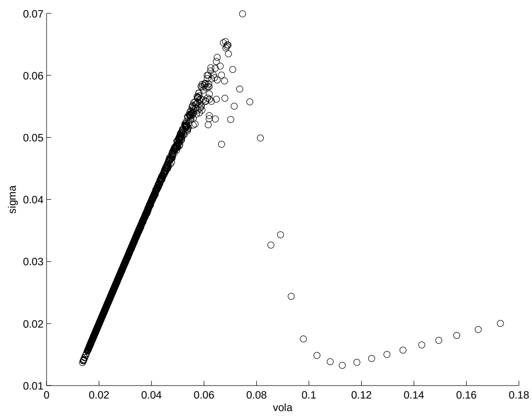


Figure 3.8: The sigma depending on the volatility

3.3 An alternative view of the model

Let $X_t = \sigma_t[(1 - B_t)\epsilon_t + B_t D_t]$, where the B_t are $\mathfrak{B}(1, p_t)$ distributed and p_t is dependent on the past. We don't assume that we have a special model for σ_t.
In the following we assume $E\epsilon_t = 0$, $E\epsilon_t^2 = \sigma_\epsilon^2$, $ED_t = \mu_D$, $VarD_t = \sigma_D^2$.
Then clearly:

$$E(X_t|\mathfrak{F}_{t-1}) = \sigma_t p_t \mu_D =: \mu_t \tag{3.4}$$

And further:

$$\begin{aligned} v_t^2 := Var(X_t|\mathfrak{F}_{t-1}) &= E((X_t - \mu_t)^2|\mathfrak{F}_{t-1}) = E(X_t^2|\mathfrak{F}_{t-1}) - \mu_t^2 \\ &= \sigma_t^2[(1 - p_t)\sigma_\epsilon^2 + p_t ED_t^2 - p_t^2\mu_D^2] \\ &= \sigma_t^2[(1 - p_t)\sigma_\epsilon^2 + p_t\sigma_D^2 + p_t(1 - p_t)\mu_D] \end{aligned} \tag{3.5}$$

Now having in mind the modeling of GARCH–like models with $\mu_t := E(X_t|\mathfrak{F}_{t-1}) \neq 0$ like ARMA–GARCH or GARCH–M, usually being define via:

$$X_t - \mu_t = \sigma_t \eta_t \tag{3.6}$$

$$\sigma_t^2 = \omega + \alpha(X_{t-1} - \mu_{t-1})^2 + \beta\sigma_{t-1}^2 \tag{3.7}$$

where η_t are iid $(0, 1)$ variables, we define

$$v_t^2 = \omega + \alpha(X_{t-1} - \mu_{t-1})^2 + \beta v_{t-1}^2 \tag{3.8}$$

In our case the $\eta_t := (1 - B_t)\epsilon_t + B_t D_t - p_t\mu_D$ are not independent. We have $E(\eta_t|\mathfrak{F}_{t-1}) = 0$ but

$$\begin{aligned} Var(\eta_t|\mathfrak{F}_{t-1}) &= E(\eta_t^2|\mathfrak{F}_{t-1}) \\ &= (1 - p_t)E\epsilon_t^2 + p_t ED_t^2 - 2p_t^2\mu_D^2 + p_t^2\mu_D^2 \\ &= (1 - p_t)\sigma_\epsilon^2 + p_t ED_t^2 - p_t^2\mu_D^2 \\ &= (1 - p_t)\sigma_\epsilon^2 + p_t\sigma_D^2 + (1 - p_t)p_t\mu_D^2 \end{aligned}$$

which is not constant in t. So v_t is not σ_t times a scaling factor.
Now we state a Lemma which motivates an alternative model in the next section and also will be helpful in the discussion of the introduced models:

Lemma 3.3.1. *Let $X_t = \mu_t + \sigma_t\eta_t$ and further μ_t and σ_t be $\mathfrak{F}_{t-1}$ measurable. If $E(\eta_t|\mathfrak{F}_{t-1}) = 0$ or $\mu_t = 0$, then $v_t^2 := Var(X_t|\mathfrak{F}_{t-1}) = \sigma_t^2 Var(\eta_t|\mathfrak{F}_{t-1})$.*

Proof.

$$\begin{aligned} Var(X_t|\mathfrak{F}_{t-1}) &= E(X_t^2|\mathfrak{F}_{t-1}) - [E(X_t|\mathfrak{F}_{t-1})]^2 \\ &= \mu_t^2 + 2\mu_t\sigma_t E(\eta_t|\mathfrak{F}_{t-1}) + \sigma_t^2 E(\eta_t^2|\mathfrak{F}_{t-1}) - \mu_t^2 - \sigma_t^2[E(\eta_t|\mathfrak{F}_{t-1})]^2 \\ &= \sigma_t^2 Var(\eta_t|\mathfrak{F}_{t-1}) \end{aligned}$$

$\square$

3.3.1 A different view of the introduced models

We now take a slightly different view on the CGARCH-S as well as the CGARCH-V model with $E\epsilon_t = 0$ and $ED_t \neq 0$. We define $\eta_t^* := (1 - B_t)\epsilon_t + B_t D_t$, $\mu_t := 0$ and $h_t^2 := Var(\eta_t^* | \mathfrak{F}_{t-1})$, and further $v_t^2 := Var(X_t | \mathfrak{F}_{t-1})$.
We have $E(X_t | \mathfrak{F}_{t-1}) = \sigma_t p_t ED_t$ and by Lemma 3.3.1 in the $\mu_t := 0$ case $\sigma_t^2 = \frac{v_t^2}{h_t^2}$.
So

$$X_t = v_t \frac{\eta_t^*}{h_t} \tag{3.9}$$

but we emphazise that $E(\eta_t^* | \mathfrak{F}_{t-1}) \neq 0$. We rather have

$$E(X_t | \mathfrak{F}_{t-1}) = \frac{v_t}{h_t} E(\eta_t^* | \mathfrak{F}_{t-1}) = \frac{v_t}{h_t} p_t ED_t \tag{3.10}$$

Therefore it makes sense to substitute σ_t by $\frac{v_t}{h_t}$.
So in the CGARCH-V model we can use this substitution directly. In the CGARCH-S model it gives us an interpretation, what the multiplicator σ_t means in terms of the conditional volatilities of the processes $\{X_t\}$ and $\{\eta_t\}$.

3.3.2 A generalization

It might be desirable that we have an additional constant in the conditional mean. We assume $ED_t = \delta \neq 0$, $E\epsilon_t = 0$, and define again:

$$\eta_t^* := (1 - B_t)\epsilon_t + B_t D_t$$

and

$$h_t^2 := Var(\eta_t^* | \mathfrak{F}_{t-1})$$

$$X_t = \mu_0 + v_t \frac{\eta_t^*}{h_t} \tag{3.11}$$

Then we get in analogy to Lemma 3.3.1

$$E(X_t | \mathfrak{F}_{t-1}) = \mu_0 + v_t \frac{E(\eta_t^* | \mathfrak{F}_{t-1})}{h_t} = \mu_0 + v_t \frac{p_t \delta}{h_t}$$

and

$$Var(X_t | \mathfrak{F}_{t-1}) = v_t^2.$$

Further, $\frac{\eta_t^*}{h_t}$ is standardized in the sense of

$$Var\left(\frac{\eta_t^*}{h_t} | \mathfrak{F}_{t-1}\right) = 1$$

Adding the constant gives us more flexibility, but certainly we are still in a GARCH–M framework.

3.4 An alternative model

The result of Lemma 3.3.1 motivates the following model, which is more flexible, because it doesn't contain a GARCH–M effect in the sense of the previous models, i.e. the conditional mean is not proportional to σ_t or v_t. The model will have the form $X_t = \mu_t + \sigma_t \eta_t$ with $\mathfrak{F}_{t-1}$ measurable μ_t and σ_t. The $\mathfrak{F}_{t-1}$ measurability assumption in Lemma 3.3.1 will cause the μ_t to rather depend on p_t directly than on B_t. To be concrete:
Let ϵ_t be iid and D_t iid, independent of each other. We assume $E(D_t) = E(\epsilon_t) = 0$ as the expected negative expectation in case of a crash is now dealt with directly via μ_t. Further, we use the notation $Var(\epsilon_t) = \sigma_\epsilon^2$ and $Var(D_t) = \sigma_D^2$. We define the model:

$$X_t = (1 - p_t)\mu + p_t\delta + \sigma_t[(1 - B_t)\epsilon_t + B_t D_t] \tag{3.12}$$

where σ_t and p_t are $\mathfrak{F}_{t-1}$ measurable, p_t lies between 0 and 1 and B_t is a $\mathfrak{B}(1, p_t)$ distributed.
To emphazise the difference to the previously introduced models we will call models of this kind ACGARCH, where A stands for alternative.

Remark 3.4.1. Modeling σ_t or v_t via a GARCH equation on and defining $p_t = f(X_{t-1}, \sigma_t)$ in the first, and $p_t = f(X_{t-1}, v_t)$ in the latter case fits into the framework. Further the GARCH equation may be centered or non centered i.e.:

$$s_t^2 = \omega + \alpha(X_{t-1} - \mu - p_{t-1}(\delta - \mu))^2 + \beta s_{t-1}^2$$

or

$$s_t^2 = \omega + \alpha X_{t-1}^2 + \beta s_{t-1}^2$$

where s may be σ or v.

Propositition 3.4.2. *Assume (3.12) and define*

1.
$$\eta_t := (1 - B_t)\epsilon_t + B_t D_t$$

2.
$$h_t^2 = (1 - p_t)\sigma_\epsilon^2 + p_t\sigma_D^2$$

3.
$$\mu_t = (1 - p_t)\mu + p_t\delta$$

such that

$$X_t = \mu_t + \sigma_t\eta_t.$$

Then

1.
$$E(\eta_t|\mathfrak{F}_{t-1}) = 0$$

2.

$$Var(\eta_t|\mathfrak{F}_{t-1}) = h_t^2$$

3.

$$E(X_t|\mathfrak{F}_{t-1}) = \mu_t$$

4.

$$v_t^2 = Var(X_t|\mathfrak{F}_{t-1}) = \sigma_t^2 h_t^2$$

Proof. μ_t and σ_t are $\mathfrak{F}_{t-1}$ measurable. Further $E(D_t) = E(\epsilon_t) = 0$ implies

$$E(\eta_t|\mathfrak{F}_{t-1}) = (1 - p_t)E\epsilon_t + p_t ED_t = 0$$

hence we can apply Lemma 3.3.1 which yields the result, together with the immediate observation that $E(\eta_t^2|\mathfrak{F}_{t-1}) = h_t^2$. □

Corollary 3.4.3. *We can rewrite model (3.12) not using σ_t in the following way:*

$$X_t = \mu_t + v_t \frac{\eta_t}{h_t} \tag{3.13}$$

In fact we will use (3.13) as an alternative definition.

Remark 3.4.4. Further, we can simplify by setting

$$\mu_t = \mu + p_t \Delta$$

with $\Delta = \delta - \mu$.

Now lets stick to the case where we model the conditional volatility v_t via a GARCH(1,1) equation, i.e. we assume

$$v_t^2 = \omega + \alpha(X_{t-1} - \mu_{t-1})^2 + \beta v_{t-1}^2. \tag{3.14}$$

We will speak of the alternative volamodel or in the short notation we introduced of the ACGARCH-V model. In fact we replaced σ_t, which in this model (and the volamodel) has no proper interpretation, by $\frac{\eta_t}{h_t}$ where both the numerator and the denominator have a clear interpretation as the "residuals" and their conditional standard deviation.
The main difference to the CGARCH-V model is that σ_t is not a factor in the conditional mean anymore.

Propositition 3.4.5. *1. If $E(B_t) = E(p_t) = p$ for all t*

$$EX_t = (1 - p)\mu + p\delta =: \mu + p\Delta \tag{3.15}$$

2.

$$Var(X_t) = Ev_t^2 + Var(\mu_t) \tag{3.16}$$

3.

$$Var(\mu_t) = Var(p_t)\Delta^2 \tag{3.17}$$

Proof. 1. Clear

2.

$$\begin{aligned} Var(X_t) &= E(X_t - \mu_t + \mu_t - EX_t)^2 \\ &= E(X_t - \mu_t)^2 - 2E(X_t - \mu_t)(EX_t - \mu_t) + E(\mu_t - EX_t)^2 = (*) \end{aligned}$$

Now $EX_t = E\mu_t$ and $EX_t\mu_t = E(\mu_t E(X_t|\mathfrak{F}_{t-1})) = E\mu_t^2$ hold.
So $E(X_t - \mu_t)(EX_t - \mu_t) = 0$, and

$$(*) = E(Var(X_t|\mathfrak{F}_{t-1})) + Var(E(X_t|\mathfrak{F}_{t-1})) = E(v_t^2) + Var(\mu_t)$$

3.

$$Var(\mu_t) = Var((1 - p_t)\mu + p_t\Delta) = Var(\mu + p_t\Delta) = Var(p_t)\Delta^2$$

□

Remark 3.4.6. $Var(X_t) = Ev_t^2 + Var(\mu_t)$ holds whenever μ_t is the conditional mean of X_t and v_t^2 is the conditional variance.

Corollary 3.4.7. *If $E(B_t) = E(p_t) = p$, $Var(p_t) = \sigma_p^2$ and $E(v_t^2) = c$ for all t, then for the model given by (3.12) and (3.14):*

1.

$$Ev_t^2 = \frac{\omega}{1 - (\alpha + \beta)}$$

2.

$$Var(X_t) = \frac{\omega}{1 - (\alpha + \beta)} + \sigma_p^2\Delta^2$$

Proof. Only point 1 remains to be shown by the standard argument using (3.14):

$$Ev_t^2 = \omega + \alpha E(X_{t-1} - \mu_{t-1})^2 + \beta Ev_{t-1}^2 = \omega + (\alpha + \beta)Ev_{t-1}^2$$

□

Chapter 4

Asymptotics in the pure ARCH case

4.1 Markov Methods

For its own sake and in order to investigate the asymptotic properties of maximum likelihood style estimator the question arises under which conditions the introduced models are stationary and ergodic and satisfy certain mixing conditions. The strategy we will use to do that will heavily use Markov chain theory. To apply the methodology of Markov chains directly, we must restrict ourselves to models of pure ARCH form. We have to do this, because in the context with genuine GARCH dynamic, X_t is not a Markov chain. Consider the GARCH(1,1) case. $\begin{pmatrix} X_t \\ \sigma_t \end{pmatrix}$ is a Markov chain, but the character of σ_t as an one step prediction makes σ_{t+1} predictable from $\begin{pmatrix} X_t \\ \sigma_t \end{pmatrix}$. So the transition function $p\left(\begin{pmatrix} X_{t+1} \\ \sigma_{t+1} \end{pmatrix} \middle| \begin{pmatrix} X_t \\ \sigma_t \end{pmatrix}\right)$ is in the second coordinate the one point measure $\delta_{\omega+\alpha X_{t-1}^2+\beta\sigma_{t-1}^2}$, while the first coordinate of it is given by a Lebesgue density. Because the σ_ts are principally defined for the positive real line, we cannot restrict the second coordinate to a countable set. Hence we cannot use the arguments we will use in the following for the ARCH case, because regularity conditions on the transition probabilities will not be satisfied.

Consider the ARCH(1) dynamic in the CGARCH-S setting: So let us restrict to the model

$$X_t = \sigma_t[(1-B_t)\epsilon_t + B_t D_t],\ \mathcal{L}(B_t|\mathfrak{F}_{t-1}) = \mathfrak{B}(1, p_t)$$

with $\sigma_t^2 = \omega + \alpha X_{t-1}^2$ and now $p_t = g(X_{t-1})$, as in the pure ARCH case σ_t is just a function of X_{t-1}, and therefore a specification of p_t as $f(X_{t-1}, \sigma_t)$ just boils down to p_t being a function of X_{t-1} only. Then we can formally write down the transition function as:

$$\begin{aligned}
&pr(X_t \in [x; x+dx]|X_{t-1}) \\
&= pr(\sigma_t \epsilon_t \in [x; x+dx], B_t = 1|X_{t-1}) + pr(\sigma_t D_t \in [x; x+dx], B_t = 0|X_{t-1}) \\
&= pr\left(\epsilon_t \in \left[\frac{x}{\sigma_t}; \frac{x+dx}{\sigma_t}\right], B_t = 1|X_{t-1}\right) \\
&\quad + pr\left(D_t \in \left[\frac{x}{\sigma_t}; \frac{x+dx}{\sigma_t}\right], B_t = 0|X_{t-1}\right)
\end{aligned}$$

So the transition density of X_t given $X_{t-1} = z$ is

$$p(x|z) = \frac{1}{\sigma(z)} f_\epsilon\left(\frac{x}{\sigma(z)}\right) g(z) + \frac{1}{\sigma(z)} f_D\left(\frac{x}{\sigma(z)}\right)(1 - g(z))$$

where f_ϵ, f_D denote the desities of ϵ_t and D_t and $\sigma^2(z) = \omega + \alpha z^2$. Using this fact we then will be able to exploit Markov chain theory.

4.2 Some Markov Chain Theory

Let $\{X_t\}$ be a Markov chain on $(\mathbb{R}^m, \mathfrak{B})$.

Definition 4.2.1. *Let ϕ be a nontrivial σ–finite measure. A Markov chain is called ϕ–irreducible if for all A with $\phi(A) > 0$, $P(X_n \in A|X_0 = x) =: P^n(x, A)$ satisfies*

$$\sum_{n=1}^{\infty} P^n(x, A) > 0$$

It is called irreducible if it is ϕ–irreducible for some ϕ.

Propositition 4.2.2. *[Ton90] If $\{X_t\}$ is irreducible there exits a maximal irreducibility measure ψ. With*

1. *$\{X_t\}$ is ψ–irreducible.*
2. *If $\{X_t\}$ is ϕ–irreducible, ϕ is absolutely continuous with respect to ψ.*
3. $$\psi(A) = 0 \Rightarrow \psi(\{x| \sum_{n-1}^{\infty} P^n(x, A) > 0\}) = 0$$

Definition 4.2.3. 1. *A set $C \in \mathfrak{B}$ is small if there exists a positive integer k, a constant $b > 0$ and a nontrivial probability measure $\varphi(.)$ such that*

$$P^k(x, A) \geq b\varphi(A), \forall x \in C, A \in \mathfrak{B} \tag{4.1}$$

2. *A set $C \in \mathfrak{B}$ is ν_a petite if there exists a probability measure a on $\mathbb{N}$ such that*

$$\sum_{n=1}^{\infty} P^n(x, A)a(n) > \nu_a(A), \forall x \in C, A \in \mathfrak{B} \tag{4.2}$$

for a nontrivial measure ν_a.

Definition 4.2.4. *Assume there exists a small set for $\{X_t\}$.*

1. *If C is small, then $I(C) := \{k|P^k(x, A) \geq b\varphi(A), x \in C, A \in \mathfrak{B}\}$.*
2. *We define $d(C)$ to be the greatest common divisor of $I(C)$. It can be shown that $d(C)$ doesn't depend on C. So we write just d.*
3. *If $d = 1$, then $\{X_t\}$ is called aperiodic.*

Definition 4.2.5. *Let $\{X_t\}$ be a Markov chain with transition density $p(x|y)$. $\{X_t\}$ has the Feller property if*

$$E(h(X_t)|X_{t-1} = y) =: Ph(y) = \int h(y)p(x|y)dx$$

is continuous and bounded for all continuous and bounded functions h.

Lemma 4.2.6. *Let $\{X_t\}$ be irreducible and aperiodic, ψ its maximal irreducibility measure. If $\{X_t\}$ has the Feller property and supp(ψ) has a non–empty interior, then all compact sets are small.*

Proof. If $\{X_t\}$ has the Feller property , is ψ irreducible and $supp(\psi)$ has a non–empty interior then by [MT93] Proposition 6.2.8 every compact set is petite. If $\{X_t\}$ is also aperiodic then by [MT93] Theorem 5.5.7 every petite set is small. So every compact set is small. □

Definition 4.2.7. *$\{X_t\}$ is geometrically ergodic if there exists a probability measure π on $(\mathbb{R}, \mathfrak{B})$, a positive constant $\rho < 0$ and a π-integrable non-negative measurable function h such that*

$$||P^n(x, .) - \pi(.)||_\tau \leq \rho^n h(x) \tag{4.3}$$

where $||.||_\tau$ is the total variation norm.

Theorem 4.2.8. *[Ton90] Let $\{X_t\}$ be aperiodic and irreducible. Suppose that there exists a small set C a non–negative measurable function h and constants $r > 1$, $\gamma > 0$ and $B > 0$ such that*

$$E(rh(X_t)|X_{t-1} = y) < h(y) - \gamma, \; y \notin C \tag{4.4}$$

and

$$E(h(X_t)|X_{t-1} = y) < B, \; y \in C \tag{4.5}$$

Then $\{X_t\}$ is geometrically ergodic.

4.3 Mixing

Definition 4.3.1. *1. Let $(\Omega, \mathfrak{A}, P)$ be a probability space and $\mathfrak{B}$, $\mathfrak{C}$ two sub σ algebras.*

$$\alpha(\mathfrak{B}, \mathfrak{C}) = sup_{B \in \mathfrak{B}, C \in \mathfrak{C}} |P(B \cap C) - P(B)P(C)|$$

2. For a process $\{X_t : t \in \mathbb{Z}\}$ and k we define:

$$\alpha(k) = \sup_{t \in \mathbb{Z}} \alpha(\sigma(X_s, s \leq t), \sigma(X_s : s \geq t + k))$$

3. $\{X_t\}$ is said to be α-mixing if

$$\lim_{k \to \infty} \alpha(k) = 0$$

Theorem 4.3.2. *[Dav73] Let $\{X_t\}$ be a geometric ergodic Markov chain and X_0 have the distribution π, where π denotes the stationary measure of $\{X_t\}$. Then $\{X_t\}$ is geometrically α-mixing. That means$\{X_t\}$ is α-mixing and*

$$\exists 0 < \rho < 1, c > 0 \text{ such that } \forall n \in \mathbb{N} : \alpha(n) \leq c\rho^n \tag{4.6}$$

Lemma 4.3.3. *[TK05] Lemma 5.11 Let $\{X_t\}$ be strictly stationary and α-mixing. Let $Y_t = f(X_{t-1}, \ldots, X_{t-p})$ where f is a measurable function. Then $\{Y_t\}$ is strictly stationary and α-mixing and its mixing coefficients decrease with the same order than that of $\{X_t\}$.*

4.4 Stationarity in the pure ARCH case of the original model (the CARCH-S model)

We work in the CGARCH-S setting. Let' s consider the special case, where we have a pure ARCH(1) dynamic on σ_t^2. That means $\sigma_t^2 = \omega + \alpha X_{t-1}^2$. Then certainly $p_t = f(X_{t-1}, \sigma_t^2) = g(X_{t-1})$. So the laws of B_t and σ_t are completely determined by X_{t-1}. But that implies $\mathcal{L}(X_t|\mathfrak{F}_{t-1}) = \mathcal{L}(X_t|X_{t-1})$. So we get the following proposition.

Propositition 4.4.1. *A process X_t of our model class which has a pure ARCH(1) dynamic on σ_t^2 is a Markov process.*

For the rest of this section we keep the assumption of having a pure ARCH(1) dynamic on σ_t^2. Now if we denote the density of ϵ_t as f_ϵ and the density of D_t as f_D and further as above $p_t = g(X_{t-1})$ we get the following formula of the transition densities $p(x|y)$ using the results of section 4.1.

$$p(x|y) = \frac{1}{(\omega + \alpha y^2)^{\frac{1}{2}}} \left[(1 - g(y)) f_\epsilon \left(\frac{x}{(\omega + \alpha y^2)^{\frac{1}{2}}} \right) + g(y) f_D \left(\frac{x}{(\omega + \alpha y^2)^{\frac{1}{2}}} \right) \right]$$

Lemma 4.4.2. *If g, f_ϵ, f_D are continuous then X_t has the Feller property.*

Proof. Let $h(x)$ be continuous and bounded. Then $Ph(y)$ is continuous if g,f_ϵ and f_D are continuous. The transformation theorem for integrals yields:

$$\frac{1}{(\omega+\alpha y^2)^{\frac{1}{2}}}\int h(x) f_*\left(\frac{x}{(\omega+\alpha y^2)^{\frac{1}{2}}}\right)dx = \int h(z(\omega+\alpha y^2)^{\frac{1}{2}}) f_*(z)dz$$

But $h(z(\omega+\alpha y^2)^{\frac{1}{2}})$ is bounded by choice of h. Let it be bounded e.g. by δ. So

$$\int h(z(\omega+\alpha y^2)^{\frac{1}{2}}) f_*(z)dz \leq \int \delta f_*(z)dz = \delta \int f_*(z)dz = \delta$$

because f_* is a probability density function. Now g is bounded because g maps into $[0,1]$ and so $Ph(y)$ is bounded.

□

Lemma 4.4.3. *Suppose one of the following assumptions hold*

1. *The support of f_ϵ and the support of f_D is $\mathbb{R}$.*
2. *The support of f_ϵ is $\mathbb{R}$ and $\forall t : p_t \leq 1-\delta < 1$.*
3. *$supp(f_\epsilon) \cup supp(f_D) = \mathbb{R}$ and $\forall t : 0 < \gamma \leq p_t \leq 1-\delta < 1$.*

Then X_t is irreducible and aperiodic.

Proof. Let A be a Borel set with $\lambda(A) > 0$ where λ denotes the Lesbegue measure. Now

$$p(A|y) = \int_A (1-g(y))\frac{1}{(\omega+\alpha y^2)^{\frac{1}{2}}} f_\epsilon\left(\frac{x-\mu(y)}{(\omega+\alpha y^2)^{\frac{1}{2}}}\right) + g(y)\frac{1}{(\omega+\alpha y^2)^{\frac{1}{2}}} f_D\left(\frac{x-\mu(y)}{(\omega+\alpha y^2)^{\frac{1}{2}}}\right)dx$$

$$= (1-g(y))\int_A \frac{1}{(\omega+\alpha y^2)^{\frac{1}{2}}} f_\epsilon\left(\frac{x-\mu(y)}{(\omega+\alpha y^2)^{\frac{1}{2}}}\right)dx + g(y)\int_A \frac{1}{(\omega+\alpha y^2)^{\frac{1}{2}}} f_D\left(\frac{x-\mu(y)}{(\omega+\alpha y^2)^{\frac{1}{2}}}\right)dx$$

If the support of f_ϵ and the support of f_D is $\mathbb{R}$ this is clearly bigger than zero. If the support of f_ϵ is $\mathbb{R}$ and $\forall t : p_t \leq 1-\delta < 1$ then

$$\int_A \frac{1}{(\omega+\alpha y^2)^{\frac{1}{2}}} f_\epsilon\left(\frac{x-\mu(y)}{(\omega+\alpha y^2)^{\frac{1}{2}}}\right)dx > 0$$

and $1-g(y) \geq \delta > 0$ So $p(A|y) > 0$. But $\sum_{n=1}^{\infty} P^n(x,A) \geq P(x,A)$ holds.
If $supp(f_\epsilon) \cup supp(f_D) = \mathbb{R}$ and $\forall t : 0 < \gamma \leq p_t \leq 1-\delta < 1$:

$$\int_A \frac{1}{(\omega+\alpha y^2)^{\frac{1}{2}}} f_\epsilon\left(\frac{x-\mu(y)}{(\omega+\alpha y^2)^{\frac{1}{2}}}\right)dx > 0$$

and $1 - g(y) \geq \delta > 0$ or

$$\int_A \frac{1}{(\omega + \alpha y^2)^{\frac{1}{2}}} f_D \left(\frac{x - \mu(y)}{(\omega + \alpha y^2)^{\frac{1}{2}}} \right) dx > 0$$

and $g(y) \geq \gamma > 0$. It follows $p(A|y) > 0$ in all three cases. But $\sum_{n=1}^{\infty} P^n(x, A) \geq P(x, A) = p(A|x) > 0$. In conclusion $\{X_t\}$ is λ–irreducible. $\square$

Remark 4.4.4. Formally it might be more correct to state the boundedness conditions on p_t in the preceding lemma in terms of $g(y)$ meaning we have to claim

$$g(y) \leq 1 - \delta < 1$$

almost everywhere or

$$0 < \gamma \leq g(y) \leq 1 - \delta < 1$$

almost everywhere. But in my opinion the form chosen seems to emphasize more clearly the intentioned meaning of the conditions.

Lemma 4.4.5. *Let the assumptions of Proposition 4.4.2 and one of the assumptions of Lemma 4.4.3 hold. Then for $\{X_t\}$ any compact set is a small set.*

Proof. $\{X_t\}$ is irreducible and aperiodic. For the maximal irreducibility measure ψ holds

$$\sum_{n=1}^{\infty} P^n(x, A) \geq P(x, A) = p(A|x) > 0$$

so by 4.2.23 $\psi(A) > 0$ holds for the maximal irreducibility measure ψ. So $supp(\psi)$ is nonempty. By Lemma 4.2.6 follows the assumption. $\square$

Theorem 4.4.6. *Let the assumptions of Proposition 4.4.2 and one of the assumptions of Lemma 4.4.3 hold. If*

$$ED_t^2 \geq E\epsilon_t^2 \text{ and } \alpha(E\epsilon_t^2 + \sup_y g(y)(ED_t^2 - E\epsilon_t^2)) < 1,$$

or

$$ED_t^2 \leq E\epsilon_t^2 \text{ and } \alpha(E\epsilon_t^2 + \inf_y g(y)(ED_t^2 - E\epsilon_t^2)) < 1$$

then X_t is geometric ergodic.

Proof. We will use the drift condition 4.2.8. Choose $h(y) = 1 + y^2$. Then $h(X_t) = 1 + (1 - B_t)\sigma_t^2\epsilon_t^2 + B_t\sigma_t^2 D_t^2$. If have $ED_t^2 \geq E\epsilon_t^2$ we define $g_\infty := \sup_y g(y)$. If we assume $ED_t^2 \leq E\epsilon_t^2$ we define $g_\infty := \inf_y g(y)$.
Then the following holds:

$$\frac{E(h(X_t)|X_{t-1} = y) - h(y)}{h(y)} = \frac{(\omega + \alpha y^2)[(1 - g(y))E\epsilon_t^2 + g(y)ED_t^2] - y^2}{1 + y^2}$$

$$\leq \frac{(\omega + \alpha y^2)[(E\epsilon_t^2 + g_\infty(ED_t^2 - E\epsilon_t^2)] - y^2}{1+y^2} = (*)$$

We define

$$A := E\epsilon_t^2 + g_\infty(ED_t^2 - E\epsilon_t^2)$$

Then the following holds:

$$(*) = \frac{(\omega + \alpha y^2)A - y^2}{1+y^2} = \frac{\omega A - (1-\alpha A)y^2}{1+y^2} \rightarrow_{y^2 \rightarrow \infty} -(1-\alpha A)$$

Therefore, for every M exists a $\Delta > 0$ such that

$$(*) \leq -(1-\alpha A) + \Delta$$

for $|y| > M$.Now if $\alpha A < 1$ then $-(1-\alpha A) < 0$.
If we choose M big enough Δ will be such that $-(1-\alpha A) + \Delta < 0$. Define $C := \{y||y| \leq M\}$ for a such an M. Then

$$(*) \leq -(1-\alpha A) + \Delta =: -A^* < 0 \text{ if } y \notin C \text{ if } \alpha A < 1$$

For $y \in C$

$$* \leq \frac{\omega A}{1+y^2} \leq \omega A$$

holds.

$$\Rightarrow E(h(X_t)|X_{t-1} = y) \leq (1+\omega A)h(y) \leq (1+\omega A)(1+M^2) =: B$$

For $y \notin C$:

$$E(h(X_t)|X_{t-1} = y) \leq (1-A^*)h(y) = (1-\frac{A^*}{2})h(y) - \frac{A^*}{2}h(y) < (1-\frac{A^*}{2})h(y) - \frac{A^*}{2}(1+M^2)$$

We define $r := (1-\frac{A^*}{2})$ and $\gamma = -\frac{A^*}{2}(1+M^2)$. $\square$

We now gather all assumptions we used to gain the geometric ergodicity result:

Assumption 4.4.7. ***Assumptions assuring geometric ergodicity***

1. *g, f_ϵ, f_D are continuous.*
2. *ϵ_t and D_t are square integrable.*
3. *One of the following conditions holds:*
   (a) *The support of f_ϵ and the support of f_D is $\mathbb{R}$.*
   (b) *The support of f_ϵ is $\mathbb{R}$ and $\forall t : p_t \leq 1-\delta < 1$.*
   (c) *$supp(f_\epsilon) \cup supp(f_D) = \mathbb{R}$ and $\forall t : 0 < \gamma \leq p_t \leq 1-\delta < 1$.*

4. One of the following conditions holds:

(a) $ED_t^2 \geq E\epsilon_t^2$ and $\alpha(E\epsilon_t^2 + \sup_y g(y)(ED_t^2 - E\epsilon_t^2)) < 1$.

(b) $ED_t^2 \leq E\epsilon_t^2$ and $\alpha(E\epsilon_t^2 + \inf_y g(y)(ED_t^2 - E\epsilon_t^2)) < 1$.

Corollary 4.4.8. *Assume Assumption 4.4.7 holds. If X_0 is distributed as the invariant measure π, $\{X_t\}$ is strongly stationary and geometric α–mixing.*

Proof. If X_0 is distributed as the invariant measure π, $\{X_t\}$ clearly is strongly stationary. Due to Theorem 4.3.2 it is also geometric α–mixing. □

Remark 4.4.9. Let us consider a process $\{X_t\}$ of our model class with an ARCH(q) dynamic on σ_t. Again we define $g_\infty := \sup_y g(y)$, if we have $ED_t^2 \geq E\epsilon_t^2$ and if we assume $ED_t^2 \leq E\epsilon_t^2$ we define $g_\infty := \inf_y g(y)$.
Then X_t is geometrically ergodic if

$$\sum_{i=1}^{q} \alpha_i (E\epsilon_t^2 - g_\infty (ED_t^2 - E\epsilon_t^2) < 1$$

For that purpose, we realize that in that case $(X_t, ..., X_{t-q-1})^T$ is a Markov chain. For irreducibility, aperiodicity and Feller property we can use the same arguments as above. And the drift condition can be verified using

$$h(y_1, \dots y_q) = \sum_{i=1}^{q} y_i^2$$

Using

$$(E\epsilon_t^2 - g_\infty (ED_t^2 - E\epsilon_t^2) \sum_{i=1}^{q} \alpha_i y_i^2 \leq (E\epsilon_t^2 - g_\infty (ED_t^2 - E\epsilon_t^2) \left(\sum_{i=1}^{q} \alpha_i \right) \left(\sum_{i=1}^{q} y_i^2 \right)$$

and

$$\left((E\epsilon_t^2 - g_\infty (ED_t^2 - E\epsilon_t^2) \sum_{i=1}^{q} \alpha_i \right) - 1 < 0$$

we can reproduce the proof of 4.4.6.

4.5 Stationarity in the pure ARCH case of the alternative volamodel (the ACARCH-V model)

We now consider the pure ARCH(1) case of the alternative model, in our short notation this is the ACARCH-V model. That means we consider:

$$X_t = \mu_t + v_t \frac{\eta_t}{h_t} \tag{4.7}$$

where

$$\eta_t = (1 - B_t)\epsilon_t + B_t D_t,\ \mathcal{L}(B_t|\mathfrak{F}_{t-1}) = \mathfrak{B}(1, p_t),\ p_t = g(X_{t-1}),\ E\epsilon_t = ED_t = 0 \tag{4.8}$$

further

$$\mu_t = (1 - p_t)\mu + p_t\delta = \mu + p_t\Delta \text{ with } \Delta = \delta - \mu \tag{4.9}$$

and

$$h_t^2 := E(\eta_t^2|\mathfrak{F}_{t-1}) \tag{4.10}$$

and the ARCH dynamic coming from

$$Var(X_t|\mathfrak{F}_{t-1}) = v_t^2 = \omega + \alpha X_{t-1}^2 \tag{4.11}$$

Remark 4.5.1. If we consider

$$v_t^2 = \omega + \alpha(X_{t-1} - \mu - p_{t-1}\Delta)^2$$

instead, then everything works out in the same way as

$$p_{t-1} = g(X_{t-2}),$$

therefore v_t^2 satisfies a nonlinear ARCH(2) relation and $(X_t, X_{t-1})^T$ is a Markov chain.

Remark 4.5.2. If, however, we consider

$$v_t^2 = \omega + \alpha(X_{t-1} - \mu - p_{t-1}\Delta)^2$$

and

$$p_t = f(X_{t-1}, v_t),$$

with explicit dependence on v_t, then this does not reduce to the case $p_t = g(X_{t-1})$. Moreover as p_{t-1} depends on X_{t-2} and v_{t-1}, v_t^2 satisfies a nonlinear GARCH(2,1)-relation So we cannot apply the Markov chain technology due to the discussion in section 4.2.

Remark 4.5.3. As the arguments for the corresponding results in Proposition 3.4.2 don't depend on the particular modelling of v_t^2 we have:

1. $E(\eta_t|\mathfrak{F}_{t-1}) = 0$
2. $h_t^2 = (1 - p_t)\sigma_\epsilon^2 + p_t\sigma_D^2$

Propositition 4.5.4. *X_t is a Markov chain with conditional density of X_t given $X_{t-1} = y$*

$$p(x|y) = (1 - g(y))\frac{h(y)}{v(y)} f_\epsilon \left(\frac{x - \mu(y)}{v(y)} h(y) \right) + g(y)\frac{h(y)}{v(y)} f_D \left(\frac{x - \mu(y)}{v(y)} h(y) \right) \tag{4.12}$$

where $\mu(y) = \mu + g(y)\Delta$, $v^2(y) = \omega + \alpha y^2$ and $h^2(y) = (1 - g(y))\sigma_\epsilon^2 + g(y)\sigma_D^2$.

Proof. The conditional density clearly is implied by the definition of the model and the relations above. It only depends on y which shows the Markov property. □

Propositition 4.5.5. *If g, f_ϵ, f_D are continuous, $\omega > 0$ then $\{X_t\}$ satisfies the Feller property.*

Proof. If $g(y)$ is continuous so are $\mu(y)$ and $h(y)$. $v(y)$ is continuous anyway and $\frac{1}{v(y)}$ is bounded if $\omega > 0$.
Now let $\psi(y)$ be a bounded continuous function with $\sup |\psi| \leq c < \infty$.

$$P\psi(y) = \int \psi(x) p(x|y) dx$$

$$= (1 - g(y))\frac{h(y)}{v(y)} \int \psi(x) f_\epsilon \left(\frac{x - \mu(y)}{v(y)} h(y) \right) dx + g(y)\frac{h(y)}{v(y)} \int \psi(x) f_D \left(\frac{x - \mu(y)}{v(y)} h(y) \right) dx$$

Obviously under the assumptions of the theorem $P\psi(y)$ is continuous. Now $0 \leq g \leq 1$ also implies that $\mu(y)$ and $h(y)$ are bounded and so also $\frac{h(y)}{v(y)}$ is bounded. So

$$\left| \frac{h(y)}{v(y)} \int \psi(x) f_\epsilon \left(\frac{x - \mu(y)}{v(y)} h(y) \right) dx \right| \leq \sup |\psi| \int f_\epsilon(u) du = c1 = c < \infty$$

and

$$\left| \frac{h(y)}{v(y)} \int \psi(x) f_D \left(\frac{x - \mu(y)}{v(y)} h(y) \right) dx \right| \leq \sup |\psi| \int f_D(u) du = c1 = c < \infty$$

From this the boundedness of g implies the boundedness of $P\psi(y)$. So $\{X_t\}$ satisfies the Feller property. □

Lemma 4.5.6. *Suppose one of the following holds:*

1. *The support of f_ϵ and the support of f_D is $\mathbb{R}$.*

2. *The support of f_ϵ is $\mathbb{R}$ and $\forall t : p_t \leq 1 - \delta < 1$.*

3. *$supp(f_\epsilon) \cup supp(f_D) = \mathbb{R}$ and $\forall t : 0 < \gamma \leq p_t \leq 1 - \delta < 1$.*

then X_t is irreducible and aperiodic.

Proof. Let A be a Borel set with $\lambda(A) > 0$ where λ denotes the Lebesgue measure.

$$p(A|y) = \int_A (1-g(y))\frac{h(y)}{v(y)} f_\epsilon \left(\frac{x-\mu(y)}{v(y)}h(y)\right) + g(y)\frac{h(y)}{v(y)} f_D \left(\frac{x-\mu(y)}{v(y)}h(y)\right) dx$$

$$= (1-g(y))\int_A \frac{h(y)}{v(y)} f_\epsilon \left(\frac{x-\mu(y)}{v(y)}h(y)\right) dx + g(y)\int_A \frac{h(y)}{v(y)} f_D \left(\frac{x-\mu(y)}{v(y)}h(y)\right) dx$$

If the support of f_ϵ and the support of f_D is $\mathbb{R}$ this is clearly bigger than zero. If the support of f_ϵ is $\mathbb{R}$ and $\forall t : p_t \leq 1-\delta < 1$ then

$$\int_A \frac{h(y)}{v(y)} f_\epsilon \left(\frac{x-\mu(y)}{v(y)}h(y)\right) dx > 0$$

and $1-g(y) \geq \delta > 0$ So $p(A|y) > 0$. If $supp(f_\epsilon) \cup supp(f_D) = \mathbb{R}$ and $\forall t : 0 < \gamma \leq p_t \leq 1-\delta < 1$.

$$\int_A \frac{h(y)}{v(y)} f_\epsilon \left(\frac{x-\mu(y)}{v(y)}h(y)\right) dx > 0$$

and $1-g(y) \geq \delta > 0$ or

$$\int_A \frac{h(y)}{v(y)} f_D \left(\frac{x-\mu(y)}{v(y)}h(y)\right) dx > 0$$

and $g(y) \geq \gamma > 0$ It follows $p(A|y) > 0$ in all three cases. Hence $\{X_t\}$ is irreducible and aperiodic. □

Lemma 4.5.7. *Let the assumptions of Proposition 4.5.5 and one of the assumptions of Lemma 4.5.6 hold, then for $\{X_t\}$ any compact set is a small set.*

Proof. Completely analogous to the proof of Lemma 4.4.5. □

Now we use theorem A1.5 of [Ton90] again.

Theorem 4.5.8. *Let the assumptions of Proposition 4.5.5 hold. If $\alpha < 1$ and one assumption of Lemma 4.5.6 is satisfied, then $\{X_t\}$ is geometric ergodic.*

Proof. Again we use the Theorem 4.2.8. Choose $\gamma(y) = 1 + y^2$. Then

$$\gamma(X_t) = 1 + (\mu_t + \frac{v_t}{h_t}[(1-B_t)\epsilon_t + B_t D_t])^2$$

Because $B_t \in \{0,1\}$ this is:

$$1 + \mu_t^2 + 2\frac{\mu_t v_t}{h_t}[(1-B_t)\epsilon_t + B_t D_t] + (1-B_t)\frac{v_t^2}{h_t^2}\epsilon_t^2 + B_t\frac{v_t^2}{h_t^2}D_t^2$$

Taking (conditional) expectation the mixed term vanishes, because $E\epsilon_t = ED_t = 0$. So

$$\frac{E(\gamma(X_t)|X_{t-1}=y)-h(y)}{\gamma(y)} = \frac{1+\mu(y)^2+\frac{v(y)^2}{h(y)^2}[(1-g(y))\sigma_\epsilon^2+g(y)\sigma_D^2]-(1+y^2)}{1+y^2} = (**) \tag{4.13}$$

Because of the equation for h_t^2 in Remark 4.5.3:

$$(**) = \frac{1+\mu(y)^2+v^2(y)-(1+y^2)}{1+y^2} = \frac{(\mu+\Delta g(y))^2+\omega+\alpha y^2-y^2}{1+y^2} = (*)$$

Because $0 \le g \le 1$, there exists an $\tilde{\omega} \ge (\mu+\Delta g(y))^2+\omega$ for all y. For such an $\tilde{\omega}$ the following holds:

$$(*) \le \frac{\tilde{\omega}+(\alpha-1)y^2}{1+y^2} \to_{|y|\to\infty} -(1-\alpha)$$

If $\alpha < 1$then $-(1-\alpha) < 0$ So we can find an M and a corresponding Δ such that $(*) < -(1-\alpha)+\Delta := A^* > 0$ for all y with $|y| > M$. Define $C := \{y||y| \le M\}$ for a such an M. The rest of the proof is completely the same as the proof of 4.4.6. ☐

Again we gather the conditions needed for the geometric ergodicity:

Assumption 4.5.9. ***Assumptions assuring geometric ergodicity***

1. *g, f_ϵ, f_D are continuous.*
2. *ϵ_t and D_t are square integrable.*
3. *One of the following conditions holds:*
   (a) *The support of f_ϵ and the support of f_D is $\mathbb{R}$.*
   (b) *The support of f_ϵ is $\mathbb{R}$ and $\forall t : p_t \le 1-\delta < 1$.*
   (c) *supp$(f_\epsilon) \cup$ supp$(f_D) = \mathbb{R}$ and $\forall t : 0 < \gamma \le p_t \le 1-\delta < 1$.*
4. *$\alpha < 1$.*

Again we can conclude:

Corollary 4.5.10. *Assume Assumption 4.5.9 holds. If X_0 is distributed as the invariant measure π, $\{X_t\}$ is strongly stationary and geometric α–mixing.*

Proof. If X_0 is distributed as the invariant measure π, $\{X_t\}$ clearly is strongly stationary. Due to Theorem 4.3.2 it is also geometric α–mixing. ☐

Remark 4.5.11. Again the geometric ergodicity can be achieved for

$$v_t^2 = \omega + \sum_{i=1}^{q} \alpha_i X_{t-i}^2$$

if

$$\sum_{i=1}^{q} \alpha_i < 1$$

the argument being the one of Remark 4.4.9.

Chapter 5

Estimation

5.1 The general model

We try to estimate the parameters of our model with a conditional maximum likelihood estimator. We maximize the conditional likelihood function which is given by:

$$\prod_{t=1}^{n} f(X_t|X_{t-1})$$

i.e. we neglect the marginal density of $f(X_0)$ in the likelihood function:

$$f(X_0)\prod_{i=1}^{n} f(X_t|X_{t-1})$$

That is done by working with the logarithm of the conditional likelihood function neglecting the starting value. If we choose the ϵ_t to be iid $N(0,1)$ distributed and the D_t to be iid the negative of lognormal distributed random variables with parameters (m, s^2) the conditional log-likelihood-function is:

$$\sum_{i=1}^{n} \log\left((1-p_t)\frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left(-\frac{X_t^2}{2\sigma_t^2}\right)\right)$$

$$+p_t\frac{1}{\sqrt{2\pi s^2 X_t^2}}\exp\left(-\frac{(\log(-X_t)-(m+\log(\sigma_t)))^2}{2s^2}\right)\mathbb{I}_{\{X_t<0\}})$$

where m and s^2 are the parameters of the chosen lognormal distribution. To apply the general theory for consistency as in [PP97] and [GW88] we have to establish first a local law of large numbers. Therefore we must show that the processes are either L_0 approximable by some α mixing process, or are near epoch dependent.

In the special case of the exogenous crash model the process $\eta_t = (1-B_t)\epsilon_t + B_t D_t$, which takes the role of the random source ϵ_t in the conventional GARCH(1,1) model, is under the corresponding condition on the exogenous variables W_t strongly stationary and mixing, and hence ergodic, we are in the situation to apply the theorem of [LH94]. This means we can get consistent estimates of the GARCH–parameters using QMLE for GARCH(1,1). But we do not get an estimator for the complete vector of parameters. We can use (and in fact I did use) QMLE for GARCH(1,1) to get initial estimate of the GARCH–sub vector. The assumptions in 1.4.3 are fulfilled so this estimate will be consistent. If additionally the assumptions in 1.4.4 hold it will also be asymptotically normal. We also get a hint, if a method estimating the whole vector works. The estimated sub vector shouldn't be to far away from the estimated values arising from QMLE for GARCH(1,1)

5.2 The methodology of Pötscher and Prucha (1997)

We now want to introduce and use the asymptotic theory for M–estimators for dependent and heterogenous processes developed in [PP97] and [GW88].
To do so we first have to do some definitions.

Definition 5.2.1. *1. Let $r \geq 1$: A process X_t is said to be L_r approximable by some basis process e_t if there exist measurable functions h_t^m such that $\lim_{n\to\infty} \sup n^{-1} \sum_{t=1}^n ||X_t - h_t^m(e_{t+m}, \dots, e_{t-m})||_r \to 0$ as $m \to \infty$.*

2. A process X_t is said to be L_0 approximable by some basis process e_t if there exist measurable functions h_t^m such that $\lim_{n\to\infty} \sup n^{-1} \sum_{t=1}^n P(|X_t - h_t^m(e_{t+m}, \dots, e_{t-m})| > \delta) \to 0$ as $m \to \infty$ for all $\delta > 0$.

Clearly L_r near epoch dependence on e_t implies L_r aproximability by it. Usually h_t^m is chosen to be $E(.|e_{t+m}, \dots, e_{t-m})$.

Definition 5.2.2. *For a given sequence of functions $f_n : T \times B \to R$, a metric ρ on B and a given sequence $\tau_n \in T$ the sequence of minimizers $\overline{\theta}_n$ of $f_n(\tau_n, \theta)$ is called identifiable unique, if for every $\epsilon > 0$*

$$\liminf_{n\to\infty} \left(\inf_{\{\theta \in B : \rho(\theta, \overline{\theta}_n) \geq \epsilon\}} (f_n(\tau_n, \theta) - f_n(\tau_n, \overline{\theta}_n)) \right) > 0$$

Definition 5.2.3. *Let $\overline{X} = (X_1, \dots, X_n)$. Let $Q(\overline{X}, \theta) = \sum_{t=1}^n q(X_t, \theta)$ be the conditional log-likelihood function, $\eta > 0$.*

1. $q_*(x, \theta, \eta) := \inf_{\rho(\theta,\theta') < \eta} q(x, \theta')$
2. $q^*(x, \theta, \eta) := \sup_{\rho(\theta,\theta') < \eta} q(x, \theta')$
3. $d(x) := \sup_{\theta \in \boldsymbol{\Theta}} |q(x, \theta)|$

4. Let H_t^X denote the law of X_t. Then we define $\overline{H}_n^X := \mathcal{L}(n^{-1}\sum_{t=1}^n X_t)$.

We now want to see when the conditional maximum likelihood estimator is consistent. For this purpose we will use the methodology used in [PP97]. We must establish:

- A local weak law of large numbers for the bracketing functions q_* and q^*.
- Strengthen this law to a uniform law of large numbers for q.
- Assure the identifiable uniqueness of the maximizing parameter θ_0.

To get the uniform law of large numbers the following assumption has to be satisfied:

Assumption 5.2.4. *1. $\mathbf{\Theta}$ is compact.*

2. For each $\theta \in \mathbf{\Theta}$ exists an $\eta(\theta)$ such that for each $0 < \eta \leq \eta(\theta)$ $q_*(x,\theta,\eta)$ and $q^*(x,\theta,\eta)$ are real valued and measurable and $q_*(X_t,\theta,\eta)$ and $q^*(X_t,\theta,\eta)$ satisfy a local law of large numbers.

Further we need one of the following assumptions:

Assumption 5.2.5. *For each $\theta \in \mathbf{\Theta}$ there exists an $\eta > 0$ such that $\rho(\theta,\theta') \leq \eta$ implies*

$$|q(X_t,\theta') - q(X_t,\theta)| \leq b_t(X_t)h(\rho(\theta,\theta'))$$

for all $t \in \mathbb{N}$, with $b_t : Z \to [0,\infty)$ being measurable and satisfying the condition $\sup_n n^{-1}\sum_{t=1}^n Eb_t(X_t) < \infty$ and $h : [0,\infty) \to [0,\infty)$ satisfies $h(y) \downarrow 0$ as $y \downarrow 0$.

Assumption 5.2.6. *1. The sequence $\overline{H}_n^X$ is tight on Ω.*

2. $q(x,\theta) = \sum_{k=1}^K r_k(x)s_k(x,\theta)$ where the r_k are measurable real functions for all $t \in N$ and $0 < k \leq K$. And the family $\{s_k(x,\theta)|t \in N\}$ are equicontinuous on $\Omega \times \mathbf{\Theta}$ for all $0 < k \leq K$ and real valued.

3. There exists a $\gamma > 0$ such that $\sup_n n^{-1}\sum_{t=1}^n E(d(X_t)^{1+\gamma}) < \infty$.

4. $\sup_n n^{-1}\sum_{t=1}^n E(|r_k(X_t)|) < \infty$ for all $0 < k \leq K$

To use Theorem 5.1 or 5.2 of [PP97] which establish uniform laws of large numbers for the estimator we have to verify 5.2.4 and either 5.2.5 or 5.2.6. To get consistency via this machinery we have also to establish identifiable uniqueness.

Assumption 5.2.7. *The maximizer θ^* of the conditional log likelihood function is identifiably unique.*

According to a remark in [PP97] this boils down to θ^* being unique in the case the function is lower semi-continuous and the parameter space is compact.
To follow this agenda we want first to establish a local law of large numbers to satisfy 5.2.4. First we will use the fact that under 7.3.3 X_t is near epoch dependent. If certain additional conditions are satisfied $q_*(X_t, \theta, \eta)$ and $q^*(X_t, \theta, \eta)$ will be also near epoch dependent(NED) on or at least L_0 approximable by an $\alpha-$mixing process. This will yield a local law of large numbers. The compactness of $\boldsymbol{\Theta}$ has to be assumed, it cannot be proven. So by argumentation and assumption we could fulfill 5.2.4. Then we have to check how we can fulfill 5.2.5 or 5.2.6.
In order to check these conditions we have to specify the likelihood function. So we have to choose the distributions of ϵ_t and D_t and the crash-probability-function f.
Generally it can be said that if we want to use 5.2.6 we can and have to assume that r_k is a constant function, because the term $\log(\ldots)$ doesn't factorize in a manner that we get a term without parameters.

5.2.1 Local law

We have seen that if 7.3.3 is satisfied, X_t is near epoch dependent on the $\alpha-$mixing process Y_t. Now we want to assure that q_* and q^* are also NED or at least L_0 approximable by Y_t. To do that we use some preservation theorems of [PP97]. The weakest condition given there is Theorem 6.13 To get a weak LLN we need

1. The compactness of $\boldsymbol{\Theta}$.
2. The equicontinuity of $\{q(x, \theta)\}$ on $\Omega \times \theta$.
3. The tightness of $\{\overline{H}_n^Z | n \in \mathbb{N}\}$ on Z where Z_t denotes the vector of all endogenous and exogenous variables.
4. The existence of a $\gamma > 0$ for which
$$\sup_n n^{-1} \sum_{t=1}^{n} E(d(X_t)^{1+\gamma}) < \infty$$

Alternatively we need:

1. The existence of measurable functions $B_t : \Omega \times \Omega \to [0, \infty)$ with
$$\limsup_{m\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} E(B_t(X_t, h_t^m(X_t))^{\epsilon}) < \infty$$
for some $\epsilon > 0$ and with
$$|q(x, \theta) - q(x', \theta)| \le B_t(x, x')|x - x'|$$

2. The equicontinuity of $\{q(x,\theta)\}$ on $\Omega \times \theta$.

3. The existence of a $\gamma > 0$ for which

$$\sup_n n^{-1} \sum_{t=1}^{n} E(d(X_t)^{1+\gamma}) < \infty$$

4. The measurability of d.

In the case where we have exogenous variables W_t we have to assume that the above assumptions hold for $Z_t := (X_t, W_t)$ instead of solely X_t. Unfortunately we cannot use the simplifications for such cases given in chapter 14 of [PP97], because the theorems there assume a normal QMLE estimator. That isn't our setup. Because we will need compactness of $\mathbf{\Theta}$ for the uniform laws of large numbers, we don't have to worry about it here. If we want to use 5.2.6 other conditions also are the same than the ones for uniform laws.
In order to check above conditions, we need to specify the likelihood function.

5.3 Application to our context

If we want to know if estimators in our context are consistent we have to consider two different cases. First the case we want to estimate a model really belonging to the model class and in the second case the estimation of real financial time-series. In the second case we can be quite sure that the observed time-series doesn't follow our model so we are in the case of misspecification. The theorems of [PP97] are formulated to work also in the latter case. To simplify matters in Theorem 7.1 of [PP97] the conditions for the different steps of the consistency proof are gathered and simplified modulo redundancy. But via σ_t the model in case of $\beta \neq 0$ depends on the infinite past. To make the theorems fit for this case we have to use the discussion on the end of chapter 6 of [PP97].
Because in the original form the presence of nuisance parameters is possible, but we don't need them here we can use the following simplified form. Because there are two alternative assumptions which can be used together with assumptions we need in any case, we state them first.

Assumption 5.3.1. *1. $\{q\}$is equicontinuous on $Z \times B$. which in our context just means q is continous on $Z \times B$.*

2. $\{\overline{H}_n^Z\}$ is is tight on Z

Remark 5.3.2. The second point is satisfied if Z is closed and

$$\sup_n n^{-1} \sum_{t=1}^{n} E|Z_t|^{\gamma'} < \infty$$

for some $\gamma' > 0$.

Assumption 5.3.3. *1. For each $\theta \in \theta$ there exists an $\eta > 0$ such that $\rho(\theta, \theta') \leq \eta$ implies*

$$|q(X_t, \theta') - q(X_t, \theta)| \leq b_t(X_t)h(\rho(\theta, \theta'))$$

for all $t \in N$, with $b_t : Z \to [0, \infty)$ being measurable and satisfying the condition $\sup_n n^{-1} \sum_{t=1}^n E(b_t(X_t)) < \infty$ and $h : [0, \infty) \to [0, \infty)$ satisfies $h(y) \downarrow 0$ as $y \downarrow 0$.

2. The existence of measurable functions $B_t : \Omega \times \Omega \to [0, \infty)$ with

$$\limsup_{m\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} E(B_t(X_t, h_t^m(X_t))^\epsilon) < \infty$$

for some $\epsilon > 0$ and with

$$|q(x, \theta) - q(x', \theta)| \leq B_t(x, x')|x - x'|$$

3. The equicontinuity of $\{q(x, \theta)\}$ on $\Omega \times \theta$.

4. The functions $q_*(x, \theta, \eta)$ and $q^*(x, \theta, \eta)$ are finite and Borel–measurable for any $\theta \in \mathbf{\Theta}$ and $\eta > 0$ small enough.

5. The functions $\sup_\theta |n^{-1} \sum_{t=1}^n q(Z_t, \theta) - n^{-1} \sum_{t=1}^n Eq(Z_t, \theta)|$ and $\sup_\theta |\nu_n(n^{-1} \sum_{t=1}^n q(Z_t, \theta)) - \nu_n(n^{-1} \sum_{t=1}^n Eq(Z_t, \theta)|$ are $\mathfrak{A}$ –measurable

Definition 5.3.4. *We define for $\omega = (Z_1, \ldots, Z_N)$*

$$R_n(\omega, \theta) := \frac{1}{n} \sum_{t=1}^{n} q(\omega, \theta)$$

and its non stochastic equivalent

$$R_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} Eq(\omega, \theta)$$

Theorem 5.3.5. *a) Let Z be the value space of $Z_t := (X_t, X_{t-1}, W_t)$, $\mathbf{\Theta}$ the parameter space. Z is a Borel subset of $\mathbb{R}^{dim(Z)}$ and $\mathbf{\Theta}$ is a compact metric space.*

b) $\{\nu_n | n \in \mathbb{N}\}$ is equicontinuous on $\mathbb{R}^{dim(q)} \times \mathbf{\Theta}$

c)

$$\sup_n n^{-1} \sum_{t=1}^{n} E(d(z)^{1+\gamma}) < \infty$$

for some $\gamma > 0$

d) The process Z_t is L_0–approximable by an α–mixing process

e) Either 5.3.1 or 5.3.3 holds.
Then $\sup_{\mathbf{\Theta}} |R_n(\omega,\theta) - \overline{R}_n(\theta)| \to 0$ *in probability as* $n \to \infty$ *and* $\{\overline{R}_n | n \in \mathbb{N}\}$ *is equicontinuous on* $\mathbf{\Theta}$. *If additionally an identifiably unique minimizers of* $\overline{R}_n$ *exists, any minimizing sequence is consistent.*

We use this form of Theorem 7.1 of [PP97]

Remark 5.3.6. In [PP97] the theorems are formulated such that time dependet summands q_t are allowed. In the the case where our models have just an ARCH-dynamic on σ_t or v_t respectively, q_t will be just q for all t. We made the notation easier for this purpose. Furthermore the equicontinuity conditions will then be just a continuity condition on q. q will also be constant over time if consider a genuine GARCH dynamic and work with the infinite past. But if we work with a start condition we will have use different q_ts for dimensional reasons.

Chapter 6

Asymptotic properties of different Quasi Maximum Likelihood Estimators in the pure ARCH case

6.1 Consistency of a Quasi–Maximum Likelihood Estimator in the CARCH-S model

Definition 6.1.1. *1. For a random variable J we define $\mu_J := EJ$.*

2. For a random variable J we define $\sigma_J^2 := Var(J)$

We consider the model discussed in section 2.1 with ARCH(1) specification of σ_t^2, i.e.:

$$X_t = \sigma_t \eta_t$$

with

$$\eta_t = (1 - B_t)\epsilon_t + B_t D_t \text{ and } \sigma_t^2 = \omega + \alpha X_{t-1}^2$$

where $\mathcal{L}(B_t|\mathfrak{F}_{t-1}) = \mathfrak{B}(1, p_t)$ and

$$p_t = f(a + bX_{t-1} + c\sigma_t^2) = f(X_{t-1}, \sigma_t) = g(X_{t-1})$$

As in section 3.3.1 we use the notation $h_t^2 := Var(\eta_t|\mathfrak{F}_{t-1})$. The specific parametrisation of f does not play a role in the proofs of this section and actually the results don't depend on it. We will use it later for the purpose of writing down the derivatives of l_t explicitly.
Pretending that the innovations η_t are normally distributed conditional on $\mathfrak{F}_{t-1}$ we maximize the conditional likelihood function and get the quasi maximum likelihood estimate. This estimator will usually be misspecified even if the data is generated by the right crash model, as we usually would not have that $\mathcal{L}(\eta_t|\mathfrak{F}_{t-1}) = N(E(\eta_t|\mathfrak{F}_{t-1}), h_t^2))$. Let

$$L_n = \sum_{t=1}^{n} l_t$$

with

$$\begin{aligned}
l_t : &= l_t(X_t, X_{t-1}, \theta) \\
&= \frac{1}{2}\log(\sqrt{2\pi}) + \frac{1}{2}(\log(\omega + \alpha X_{t-1}^2) + \log((1-p_t)\mu_{\epsilon^2} + p_t\mu_{D^2} - [(1-p_t)^2\mu_\epsilon^2 + p_t^2\mu_D^2])) \\
&\quad + \frac{(x - \sqrt{(\omega + \alpha X_{t-1}^2)}(\mu_\epsilon + p_t(\mu_D - \mu_\epsilon)))^2}{(\omega + \alpha X_{t-1}^2)((1-p_t)\mu_{\epsilon^2} + p_t\mu_{D^2} - [(1-p_t)\mu_\epsilon + p_t\mu_D]^2)}
\end{aligned}$$

The l_t can be written in the short form:

$$l_t = \frac{1}{2}\log(\sqrt{2\pi}) + \frac{1}{2}(\log(\sigma_t^2) + \log(h_t^2)) + \frac{(x - \sigma_t(\mu_\epsilon + p_t(\mu_D - \mu_\epsilon)))^2}{\sigma_t^2 h_t^2}$$

where $h_t^2 = (1-p_t)\mu_{\epsilon^2} + p_t\mu_{D^2} - [(1-p_t)\mu_\epsilon + p_t\mu_D]^2 = Var(\eta_t|\mathfrak{F}_{t-1})$.

We have to keep in mind that generally this estimator will use a misspecified model. So in fact we just minimize a functional. And even if the minimizer is consistent, that doesn't mean that the limit has any reasonable interpretation. Let $\theta \in \mathbf{\Theta}$ be the parameter vector of our model.
We assume in the following:

Assumption 6.1.2. *The parameter space $\mathbf{\Theta}$ is compact.*

Lemma 6.1.3. *For any $\gamma > 0$: If $E|X_t|^{2+2\gamma} < \infty$ then $E\sup_\Theta|\log(\sigma_t^2(\theta))|^{1+\gamma} < \infty$.*

Proof.

1. We see: $\log(\sigma_t^2) \leq \sigma_t^2 = \omega + \alpha X_{t-1}^2$ The compactness of the parameter space yields the existence of $\bar{\omega}$,$\bar{\alpha}$ such that

$$\log(\sigma_t^2(\theta)) \leq \bar{\omega} + \bar{\alpha}X_{t-1}^2$$

for all $\theta \in \mathbf{\Theta}$.

2. For all $\theta \in \mathbf{\Theta}$ and all X_{t-1} $\sigma_t^2(X_{t-1}, \theta) \geq \omega(\theta)$ holds. But $\omega > 0$ together with the compactness of $\mathbf{\Theta}$ implies the existence of an $\tilde{\omega}$ such that $\sigma_t^2(X_{t-1}, \theta) \geq \tilde{\omega}$ for all $\theta \in \mathbf{\Theta}$ and all X_{t-1}. So $\log(\sigma_t^2(X_{t-1}, \theta)) \geq \log(\tilde{\omega})$ for all $\theta \in \mathbf{\Theta}$ and all X_{t-1}

3. $$\log(\tilde{\omega}) \leq \log(\sigma_t^2(\theta)) \leq \log(\bar{\omega} + \bar{\alpha}|X_{t-1}|^2) \leq \bar{\omega} + \bar{\alpha}|X_{t-1}|^2$$

Therefore,
$$|\log(\sigma_t^2(\theta))| \leq |\log(\tilde{\omega})| + \bar{\omega} + \bar{\alpha}|X_{t-1}|^2$$
Hence, we have a constant k such that
$$|\log(\sigma_t^2(\theta))| \leq k + \bar{\alpha}|X_{t-1}|^2$$
This implies for any $\delta \geq 1$:
$$|\log(\sigma_t^2(\theta))|^\delta \leq 2^\delta(k^\delta + \bar{\alpha}^\delta|X_{t-1}|^{2\delta}) := d_1 + d_2|X_{t-1}|^{2\delta}$$
for positive constants d_1, d_2.

This finally yields:
$$E\sup_\Theta|\log(\sigma_t^2(\theta))|^\delta \leq d_1 + d_2E|X_{t-1}|^{2\delta} \tag{6.1}$$
So if $|X_{t-1}|^{2\delta} < \infty$ we have $E\sup_\Theta|\log(\sigma_t^2(\theta))|^\delta < \infty$ Taking $\delta = 1+\gamma$ yields the statement.

□

Corollary 6.1.4. *There exist constants k_1 and k_2 such that for all $\theta \in \mathbf{\Theta}$*
$$\sigma_t^2 \leq k_1 + k_2X_{t-1}^2 \tag{6.2}$$
Lemma 6.1.5. 1. *For all $\theta \in \mathbf{\Theta}$ and all X_{t-1}:*
$$(1-p_t)\mu_{\epsilon^2} + p_t\mu_{D^2} - [(1-p_t)\mu_\epsilon + p_t\mu_D]^2 \geq (1-p_t)\sigma_\epsilon^2 + p_t\sigma_D^2 \geq \min(\sigma_\epsilon^2, \sigma_D^2)$$
2. *For all $\theta \in \mathbf{\Theta}$ and all X_{t-1}.*
$$\max(\log(\sigma_\epsilon^2), \log(\sigma_D^2)) \geq \log((1-p_t)\sigma_\epsilon^2 + p_t\sigma_D^2) \geq \min(\log(\sigma_\epsilon^2), \log(\sigma_D^2))$$
3. *For all $\theta \in \mathbf{\Theta}$ and all X_{t-1}.*
$$(1-p_t)\mu_{\epsilon^2} + p_t\mu_{D^2} - [(1-p_t)\mu_\epsilon + p_t\mu_D]^2 \leq \mu_{\epsilon^2} + \mu_{D^2}$$
Proof. 1. Because of the convexity of $x \to x^2$ we have:
$$[(1-p_t)\mu_\epsilon + p_t\mu_D]^2 \leq (1-p_t)\mu_\epsilon^2 + p_t\mu_D^2$$
Therefore
$$\begin{aligned}&(1-p_t)\sigma_\epsilon^2 + p_t\sigma_D^2\\ &\quad= (1-p_t)\mu_{\epsilon^2} + p_t\mu_{D^2} - [(1-p_t)\mu_\epsilon^2 + p_t\mu_D^2]\\ &\quad\leq (1-p_t)\mu_{\epsilon^2} + p_t\mu_{D^2} - [(1-p_t)\mu_\epsilon + p_t\mu_D]^2\end{aligned}$$
holds. The last inequality is clear.

2. Let without loss of generality $\sigma_D \geq \sigma_\epsilon > 0$ Because p_t is between 0 and 1, $\sigma_\epsilon^2 \leq (1-p_t)\sigma_\epsilon^2 + p_t\sigma_D^2 \leq \sigma_D^2$. So $\log(\sigma\epsilon^2) \leq \log((1-p_t)\sigma_\epsilon^2 + p_t\sigma_D^2) \leq \log(\sigma_D^2)$. In fact
$$\log((1-p_t)\sigma_\epsilon^2 + p_t\sigma_D^2) \geq \min(\log(\sigma_\epsilon^2), \log(\sigma_D^2))$$
holds. This clearly does not depend on parameters concerning $p_t(\theta)$. Hence, this holds for all $\theta \in \mathbf{\Theta}$ and all X_{t-1}.

3. Clear because p_t is between 0 and 1 and μ_{ϵ^2},μ_{D^2} and $[(1-p_t)\mu_\epsilon + p_t\mu_D]^2$ are positive. .

□

Corollary 6.1.6. *Whenever ϵ_t and D_t are square integrable then there exist constants c_0, c_1, c_2, c_3 such that for all $\theta \in \mathbf{\Theta}$ and all X_{t-1}:*

1.
$$\infty > c_0 \geq (1-p_t)\mu_{\epsilon^2} + p_t\mu_{D^2} - [(1-p_t)\mu_\epsilon + p_t\mu_D]^2 \geq c_1 > 0$$

2.
$$\infty > c_2 \geq \log((1-p_t)\mu_{\epsilon^2} + p_t\mu_{D^2} - [(1-p_t)\mu_\epsilon + p_t\mu_D]^2) \geq c_3 > 0$$

Proof. When ϵ_t and D_t are square integrable all right hand sides of the inequalities in the previous lemma are constants greater than 0 and smaller than ∞. And so is the left hand side of point 2 in the previous lemma. □

Corollary 6.1.7.
$$E\sup_\theta |\log(h_t^2)|^{1+\gamma} < \infty$$

Proof. By the previous lemma there exist constants c_2, c_3 such that
$$\sup_\theta |\log(h_t^2)|^{1+\gamma} \leq (c_2 + c_3)^{1+\gamma}$$
so
$$E\sup_\theta |\log(h_t^2)|^{1+\gamma} \leq (c_2 + c_3)^{1+\gamma}$$

□

Remark 6.1.8. If we want to treat the moments of ϵ_t and D_t as nuisance parameters we get the inequalities by assuming compactness on their parameter spaces and $\sigma_\epsilon > 0$ and $\sigma_D > 0$. In practice we will have to choose them a priori anyway, because when trying to estimate them we firstly won't model crashes because these estimates lead to non extreme distributions and secondly, because of that fact we will run into flat areas of the parameter space. Nonetheless here and in the section concerning the alternative model we mention and treat them as estimable parameters when talking of moment conditions, because they don't make any problems in this sense. The only problem with regularity conditions is when we want to treat the means of ϵ_t and D_t as nuisance parameters, because they won't satisfy condition 11.3 b in [PP97]

Lemma 6.1.9. *Let $\gamma > 0$. Suppose that $E|X_t|^{2+2\gamma} < \infty$ then:*

$$E \sup_{\theta} \left(\frac{|X_t - \sigma_t(\mu_\epsilon + p_t(\mu_D - \mu_\epsilon))|}{\sigma_t h_t} \right)^{2+2\gamma} < \infty$$

Proof.

1. If we assume μ_ϵ and μ_D to be fixed or coming from a compact nuisance parameter space we get in either case a constant M such that

$$|\mu_\epsilon + p_t(\mu_D - \mu_\epsilon)| \leq M$$

the argument being that p_t is uniformly bounded to be between 0 and 1 no matter what its parameters are.

2. By corollary 6.1.4 we get have positive constants k_1 and k_2 such that for all $\theta \in \mathbf{\Theta}$ $\sigma_t^2 \leq k_1 + k_2|X_t|^2$. So $\sigma_t \leq \sqrt{(k_1)} + \sqrt{k_2}|X_t|$ because the constants are positive. Taking 1. into account we get positive constants M_1, M_2 such that:

$$\sigma_t|\mu_\epsilon + p_t(\mu_D - \mu_\epsilon)| < M_1 + M_2|X_{t-1}|$$

Therefore, we get

$$|X_t - \sigma_t(\mu_\epsilon + p_t(\mu_D - \mu_\epsilon))| \leq |X_t| + M_1 + M_2|X_{t-1}|.$$

Now we get for any $\delta > 1$:

$$|X_t - \sigma_t(\mu_\epsilon + p_t(\mu_D - \mu_\epsilon))|^\delta \leq 2^\delta|X_t|^\delta + 2^{2\delta}M_1^\delta + 2^{2\delta}M_2^\delta|X_{t-1}|^\delta$$

Which yields:

$$E \sup_{\theta} |X_t - \sigma_t(\mu_\epsilon + p_t(\mu_D - \mu_\epsilon))|^\delta \leq 2^\delta E|X_t|^\delta + 2^{2\delta}M_1^\delta + 2^{2\delta}M_2^\delta E|X_{t-1}|^\delta < \infty$$

3. We have seen that there exist constants e^2 and c^2 that bound σ_t^2 and h_t^2 uniformly from 0. So:

$$E\left(\sup_{\theta} \frac{|X_t - \sigma_t(\mu_\epsilon + p_t(\mu_D - \mu_\epsilon))|}{\sigma_t h_t}\right)^\delta \leq \frac{2^{\delta+1}E|X_t|^\delta + 2^{2\delta+1}M_1 + 2^{2\delta+1}M_2 E|X_{t-1}|^\delta}{e^\delta c^\delta} < \infty$$

taking $\delta = 2 + 2\gamma$ yields the result.

□

Corollary 6.1.10. *Let $\gamma > 0$. Suppose that $E|X_t|^{2+2\gamma} < \infty$ then:*

1.

$$\sup_n \frac{1}{n} \sum_{t=1}^{n} E \sup_{\theta} \left(\frac{|X_t - \sigma_t(\mu_\epsilon + p_t(\mu_D - \mu_\epsilon))|}{\sigma_t h_t} \right)^{2+2\gamma} < \infty$$

2.
$$\sup_n \frac{1}{n}\sum_{t=1}^{n} E\sup_{\Theta} |\log(\sigma_t^2(\theta))|^{1+\gamma} = E\sup_{\Theta} |\log(\sigma_t^2(\theta))|^{1+\gamma} < \infty$$

3.
$$\sup_n \frac{1}{n}\sum_{t=1}^{n} E\sup_{\Theta} |\log(h_t^2(\theta))|^{1+\gamma} = E\sup_{\Theta} |\log(h_t^2(\theta))|^{1+\gamma} < \infty$$

Proof. For (1) we use Lemma 6.1.9 and the following argumentation: Due to the stationarity of X_t,

$$E\sup_{\theta}\left(\frac{|X_t - \sigma_t(\mu_\epsilon + p_t(\mu_D - \mu_\epsilon))|}{\sigma_t h_t}\right)^{2+2\gamma}$$

does not depend on t and is finite say a constant H. But

$$\sup_n \frac{1}{n}\sum_{t=1}^{n} H = H < \infty$$

The two other points are obtained analogously from Lemmas 6.1.3 and 6.1.7. $\square$

Assumption 6.1.11. *1. $E|X_t|^{2+2\gamma} < \infty$ for a $\gamma > 0$.*

2. *g, f_ϵ and f_D are continuous.*
3. *$\omega > 0$*
4. *The parameter space $\mathbf{\Theta}$ is compact and doesn't contain 0 in the direction of the ω-coordinate.*
5. *One of the following conditions holds:*
   (a) *The support of f_ϵ and the support of f_D is $\mathbb{R}$.*
   (b) *The support of f_ϵ is $\mathbb{R}$ and $\forall t : p_t \leq 1 - \delta < 1$.*
   (c) *$supp(f_\epsilon) \cup supp(f_D) = \mathbb{R}$ and $\forall t : 0 < \gamma \leq p_t \leq 1 - \delta < 1$.*
6. *One of the following conditions holds:*
   (a) *$ED_t^2 \geq E\epsilon_t^2$ and $\alpha(E\epsilon_t^2 + \sup_y g(y)(ED_t^2 - E\epsilon_t^2)) < 1$.*
   (b) *$ED_t^2 \leq E\epsilon_t^2$ and $\alpha(E\epsilon_t^2 + \inf_y g(y)(ED_t^2 - E\epsilon_t^2)) < 1$.*
7. *There is an identifiably unique sequence of minimizer $\bar{\theta}_n$ of $\frac{1}{n}\sum_{t=1}^{n} El_t(\theta)$.*

Theorem 6.1.12. *If the Assumptions of 6.1.11 number 1–6 are satisfied then*

$$\sup_{\theta \in \boldsymbol{\Theta}} \left| \frac{1}{n} L_n(\theta) - \frac{1}{n} \sum_{t=1}^{n} El_t(\theta) \right| \to_p 0$$

as $n \to \infty$. And $\{\frac{1}{n}\sum_{t=1}^{n} El_t(\theta)\}$ is equicontinuous on $\boldsymbol{\Theta}$. If additionally Assumption 6.1.11 number 7 is satisfied then for any sequence $\hat{\theta}_n$ of minimizers of $\frac{1}{n}L_n(\theta)$

$$|\hat{\theta}_n - \bar{\theta}_n| \to_p 0$$

for $n \to \infty$. That means the estimator is consistent for $\bar{\theta}_n$.

Proof. We use Theorem 14.1 of [PP97].
Therefore, we write:

$$\frac{X_t - E(X_t|\mathfrak{F}_{t-1})}{\sqrt{Var(X_t|\mathfrak{F}_{t-1})}} = \frac{\eta_t - E(\eta_t|\mathfrak{F}_{t-1})}{h_t} =: F(X_t, X_{t-1})$$

with

$$\begin{aligned} F_t(x,y,\theta) &= F(x,y,\theta) \\ &= \frac{x - \sqrt{\omega + \alpha y^2}(\mu_\epsilon + g(y)(\mu_D - \mu_\epsilon))}{\sqrt{(\omega + \alpha y^2)((1-g(y))\mu_{\epsilon^2} + g(y)\mu_{D^2} - [(1-g(y))\mu_\epsilon + g(y)\mu_D]^2)]}} \end{aligned}$$

Then

$$\frac{\partial F}{\partial x}(x,y,\theta) = \frac{1}{\sqrt{((1-g(y))\mu_{\epsilon^2} + g(y)\mu_{D^2} - [(1-g(y))\mu_\epsilon + g(y)\mu_D]^2)(\omega + \alpha y^2)}}$$

and we can write

$$\begin{aligned} l_t(X_t, X_{t-1}, \theta) &= \log(2\pi) - \log\left(\frac{\partial F}{\partial x}(X_t, X_{t-1}, \theta)\right) + \frac{1}{2}F^2(X_t, X_{t-1}, \theta) \\ &= \log(2\pi) - \log\left(\left|\det \frac{\partial F}{\partial x}(X_t, X_{t-1}, \theta)\right|\right) + \frac{1}{2}F^2(X_t, X_{t-1}, \theta) \qquad (6.3) \end{aligned}$$

Now Lemmas 6.1.3 and 6.1.7 imply $\infty > \frac{1}{2}(E\sup_\theta|\log(\sigma_t^2)|^{1+\gamma} + E\sup_\theta|\log(h_t^2)|^{1+\gamma}) = E\sup_\theta|\log(|\det\frac{\partial F}{\partial x}(X_t, X_{t-1}, \theta)|)|$. Which by Corollary 6.1.10 (2) and (3)implies the same statements for the $\sup_n \frac{1}{n}\sum_{t=1}^{n} E\sup_\theta|\log(|\det\frac{\partial F}{\partial x}(X_t, X_{t-1}, \theta)|)| < \infty$. And Corollary 6.1.10 (1) explicitly states the moment condition on $F_t(X_t, X_{t-1}, \theta)$. So we verified Assumption 14.1. (c) of [PP97].
Surely $\frac{\partial F}{\partial x}(x,y,\theta)$ is continuous. Also $F_t(x,y,\theta) = F(x,y,\theta)$ is continuous and not depending on time. So the families $\{\frac{\partial F}{\partial x}(x,y,\theta) : t \in \mathbb{N}\}$ and $\{F(x,y,\theta) : t \in \mathbb{N}\}$ are equicontinuous verifying Assumption 14.1 (f) of [PP97]. By assumption $\boldsymbol{\Theta}$ is compact. By construction we have only one $s \in S$ so it is compact too. This

yields Assumption 14.1. (a) and (b) of [PP97]. We don't have to verify (e), because we use no external variables. Assumption 14.1 (d) boils down to

$$E|X_t|^\delta < \infty$$

for a $\delta > 0$ because we are working in a stationary context. But this is surely true because we just assumed that for $\delta = 2 + \gamma$.
Finally, in our case $\sup_t |F_t(x, y, \theta)| = F(x, y, \theta) < \infty$ for every tuple $(x, y, \theta) \in Im(X_t, X_{t-1}) \times \mathbf{\Theta}$ because F_t is time invariant. This yields 14.1(g) of [PP97].
Assumption 14.2 of [PP97] holds because we established $\{X_t\}$ to be mixing under our assumptions because they imply Assumption 4.4.7.
We assumed identifiable uniqueness and so we can apply Theorem 14.1 of [PP97].
□

6.2 Consistency results for the Conditional Maximum Likelihood estimator of the CARCH-S model

In the previous section we examined an estimator which will be just a quasi maximum likelihood estimator.
It seems to be reasonable to establish conditions when we will get consistent estimates using the real conditional likelihood under the assumption that the data are really generated by a mechanism belonging to our model class. The conditions we will get are hard to interpret, but will immediately lead to conditions comparable to the previous section, if we assume that both the crash and the non crash distributions are normal distributions.
We maximize the conditional likelihood function which is given by:

$$\prod_{t=1}^{n} p(X_t|X_{t-1})$$

We do this via maximizing the log-likelihood function

$$\sum_{t=1}^{n} \log(p(X_t|X_{t-1}))$$

which is in our case:

$$
\begin{aligned}
&\sum_{t=1}^{n} \log\left(p\left(X_t|X_{t-1}\right)\right) && (6.4)\\
&= \sum_{t=1}^{n} \log\left(\frac{1}{\left(\omega+\alpha X_{t-1}^2\right)^{\frac{1}{2}}}\left[\left(1-g\left(X_{t-1}\right)\right) f_\epsilon\left(\frac{X_t}{\left(\omega+\alpha X_{t-1}^2\right)^{\frac{1}{2}}}\right)\right.\right.\\
&\quad \left.\left.+\, g\left(X_{t-1}\right) f_D\left(\frac{X_t}{\left(\omega+\alpha X_{t-1}^2\right)^{\frac{1}{2}}}\right)\right]\right)\\
&= \sum_{t=1}^{n}\left(\log\left(\frac{1}{\left(\omega+\alpha X_{t-1}^2\right)^{\frac{1}{2}}}\right)\right.\\
&\quad \left.+\log\left[\left(1-g\left(X_{t-1}\right)\right) f_\epsilon\left(\frac{X_t}{\left(\omega+\alpha X_{t-1}^2\right)^{\frac{1}{2}}}\right)+g\left(X_{t-1}\right) f_D\left(\frac{X_t}{\left(\omega+\alpha X_{t-1}^2\right)^{\frac{1}{2}}}\right)\right]\right)
\end{aligned}
$$

To improve readability, we do not include the parameters of the crash probability function g explicitly in our notation. However, we have to recall that g depends also on parameters which constitute a part of the total parameter vector θ and which have to be estimated. We define:

$$
\begin{aligned}
q_t :=& \log\left(\frac{1}{\left(\omega+\alpha X_{t-1}^2\right)^{\frac{1}{2}}}\right) && (6.5)\\
&+\log\left[\left(1-g\left(X_{t-1}\right)\right) f_\epsilon\left(\frac{X_t}{\left(\omega+\alpha X_{t-1}^2\right)^{\frac{1}{2}}}\right)+g\left(X_{t-1}\right) f_D\left(\frac{X_t}{\left(\omega+\alpha X_{t-1}^2\right)^{\frac{1}{2}}}\right)\right]
\end{aligned}
$$

That means in short form:

$$
q_t = \frac{-1}{2}\log\left(\sigma_t^2\right)+\log\left(\left(1-p_t\right) f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_t f_D\left(\frac{X_t}{\sigma_t}\right)\right) \tag{6.6}
$$

To get the desired objective function we set:

$$
Q_n = \sum_{i=1}^{n} q_t \tag{6.7}
$$

We first want to find a set of general conditions to get consistency of the maximum likelihood estimator.

Assumption 6.2.1. *1. The assumptions of 4.4.7 hold and $\{X_t\}$ is a stationary version of the process.*

2. *The parameter space $\mathbf{\Theta}$ is compact.*

3. *g,f_ϵ and f_D are continuous.*

4. $$E(\sup_\theta |q_t(X_t, X_{t-1}, \theta)|)^{1+\gamma} := H < \infty$$

5. *There exists $\bar{\theta}_n$, an identifiable unique minimizer of $\frac{1}{n}\sum_{t=1}^n Eq_t(\theta)$.*

Remark 6.2.2. Point 3 of the preceeding assumption is redundant if we assume point 1, but we stated it, because if we could get a mixing version of X_t for other reasons than point 1, we would have assume 2 independently.

Remark 6.2.3. In our setting where q_t does not depend on t and so $\frac{1}{n}\sum_{t=1}^n Eq_t(\theta) = Eq_t(\theta)$ does not depend on t or n either and is furthermore continuous, the existence of an identifiable unique minimizer $\bar{\theta}_n$ of $\frac{1}{n}\sum_{t=1}^n Eq_t(\theta)$ is equivalent to the existence of a unique minimizer θ_0 of $Eq_t(\theta)$.

Theorem 6.2.4. *Assume Assumptions 6.2.1 1 to 4 hold, then*

$$\sup_{\theta\in\mathbf{\Theta}} \left| \frac{1}{n}Q_n(\theta) - \frac{1}{n}\sum_{t=1}^n Eq_t(\theta) \right| \to_p 0$$

as $n \to \infty$, and $\{\frac{1}{n}\sum_{t=1}^n Eq_t(\theta)\}$ is equicontinuous on θ. If additionally Assumption 6.2.1 5 is satisfied then for any sequence $\hat{\theta}_n$ of minimizers of $\frac{1}{n}Q_n(\theta)$

$$|\hat{\theta}_n - \bar{\theta}_n| \to_p 0$$

for $n \to \infty$, that means the estimator is consistent for $\bar{\theta}_n$.

Proof. Assuming compactness of $\mathbf{\Theta}$ Assumption 7.1.(a) of [PP97] is clear.
If

$$E(\sup_\theta |q_t(X_t, X_{t-1}, \theta)|)^{1+\gamma} := H < \infty$$

Then for all $n \in \mathbb{N}$:

$$\frac{1}{n}\sum_{t=1}^n E(\sup_\theta |q_t(X_t, X_{t-1}, \theta)|)^{1+\gamma} = H < \infty$$

So

$$\sup_n \frac{1}{n}\sum_{t=1}^n E(\sup_\theta |q_t(X_t, X_{t-1}, \theta)|)^{1+\gamma} = H < \infty$$

verifying Assumption 7.1.(c) of [PP97]. (d) is verified because (X_t, X_{t-1}) is mixing and hence L_0 approximable by itself (b) is naught in our context, because we

just work without transformation. (e) likewise as long as we work with fixed distributions of ϵ_t and D_t.
Now if g, f_ϵ and f_D are continuous q_t is continuous. Because $q_t = l$ for all t this means the family $\{q_t\}$ is equicontinuous. On the other hand $E|X_t| < \infty$ and all X_t have the same mean. So

$$\sup_n \frac{1}{n} \sum_{t=1}^{n} E|X_t| < \infty$$

Taking into account that the distributional assumptions assuring the geometric ergodicity of X_t imply that $Im(X_t) = \mathbb{R}$, we see that $Im((X_t, X_{t-1})) = \mathbb{R}^2$ which is closed in $\mathbb{R}^2$. So using Lemma C1 of [PP97] $\{H_t^n\}$ is tight satisfying Assumption 7.2 of [PP97]. Therefore, we can apply Theorem 7.1 of [PP97] yielding the result via the assumed identifiable uniqueness.
This shows the Consistency.

□

The following proposition reduces moment conditions on the weighted sum of the crash– and non–crash–distributions to conditions concerning these distributions itself.

Propositition 6.2.5. *Let $\gamma > 0$. If*

1. $$0 < M_1 \leq g(y) \leq M_2 < 1$$

2. $$E|X_t|^{2+\gamma} < \infty$$

3. $$E\left(\sup_{(\omega,\alpha)} \left|\log\left(f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right)\right|\right)^{1+\gamma} < \infty \tag{6.8}$$

4. $$E\left(\sup_{(\omega,\alpha)} \left|\log\left(f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right|\right)^{1+\gamma} < \infty \tag{6.9}$$

   Then

$$E(\sup_\theta |q_t(X_t, X_{t-1}, \theta)|)^{1+\gamma} < \infty$$

Proof. First we remark that if the claimed expectations exist they don't depend on t. This is due to the stationarity of X_t which implies the stationarity of $q_t(X_t, X_{t-1}, \theta)$.

$$|q_t(X_t, X_{t-1}, \theta)| = \frac{-1}{2}\log(\sigma_t^2) + \log\left((1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)\right)$$

First:

$$|\log(\sigma_t^2)| \leq c_1 + c_2 X_{t-1}^2$$

which is shown completely analogously to the proof of Lemma 6.1.3. And this holds for all $\theta \in \theta$. So

$$\sup_\theta |\log(\sigma_t^2)| \leq c_1 + c_2 X_{t-1}^2$$

for positive constants c_1, c_2. which yields

$$(\sup_\theta |\log(\sigma_t^2)|)^{1+\gamma} \leq c_4 + c_5 X_{t-1}^{2+2\gamma}$$

for positive constants c_4, c_5. Now we can conclude if $E|X_t|^{2+2\gamma} < \infty$ then $E(\sup_\theta |\log(\sigma_t^2)|)^{1+\gamma} < \infty$. So we have just to deal with

$$\log\left((1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)\right)$$

Now because f_ϵ and f_D are continuous density functions there exist constants M_ϵ and M_D. such that $f_\epsilon(x) \leq M_\epsilon$ and $f_D(x) \leq M_D$ for all $x \in \mathbb{R}$ and so we get:

$$\begin{aligned}
& \log\left((1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)\right) \\
\leq\ & (1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right) \\
\leq\ & (1-M_1)f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + M_2 f_D\left(\frac{X_t}{\sigma_t}\right) \\
\leq\ & (1-M_1)M_\epsilon + M_2 M_D =: K.
\end{aligned} \tag{6.10}$$

On the other hand:

$$\begin{aligned}
& \log\left((1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t(f_D\left(\frac{X_t}{\sigma_t}\right))\right) \\
\geq\ & (1-p_t)\log\left(f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right) + p_t \log\left(f_D\left(\frac{X_t}{\sigma_t}\right)\right) \\
\geq\ & (1-M_2)\log\left(f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right) + M_1 \log\left(f_D\left(\frac{X_t}{\sigma_t}\right)\right)
\end{aligned} \tag{6.11}$$

so we get:

$$\begin{aligned}
& \left|\log\left((1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right| \\
\leq\ & |K| + (1-M_2)\left|\log\left(f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right)\right| + M_1\left|\log\left(f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right|
\end{aligned} \tag{6.12}$$

Because this holds for all $\theta \in \mathbf{\Theta}$ and the constants don't depend on a particular θ we get:

$$\begin{aligned}
&\sup_\theta \left|\log\left((1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right| \qquad (6.13)\\
&\le \quad |K| + (1-M_2)| \left[\sup_\theta \left|\log\left(f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right)\right|\right] + M_1 \left[\sup_\theta \left|\log\left(f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right|\right]
\end{aligned}$$

and hence exist constants K_1, K_2, K_3, such that:

$$\begin{aligned}
&\left[\sup_\theta \left|\log\left((1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right|\right]^{1+\gamma} \qquad (6.14)\\
&\le \quad K_1 + K_2 \left[\sup_\theta \left|\log\left(f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right)\right|\right]^{1+\gamma} + K_3 \sup_\theta \left[\sup_\theta \left|\log\left(f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right|\right]^{1+\gamma}
\end{aligned}$$

Therefore, if $E\left(\sup_\theta \left|\log(f_D\left(\frac{X_t}{\sigma_t}\right)\right|\right)^{1+\gamma} < \infty$ and $E\left(\sup_{(\theta)} \left|\log(f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right|\right)^{1+\gamma} < \infty$ then

$$E\left(\sup_\theta \left|\log\left((1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right|\right)^{1+\gamma} < \infty$$

Obviously, for $* = \epsilon$, D,

$$\sup_\theta \left|\log\left(f_*\left(\frac{X_t}{\sigma_t}\right)\right)\right| = \sup_{(\omega,\alpha)} \left|\log\left(f_*\left(\frac{X_t}{\sigma_t}\right)\right)\right|$$

because the parameters concerning p_t are not part of the term.
Using $\sigma_t^2 \ge \omega \ge \tilde{\omega} > 0$ for all $\theta \in \mathbf{\Theta}$ we get together with the first statement of the proof constants C_1 and C_2 such that:

$$\begin{aligned}
&E(\sup_\theta |q_t(X_t, X_{t-1}, \theta)|)^{1+\gamma} \qquad (6.15)\\
&\le \quad C_1 E(\sup_\theta |\log(\sigma_t^2)|)^{1+\gamma}\\
&\quad +C_2 E\left(\sup_\theta \left|\log\left((1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right|\right)^{1+\gamma} =: H < \infty
\end{aligned}$$

□

Corollary 6.2.6. *If Assumptions 6.2.1 number 1 to 3 and 5 hold and additionally the assumptions of Proposition 6.2.5, then the conditional maximum likelihood estimator is consistent.*

Remark 6.2.7. The inequality in 6.11 actually permits only densities which are not vanishing on the whole real line. But we can get control of these densities in other cases, too. To illustrate this we consider f_ϵ to be greater than zero on the whole real line and f_D to be greater than zero on the negative real line and zero elsewhere. Then we can calculate:

$$\begin{aligned}
&\log\left((1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_t(f_D\left(\frac{X_t}{\sigma_t}\right))\right)\\
=\;& \mathbb{I}_{\{x\geq 0\}}(X_t)\log\left((1-p_t)f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right)\\
& \mathbb{I}_{\{x<0\}}(X_t)\log\left((1-p_t)f_\epsilon\left(\frac{-|X_t|}{\sigma_t}\right)+p_t(f_D\left(\frac{X_t}{\sigma_t}\right))\right)\\
\geq\;& (1-p_t)\log\left(f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right)+\mathbb{I}_{\{x<0\}}p_t\log\left(f_D\left(\frac{-|X_t|}{\sigma_t}\right)\right)\\
=\;& (1-p_t)\log\left(f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right)+\mathbb{I}_{\{x<0\}}p_t\log\left(f_D\left(\frac{-|X_t|}{\sigma_t}\right)\right)=(*)
\end{aligned}$$

Except for the null set $\{0\}$ then

$$\begin{aligned}
&(*)\\
\geq\;& (1-M_2)\log\left(f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right)+\mathbb{I}_{\{x<0\}}(X_t)M_1\log\left(f_D\left(\frac{-|X_t|}{\sigma_t}\right)\right)
\end{aligned}$$

So the rest of the proof of Proposition 6.2.5 will go through if we replace

$$E\left(\sup_{(\omega,\alpha)}\left|\log\left(f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right|\right)^{1+\gamma}<\infty \tag{6.16}$$

by

$$E\left(\sup_{(\omega,\alpha)}\left|\log\left(f_D\left(\frac{-|X_t|}{\sigma_t}\right)\right)\right|\right)^{1+\gamma}<\infty \tag{6.17}$$

While sticking to the rest of the original conditions we get the result of Proposition 6.2.5 and finally consistency.

Corollary 6.2.8. *If ϵ_t and D_t are both normally distributed, p_t bounded away from 0 and 1 and $E|X_t|^{2+2\gamma}<\infty$ for a $\gamma>0$. then*

$$E(\sup_\theta|q_t(X_t,X_{t-1},\theta)|)^{1+\gamma}<\infty$$

Proof. To show that $E\sup_{(\omega,\alpha)}|\log(f_*(\frac{X_t}{\sigma_t})|^{1+\gamma} < \infty$ for $* = D, \epsilon$ we can use similar argumentation as in the proof of Lemma 6.1.9:

$$\left| -\frac{1}{2}\log(2\pi\sigma_\epsilon^2) \ - \frac{(X_t-\mu_\epsilon\sigma_t)^2}{2\sigma_\epsilon^2\sigma_t^2} \right| \leq \frac{c_1+c_2|X_t|^2}{2\sigma_\epsilon^2\tilde{\omega}}$$

This again holds for all $\theta \in \mathbf{\Theta}$ and so

$$\sup_\theta \left| f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right| \leq \frac{c_1+c_2|X_t|^2}{2\sigma_\epsilon^2\tilde{\omega}}$$

$$\Rightarrow \left(\sup_\theta \left| f_\epsilon\left(\frac{X_t}{\sigma_t}\right)\right|\right)^{1+\gamma} \leq c_3+c_4|X_t|^{2+2\gamma}$$

for positive constants $c_1, c_2, c_3\ c_4$. So if $E|X_t|^{2+2\gamma}$ is finite so is $E(\sup_\theta|f_\epsilon(\frac{X_t}{\sigma_t})|)^{1+\gamma}$. The same argument holds for f_D too. Therefore Proposition 6.2.5 yields the result. □

So we get using Theorem 6.2.4 the following result, which we want to state as a theorem.

Theorem 6.2.9. *If*

1. *the parameter space* $\mathbf{\Theta}$ *is compact and the assumptions of 4.4.7 hold,*
2. ϵ_t *and* D_t *are normally distributed,*
3. $0 < M_1 \leq g(y) \leq M_2 < 1$,
4. $E|X_t|^{2+\gamma} < \infty$,
5. *the true parameter* θ_0 *is identifiable unique,*

then the conditional maximum likelihood estimator is consistent.

Propositition 6.2.10. *If* ϵ_t *is normally distributed and* D_t *lognormally distributed with parameters (0,1),* p_t *bounded away from away* 0 *and* 1 *and*

1. $E|X_t|^{2+2\gamma} < \infty$ *for a* $\gamma > 0$.
2. $E|\log(|X_t|)|^{2+2\gamma} < \infty$ *for a* $\gamma > 0$.

then

$$E(\sup_\theta|q_t(X_t, X_{t-1}, \theta)|)^{1+\gamma} < \infty$$

and so under these conditions and assuming

1. *the parameter space* $\mathbf{\Theta}$ *is compact and the assumptions of 4.4.7 hold,*
2. $0 < M_1 \leq g(y) \leq M_2 < 1$,

3. the true parameter θ_0 is identifiable unique,

we have consistency of the maximum likelihood estimator.

Proof. Having done the calculations for the normal distribution we have to check, using Remark 6.2.7 that

$$E\left(\sup\left|-\log\left(\sqrt{2\pi}\left|\frac{X_t}{\sigma_t}\right|\right)-\frac{1}{2}(\log(|X_t|)-\log(\sigma_t))^2\right|\right)^{1+\gamma}<\infty$$

Now

$$\begin{aligned}&-\log(\sqrt{2\pi})-(\log(|X_t|)-\log(\sigma_t))-\frac{1}{2}(\log(|X_t|)-\log(\sigma_t))^2\\ \leq\ & const.+|\log(|X_t|)|+|\log(\sigma_t)|+const.[\log(|X_t|)^2+\log(\sigma_t)^2]\end{aligned}$$

Now $\log(\sigma_t)=\frac{1}{2}\log(\sigma_t^2)$, $\log(\sigma_t^2)\geq\omega$ and

$$\log(\sigma_t^2)=\log(\omega+\alpha X_{t-1}^2)=\log(\omega)+\log\left(1+\frac{\alpha}{\omega}X_{t-1}^2\right)\leq\log(\omega)+\frac{\alpha}{\omega}X_{t-1}^2$$

Using the compactness of $\boldsymbol{\Theta}$ and, in particular, $\omega\geq\tilde{\omega}>0$ we get:

$$\begin{aligned}&\sup_{(\omega,\alpha)}\left|-\log\left(\sqrt{2\pi}\left|\frac{X_t}{\sigma_t}\right|\right)-\frac{1}{2}(\log(|X_t|)-\log(\sigma_t))^2\right|\\ \leq\ & const.+const.\log(|X_t|)+\log(X_t^2)+const.X_{t-1}^2\end{aligned}$$

But the expectation of the $(1+\gamma)$th moment of the right hand side is finite by our assumptions.

□

6.3 Consistency of a Quasi–Maximum Likelihood Estimator in the ACARCH-V model

Now we consider the following model:

$$X_t=(1-p_t)\mu+p_t\delta+\sigma_t[(1-B_t)\epsilon_t+B_tD_t] \tag{6.18}$$

with where $\mathcal{L}(B_t|\mathfrak{F}_{t-1})=\mathfrak{B}(1,p_t)$,

$$p_t=f(X_{t-1},v_t)=g(X_{t-1})$$

and ϵ_t and D_t both have mean 0. We use like in section 3.4 the notation $\mu_t=(1-p_t)\mu+p_t\delta=\mu+p_t\Delta$, $\eta_t=(1-B_t)\epsilon_t+B_tD_t$ $v_t=Var(X_t|\mathfrak{F}_{t-1})$ and $h_t^2=Var(\eta_t|\mathfrak{F}_{t-1})$.

Further we assume that v_t follows an ARCH(1)–dynamic:

$$\omega + \alpha X_{t-1}^2$$

The specific parametrisation of f does not play a role in the proofs of this section and actually the results don't depend on it. We will use it later for the purpose of writing down the derivatives of l_t explicitly. To avoid black box conditions we may choose an appropriate quasi likelihood function. We do so by modeling $\{\eta_t\}$ to be conditionally normal i.e . $\mathcal{L}(\frac{\eta_t}{h_t}|\mathfrak{F}_{t-1}) = N(0,1)$. i.e. conditional on the past:

$$\mathcal{L}(\eta_t|\mathfrak{F}_{t-1}) = N(0, h_t^2) = N(0, (1-p_t)\sigma_\epsilon^2 + p_t\sigma_D^2) \tag{6.19}$$

Then the objective function to be minimized is

$$L_n = \sum_{t=1}^{n} l_t$$

with

$$l_t = \frac{1}{2}\log(\sqrt{2\pi}) + \frac{1}{2}\log(v_t^2) + \frac{(X_t - \mu - p_t\Delta)^2}{2v_t^2}$$

$$= \frac{1}{2}\log(\sqrt{2\pi}) + \frac{1}{2}\log(\omega + \alpha X_{t-1}^2) + \frac{(X_t - \mu - g(X_{t-1})\Delta)^2}{2(\omega + \alpha X_{t-1}^2)} \tag{6.20}$$

Lemma 6.3.1. *For any $\gamma > 0$, if $E|X_t|^{2+2\gamma} < \infty$ then $E\sup_\Theta |\log(v_t^2(\theta))|^{1+\gamma} < \infty$.*

Proof. Completely like the proof of Lemma 6.1.3. □

Corollary 6.3.2. *If $\omega > 0$ and $\{X_t\}$ is square integrable, then $E|\log(v_t^2)| < \infty$*

Lemma 6.3.3. *For any $\gamma > 0$, if $E|X_t|^{2+2\gamma} < \infty$ then*

$$E\sup_{\mathbf{\Theta}} \left|\frac{(X_t - \mu + p_t\Delta)}{v_t}\right|^{2+2\gamma} < \infty$$

Proof.

1. For all $\theta \in \mathbf{\Theta}$ and all X_{t-1} $v_t^2(X_{t-1},\theta) \geq \omega(\theta)$ holds. But $\omega > 0$ together with the compactness of $\mathbf{\Theta}$ implies the existence of an ω_0 such that $v_t^2(X_{t-1},\theta) \geq \omega_0$ for all $\theta \in \mathbf{\Theta}$ and all X_{t-1}. We define $\epsilon^2 := \omega_0$, then for any $\delta > 0$

$$(v_t^2)^{-\delta} \geq \epsilon^{-2\delta}$$

holds.

2. Because p_t is between 0 and 1 $\|\mu + p_t\Delta| \leq |\mu| + |\delta|$ and since we assumed the parameter space to be compact, there exists a constant $c \geq 0$ such that. $|\mu| + |\delta| \leq c$. Hence $|X_t - \mu + p_t\Delta| \leq |X_t| + |\mu + p_t\Delta| \leq |X_t| + c$. It follows for any $\delta \geq 1$ $|X_t - \mu + p_t\Delta|^\delta \leq (|X_t| + c)^\delta = (c_1 + c_2|X_t|^\delta)$.

   for constants c_1 and c_2 which we get like in the proof of Lemma 6.3.1.

$$E(\sup_{\mathbf{\Theta}} \frac{(X_t - \mu + p_t\Delta)^2}{v_t})^\delta \leq \frac{1}{\epsilon^\delta}(c_1 + c_2 E(|X_t|^\delta)) < \infty \tag{6.21}$$

   We use this argument for $\delta = 2 + 2\gamma$ and get the desired result.

Corollary 6.3.4. *Let $\gamma > 0$. Suppose that $E|X_t|^{2+2\gamma} < \infty$, then:*

1.
$$\sup_n \frac{1}{n} \sum_{t=1}^{n} E \sup_{\mathbf{\Theta}} \left| \frac{X_t - \mu + p_t\Delta}{v_t} \right|^{2+2\gamma}$$

2.
$$\sup_n \frac{1}{n} \sum_{t=1}^{n} E \sup_{\Theta} |\log(v_t^2(\theta))|^{1+\gamma} < \infty$$

Proof. Due to the stationarity of X_t the weighted sum reproduces

$$E \sup_{\mathbf{\Theta}} \left| \frac{(X_t - \mu + p_t\Delta)^2}{2v_t^2} \right|^{1+\gamma}.$$

Sinnce the supremum of a constant is a constant, Lemma 6.3.3 yields the desired result. Analogously we get the second point by using Lemma 6.3.1. □

If we assume to have a stationary and ergodic version of $\{X_t\}$ which exists by the geometric ergodicity we can apply a local law of large numbers assuming $\{X_t\}$ is square integrable. This is the case if $\alpha < 1$ and $\{\epsilon_t\}$ and $\{D_t\}$ are square integrable. □

Assumption 6.3.5. *1. $E|X_t|^{2+2\gamma} < \infty$ for a $\gamma > 0$.*

2. g,f_ϵ and f_D are continuous.

3. $\alpha < 1$, $\omega > 0$.

4. The parameter space $\mathbf{\Theta}$ is compact and doesn't contain 0 in the ω-coordinate direction.

5. One of the following conditions holds:

   (a) The support of f_ϵ and the support of f_D is $\mathbb{R}$.

   (b) The support of f_ϵ is $\mathbb{R}$ and $\forall t : p_t \leq 1 - \delta < 1$.

(c) $supp(f_\epsilon) \cup supp(f_D) = \mathbb{R}$ *and* $\forall t: 0 < \gamma \leq p_t \leq 1 - \delta < 1$.

6. *There is an identifiably unique sequence of minimizer* $\bar{\theta}_n$ *of* $\frac{1}{n}\sum_{t=1}^{n} El_t(\theta)$.

Theorem 6.3.6. *If the Assumptions of 6.3.5 1–5 are satisfied then*

$$\sup_{\theta\in\mathbf{\Theta}}\left|\frac{1}{n}L_n(\theta) - \frac{1}{n}\sum_{t=1}^{n} El_t(\theta)\right| \to_p 0$$

as $n \to \infty$. *Furthermore,* $\{\frac{1}{n}\sum_{t=1}^{n} El_t(\theta)\}$ *is equicontinuous on* θ. *If additionally Assumption 6.3.5 6 is satisfied then for any sequence* $\hat{\theta}_n$ *of minimizers of* $\frac{1}{n}L_n(\theta)$

$$|\hat{\theta}_n - \bar{\theta}_n| \to_p 0$$

for $n \to \infty$. *That means the estimator is consistent for* $\bar{\theta}_n$.

Proof. We use Theorem 14.1 of [PP97].
Therefore we write $\frac{\eta_t}{h_t}$ again as $F(X_t, X_{t-1}, \theta)$ with

$$F(x, y, \theta) = \frac{x - \mu - g(y)\Delta}{\sqrt{\omega + \alpha y^2}}$$

Then

$$\frac{\partial F}{\partial x}(x, y, \theta) = \frac{1}{\sqrt{\omega + \alpha y^2}}$$

and we can write

$$l_t(X_t, X_{t-1}, \theta) = \log(2\pi) - \log\left(\frac{\partial F}{\partial x}(X_t, X_{t-1}, \theta)\right) + \frac{1}{2}F^2(X_t, X_{t-1}, \theta)$$

$$= \log(2\pi) - |\log\left(\left|\det\frac{\partial}{F}\partial x(X_t, X_{t-1}, \theta)\right)\right| + \frac{1}{2}F^2(X_t, X_{t-1}, \theta) \qquad (6.22)$$

Now Corollary 6.3.4 (2) implies $\infty > \frac{1}{2}\sup_n \frac{1}{n}\sum_{t=1}^{n} E\sup_\theta |\log(v_t^2)|^{1+\gamma}$
$= E\sup_\theta|\log(|\det\frac{\partial F}{\partial x}(X_t, X_{t-1}, \theta)|$.
And Corollary 6.3.4 (1) explicitly states the moment condition on $F(X_t, X_{t-1}, \theta)$.
So we verified Assumption 14.1. (c) of [PP97].
Obviously $\frac{\partial F}{\partial x}(x, y, \theta)$ is continuous and does not depend on t. Also $F(x, y, \theta)$ is continuous and not depending on time. So the families $\{\frac{\partial F}{\partial x}(x, y, \theta) : t \in \mathbb{N}\}$ and $\{F(x, y, \theta) : t \in \mathbb{N}\}$ are equicontinuous verifying Assumption 14.1 (f) of [PP97].
By assumption $\mathbf{\Theta}$ is compact and by construction we have $S = \{s\}$, so it is compact too. This yields Assumption 14.1. (a) and (b) of [PP97]. We don't have to verify (e), because we use no external variables. Assumption 14.1 (d) is boils down to

$$E|X_t|^\delta < \infty$$

for a $\delta > 0$ because we are working in a stationary context. But this is surely true because we just assumed that for $\delta = 2 + \gamma$.
Finally, in our case $\sup_t |F_t(x, y, \theta)| = F(x, y, \theta) < \infty$ for every tuple $(x, y, \theta) \in Im(X_t, X_{t-1}) \times \mathbf{\Theta}$ because F is time invariant. This yields 14.1(g) of [PP97]. Our assumptions imply Assumption 4.5.9. So we can assume that $\{X_t\}$ is a mixing version of our process. This yields Assumption 14.2 of [PP97]. We assumed identifiable uniqueness and so we can apply Theorem 14.1 of [PP97].

□

6.4 Consistency results for the Conditional Maximum Likelihood estimator of the ACARCH-V model

We consider the same model as in section 6.3 and we want to explore the likelihood function given by

$$Q_n := \sum_{t=1}^{n} \log(q_t(X_t|X_{t-1}))$$

with

$$\begin{aligned} & q_t(X_t|X_{t-1}) \\ = \; & (1 - g(X_{t-1}))\frac{h(X_{t-1})}{v(X_{t-1})} f_\epsilon \left(\frac{X_t - \mu(X_{t-1})}{v(X_{t-1})} h(X_{t-1}) \right) \\ & + g(X_{t-1}) \frac{h(X_{t-1})}{v(X_{t-1})} f_D \left(\frac{X_t - \mu(X_{t-1})}{v(X_{t-1})} h(X_{t-1}) \right). \end{aligned} \tag{6.23}$$

We remark that $q_t(x, y)$ does not depend on t.
In analogy to section 6.2 we get a first set of conditions for consistency of the estimator, defined as the minimizer of Q_n.

Assumption 6.4.1. *1. The assumptions of 4.4.7 hold and $\{X_t\}$ is a stationary version of the process.*

2. The parameter space $\mathbf{\Theta}$ is compact.

3. g,f_ϵ and f_D are continuous.

4.
$$E(\sup_\theta |q_t(X_t, X_{t-1}, \theta)|)^{1+\gamma} := H < \infty$$

5. The true parameter θ_0 is identifiable unique.

Remark 6.4.2. Assuming point 1 point 3 is redundant. Again we stated 3 to make clear what kind of conditions arise even if we had mixing or L_0–approximability established via a different way.

Theorem 6.4.3. *Assume Assumptions 6.4.1 number 1 to 4 hold, then*

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}Q_n(\theta)-\frac{1}{n}\sum_{t=1}^{n}Eq_t(\theta)\right|\to_p 0$$

as $n\to\infty$, *and* $\{\frac{1}{n}\sum_{t=1}^{n}Eq_t(\theta)\}$ *is equicontinuous on* θ. *If additionally Assumption 6.4.1 number 5 is satisfied then for any sequence* $\hat{\theta_n}$ *of minimizers of* $\frac{1}{n}Q_n(\theta)$

$$|\hat{\theta_n}-\bar{\theta_n}|\to_p 0$$

for $n\to\infty$. *that means the estimator is consistent for* $\bar{\theta_n}$ *.*

Proof. Due to the general nature of the statement the proof is actually word by word the same as that of Theorem 6.2.4. ☐

Again we state a Proposition, which gets rid of the sum in the logarithm in order to eventually get more explicit assumptions.

Propositition 6.4.4. *Let* $\gamma>0$. *If*

1. $$0<M_1\le g(y)\le M_2<1$$

2. $$E|X_t|^{2+\gamma}<\infty$$

3. $$E\left(\sup_{\theta}\left|\log\left(f_\epsilon\left(\frac{(X_t-\mu_t)h_t}{v_t}\right)\right)\right|\right)^{1+\gamma}<\infty \tag{6.24}$$

4. $$E\left(\sup_{\theta}\left|\log\left(f_D\left(\frac{(X_t-\mu_t)h_t}{v_t}\right)\right)\right|\right)^{1+\gamma}<\infty \tag{6.25}$$

   Then

$$E(\sup_{\theta}|\log(q_t(X_t,X_{t-1},\theta))|)^{1+\gamma}<\infty$$

Proof. We can use the analogous argumentation of the proof of Proposition 6.2.5 and have to show additionally $E(\sup_\theta|(\log(h_t))|)^{1+\gamma}<\infty$.

Now

$$
\begin{aligned}
&(1-M_2)\log(\sigma_\epsilon^2)+M_1\log(\sigma_D^2)\\
&\quad\leq\ (1-p_t)\log(\sigma_\epsilon^2)+p_t\log(\sigma_D^2)\\
&\quad\leq\ \log(h_t^2)\leq(1-p_t)\sigma_\epsilon^2+p_t\sigma_D^2\\
&\quad\leq\ (1-M_1)\sigma_\epsilon^2+M_2\sigma_D^2
\end{aligned}
$$

Hence,

$$
|\log(h_t)|=\frac{1}{2}|\log(h_t^2)|\leq C
$$

for a constant C. Therefore,

$$
\sup_\theta|\log(h_t)|^{1+\gamma}\leq C^{1+\gamma}
$$

$E(\sup_\theta|(\log(h_t))|)^{1+\gamma}<\infty$ follows because this constant does not depend on θ.

□

Remark 6.4.5. The stationarity of X_t implies that if

$$
E(\sup_\theta|\log(q_t(X_t,X_{t-1},\theta))|)^{1+\gamma}<\infty
$$

then also

$$
\sup_n\frac{1}{n}\sum_{t=1}^{n}E(\sup_\theta|\log(q_t(X_t,X_{t-1},\theta))|)^{1+\gamma}<\infty
$$

Choosing appropriate distributions for ϵ_t and D_t we get an explicit moment condition to get consistency.

Corollary 6.4.6. *Assume Assumption 6.4.1 1–3 and 5 hold. If ϵ_t and D_t are both normally distributed and $E|X_t|^{2+2\gamma}<\infty$ for a $\gamma>0$, then the maximum-likelihood estimator is consistent.*

Proof. Because p_t is between 0 and 1, $|\mu+p_t\Delta|\leq|\mu|+|\delta|$. Because we assumed the parameter space to be compact, there exists a constant $c\geq 0$ such that $|\mu|+|\delta|\leq c$. Therefore $|X_t-\mu+p_t\Delta|\leq|X_t|+|\mu+p_t\Delta|\leq|X_t|+c$ holds and we can conclude $|X_t-\mu+p_t\Delta|^2\leq(|X_t|+c)^2=(c_1+c_2|X_t|^2)$ for positive constants c_1, c_2. On the other hand $|h_t^2|\leq\max(\sigma_\epsilon^2,\sigma_D^2)$. So $|h_t^2|\leq c_3$ for a positive constant c_3. It follows

$$
|(X_t-\mu+p_t\Delta)^2h_t^2|\leq c_1c_3+c_2c_3|X_t|^2.
$$

With the notation used before we have $|v_t^2|\geq\tilde{\omega}>0$ for all $\theta\in\boldsymbol{\Theta}$.
So

$$
\left|-\frac{1}{2}\log(2\pi\sigma_\epsilon^2)-\frac{(X_t-\mu+p_t\Delta)^2h_t^2}{2\sigma_\epsilon^2v_t^2}\right|\leq c_4\frac{c_1c_3+c_2c_3|X_t|^2}{2\sigma_\epsilon^2\tilde{\omega}}
$$

for a positive constant c_4. Argumentation like e.g. that in the proof of Corollary 6.2.8 yields

$$E\left(\sup_{(\theta)}\left|\log\left(f_\epsilon\left(\frac{(X_t-\mu_t)h_t}{v_t}\right)\right)\right|\right)^{1+\gamma} < \infty.$$

We use the same argumentation to show

$$E\left(\sup_{(\theta)}\left|\log(f_D(\frac{(X_t-\mu_t)h_t}{v_t})\right|\right)^{1+\gamma} < \infty$$

and we can then use Proposition 6.4.4 to conclude with Theorem 6.4.3. $\square$

6.5 Some generic results concerning Asymptotic Normality

We want to state a generic result when

$$\frac{1}{\sqrt{n}}\frac{\partial L_n}{\partial\theta}|_{\theta_0} \to^{\mathcal{D}} N(0,B)$$

for a positive definite matrix B. The following Lemma is formulated for the first introduced quasi maximum likelihood estimators. For the estimators given by functions still including f_ϵ and f_D a corresponding result holds assuming the continuous differentiability of these functions.

Lemma 6.5.1. *If*

1. *g is two times continuously differentiable.*

2. $$E\left|\frac{\partial L_n}{\partial\theta_i}|_{\theta_0}\right|^{2+\delta} < \infty$$

   for a $\delta > 0$ and all parameters θ_i.

Then

$$\frac{1}{\sqrt{n}}\frac{\partial L_n}{\partial\theta}|_{\theta_0} \to^{\mathcal{D}} N(0,B)$$

for a positive definite matrix B.

Proof.

1. The first condition is just needed to ensure the very existence of the partial derivatives.

2. The moment condition implies the existence of $E|\frac{\partial L_n}{\partial\theta_i}|_{\theta_0}|$ So we can apply the ergodic theorem for stationary processes and have

$$E\frac{\partial L_n}{\partial\theta_i}|_{\theta_0} = \frac{\partial}{\partial\theta_i}E(L_n)|_{\theta_0} = 0$$

3. We have shown that under the assumptions $\{X_t\}$ is strongly mixing with geometric coefficients. $\frac{\partial L_n}{\partial \theta}|_{\theta_0}$ is a function of X_n and X_{n-1}. So is for every $\lambda \in R^{length(\theta)}$:
$$\lambda' \frac{\partial L_n}{\partial \theta}|_{\theta_0} = \sum \lambda_i \frac{\partial L_n}{\partial \theta_i}$$
So it is also mixing with geometric coefficients, i.e. we have $a > 0$, $\rho < 1$ such that
$$\alpha(n) \leq c\rho^n \Rightarrow \alpha(n)^{\frac{2+\delta}{\delta}} \leq c^{\frac{2+\delta}{\delta}} (\rho^{\frac{2+\delta}{\delta}})^n \leq c_1 n^{-\frac{2+\delta}{\delta}+b}$$
for any positive constants c_1, b if n is large enough.

4. Define $Y_t(\lambda) = \sum \lambda_i \frac{\partial L_n}{\partial \theta_i}$.
Since
$$\left|\sum \lambda_i \frac{\partial L_n}{\partial \theta_i}\right|^2 \leq \left|\sum \lambda_i^2\right| \left(\sum \left|\frac{\partial L_n}{\partial \theta_i}\right|^2\right)$$
$E|Y_t(\lambda)^{2+\delta}| < \infty$ holds. Because we have exponentially decreasing mixing coefficients
$$\sum_{k \geq 1} \alpha(k)^{1-\frac{2}{2+\gamma}}$$
holds. We can apply Theorem 1.5 of [Bos96]. So $Cov(Y_0(\lambda), Y_t(\lambda))$ is absolutely summable and
$$\sum_{k \in \mathbb{Z}} Cov(Y_0(\lambda), Y_t(\lambda)) > 0$$

5. First we assume $Y_t(\lambda)$ to have zero mean. Now we can apply a central limit theorem for mixing processes to be precise Theorem 1.7 in [Bos96]. So $\frac{1}{\sqrt{n}} \lambda' \frac{\partial L_n}{\partial \theta}|_{\theta_0} \to^{\mathcal{D}} N(0, s_\lambda^2)$ for every $\lambda \neq 0$. The general result holds if we apply this theorem to the translated process. Finally, by the Cramer Wold theorem this implies the statement of the lemma.

□

Theorem 6.5.2. *Let some assumptions hold ensuring that $\hat{\theta}_n$ is a consistent estimator for θ_0. Assume additionally:*

1. *g is two times continuously differentiable.*

2. $$\frac{1}{n} \frac{\partial^2 L_n}{\partial \theta \partial \theta'}|_{\theta^*} \to_p A(\theta_0) \text{ for } \theta^* \to_p \theta_0$$
for an invertible matrix $A(\theta_0)$.

Then

$$\sqrt{N}(\hat{\theta}_n - \theta_0) \to^{\mathcal{D}} N(0, A^{-1}BA^{-1}) \tag{6.26}$$

where $A = A(\theta_0)$ and B is the matrix from Lemma 6.5.1.

Proof. We make use of the Taylor expansion of $\frac{\partial L_n}{\partial \theta}|_{\hat{\theta}_n}$

$$\frac{\partial L_n}{\partial \theta}|_{\hat{\theta}_n} = \frac{\partial L_n}{\partial \theta}|_{\theta_0} + \frac{\partial^2 L_n}{\partial \theta \partial \theta'}|_{\theta^*}(\hat{\theta}_n - \theta_0) \tag{6.27}$$

$\frac{\partial L_n}{\partial \theta}|_{\hat{\theta}_n} = 0$ because it is a minimum of L_t in the interior of the parameter space and is identifiable unique.

$$\Rightarrow \sqrt{N}(\hat{\theta}_n - \theta_0) = (\frac{1}{n}\frac{\partial^2 L_n}{\partial \theta \partial \theta'}|_{\theta^*})^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_n}{\partial \theta}|_{\theta_0} \tag{6.28}$$

$$\to^{\mathcal{D}} A^{-1}G(\theta_0) \text{ with } G =^{\mathcal{D}} N(0, B) \tag{6.29}$$

Where $G =^{\mathcal{D}} N(0, B)$ is implied by Lemma 6.5.1. Then

$$Cov(A^{-1}G, A^{-1}G) = E(A^{-1}GG'A^{-1}) = A^{-1}E(GG')A^{-1} = A^{-1}BA^{-1}$$

So $A^{-1}G$ has the desired distribution. □

6.6 Asymptotic Normality and alternative Consistency for the Quasi–Maximum Likelihood Estimator in the ACARCH-V model

We now switch back to the quasi maximum likelihood estimator in the model described at the beginning of section 6.3.
We do some calculations on the derivatives of l_t in order to establish asymptotic normality. On the way we get also some alternative approach to consistency. This consistency result is a bit nicer than the one established before, because the moment condition is weaker and implied by the stationarity condition of the model before choosing the possibly misspecified likelihood function.
We assume that $g(y)$ has the following parametric form

$$g(y) = f(a + by + c(\omega + \alpha y^2)) \tag{6.30}$$

where f is continuously differentiable. In order to do proper calculation on the derivatives of l_t we postulate:

Assumption 6.6.1. *f is continuously differentiable.*

Further we assume that $\alpha > 0$ for all $(\omega, \alpha, a, b, c) \in \mathbf{\Theta}$. Then we define $\tilde{\alpha}$ as the infimum of all α with $(\omega, \alpha, a, b, c) \in \mathbf{\Theta}$. Due to compactness we have $\tilde{\alpha} > 0$.
We calculate the first derivatives:

$$\begin{aligned} &\left(\frac{\partial l_t}{\partial \omega}, \frac{\partial l_t}{\partial \alpha}\right) \qquad (6.31) \\ &= \frac{1}{2v_t^2}\left(\frac{(X_t - \mu - p_t\Delta)^2}{v_t^2} - 1 + 2(X_t - \mu - p_t\Delta)f'(a + bX_{t-1} + cv_t^2)\Delta c\right)(1, X_{t-1}^2) \end{aligned}$$

$$\left(\frac{\partial l_t}{\partial a}, \frac{\partial l_t}{\partial b}, \frac{\partial l_t}{\partial c}\right) = -1\frac{(X_t - \mu - p_t\Delta)f'(a + bX_{t-1} + cv_t^2)\Delta}{v_t^2}(1, X_{t-1}, v_t^2) \quad (6.32)$$

$$\frac{\partial l_t}{\partial \mu} = -1\frac{(X_t - \mu - p_t\Delta)\mu}{v_t^2} \quad (6.33)$$

Propositition 6.6.2. *If $\mathbf{\Theta}$ is compact and $\alpha > 0$ for all $(\omega, \alpha, a, b, c) \in \mathbf{\Theta}$, then $E|X_t|^{2\delta} < \infty$ for $\delta \geq 1$ implies*

$$E \sup_\theta \left|\frac{\partial l_t}{\partial \theta_i}\right|^\delta < \infty$$

Proof. Now f is bounded, defined on $\mathbb{R}$ and continuously differentiable. So f' is bounded. We have already seen that

$$\left|\frac{(X_t - \mu - p_t\Delta)^2}{v_t^2}\right| \leq \frac{1}{\tilde{\omega}^2}(MX_t^2 + k)$$

for constants $\tilde{\omega}$,M and k. Further

$$\left|\frac{1}{v_t^2}\right| \leq \frac{1}{\tilde{\omega}}$$

and

$$\left|\frac{X_{t-1}^2}{v_t^2}\right| \leq \frac{1}{\alpha}$$

$$\left|\frac{1}{v_t}\right| \leq \frac{1}{\sqrt{\tilde{\omega}}}$$

and

$$\frac{X_{t-1}^n}{v_t^n} \leq \frac{1}{\tilde{\alpha}^{\frac{n}{2}}}$$

Here $\tilde{\alpha}$ denotes the infimum of α such that $(\omega, \alpha, a, b, c) \in \mathbf{\Theta}$ for some ω, a, b, c. By the compactness of $\mathbf{\Theta}$ and our assumption that $\alpha > 0$ for all tuples in $\mathbf{\Theta}$ $\tilde{\alpha} > 0$ holds. This yields for the two most extreme cases:

$$\left|\frac{\partial l_t}{\partial \alpha}\right| \leq \left|\frac{1}{2\tilde{\omega}}\frac{k+k_1X_t^2}{\tilde{\alpha}}\right| + \left|\frac{1}{\tilde{\alpha}}\right| + \left|\frac{(2X_t+k_2)k_3}{\tilde{\alpha}}\right|$$

for constants k, k_1, k_2,k_3. and

$$\left|\frac{\partial l_t}{\partial c}\right| \leq (2X_t+k_2)k_4$$

for an additional constant k_4. □

Corollary 6.6.3. *If* $E|X_t|^2 < \infty$, *then* $E\sup_\theta|\frac{\partial l_t}{\partial \theta_i}| < \infty$.

This gives us a slightly different version of the consistency result:

Theorem 6.6.4. *If we replace Assumption 6.3.5 numbers 1 and 2 with*

1'. $EX_t^2 < \infty$,

2'. f *is continuously differentiable,* f_D *and* f_ϵ *are continuous*

then the quasi maximum likelihood estimator is consistent.

Proof. First by the argument of the proofs in section 6.3 we can bound $|l_t(\theta)|$ uniformly for all $\theta \in \mathbf{\Theta}$ by

$$|\log(\tilde{\omega})| + \bar{\omega} + \bar{\alpha}|X_{t-1}|^2) + \frac{1}{\tilde{\omega}}(c_1 + c_2E(|X_t|^2))$$

where c_1, c_2 are positive constants, $\tilde{\omega}, \bar{\omega}, \bar{\alpha}$ are constants defined analogously to the proof of Lemma 6.1.3.
So $EX_t^2 < \infty$ implies $E|l_t| < \infty$. Because the result holds uniformly for all $\theta \in \mathbf{\Theta}$ we also have $E|l_t^*| < \infty$ and $E|l_{t*}| < \infty$ where we use the notation of Definition 5.2.3. Now we work with a strongly stationary and ergodic version of $\{X_t\}$. As functions of finitely many X_t l_t, l_t^* and l_{t*} are stationary and ergodic. Therefore, we can get a local law of large numbers via the ergodic theorem for stationary processes. Secondly $EX_t^2 < \infty$ implies $E\sup_\theta|\frac{\partial l_t}{\partial \theta_i}| < \infty$ which provides us with a Lipschitz condition: Taylor expansion yields

$$l_t(\theta_1) - l_t(\theta_2) = \frac{\partial l_t}{\partial \theta}|_{\theta^*}(\theta_1 - \theta_2)$$

This yields

$$|l_t(\theta_1) - l_t(\theta_2)| \leq \left|\left|\frac{\partial l_t}{\partial \theta}|_{\theta^*}\right|\right| \, ||(\theta_1 - \theta_2)||$$

This implies

$$|l_t(\theta_1) - l_t(\theta_2)| \leq \sup_\theta \left|\left|\frac{\partial l_t}{\partial\theta}\right|\right| ||(\theta_1 - \theta_2)|| \tag{6.34}$$

where $||(x_1, \ldots x_n)|| = \sum_{i=1}^n |x_i|$.
And we have because of the integrability of the derivatives (and stationarity):

$$\sup_n \frac{1}{n}\sum_{t=1}^n E\sup_\theta ||\frac{\partial l_t}{\partial\theta}|| < \infty$$

We set $b_t(y) = \sup_\theta \left|\left|\frac{\partial l_t}{\partial\theta}\right|\right|$,$d(\theta_1, \theta_2) = ||(\theta_1 - \theta_2)||$ and $h(z) = z$. Because we are fulfilling conditions A, 5.1 and 5.2 of that theorem, we can conclude with Theorem 5.1 of [PP97] that

$$\sup_{\theta\in\mathbf{\Theta}} \left|n^{-1}\sum_{t=1}^n [l_t(X_t, X_{t-1}, \theta) - El_t(X_t, X_{t-1}, \theta)]\right| \to_p 0$$

as $n \to \infty$ and

$$\{n^{-1}\sum_{t=1}^n El_t(X_t, X_{t-1}, \theta) : n \in \mathbb{N}\}$$

is equicontinuous on $\mathbf{\Theta}$.
The assumed uniqueness of θ_0 yields the consistency. □

This version seems to be nicer than the one established in Theorem 6.1.12 , because the moment condition in the case that the model before assuming normality was not misspecified is implied by the stationarity condition. On the other hand f has to be differentiable, we have to use the compactness of $\mathbf{\Theta}$ in the dimension corresponding to c and we have to claim that $\alpha > 0$ for all $(\ldots, \alpha, \ldots) \in \mathbf{\Theta}$.

Corollary 6.6.5. *If $E|X_t|^{4+\gamma'} < \infty$ for a $\gamma' > 0$, then the moment condition in Theorem 6.5.1 is fulfilled.*

Now switching to the second derivatives we first make an assumption:

Assumption 6.6.6. *f is two times continuously differentiable*

We calculate:

$$\begin{aligned}
&\left(\frac{\partial^2 l_t}{\partial^2\omega^2}, \frac{\partial^2 l_t}{\partial\omega\partial\alpha}, \frac{\partial^2 l_t}{\partial^2\alpha^2}\right) && (6.35)\\
= \ & \left[\frac{1}{2v_t^4}\left(\frac{(X_t - \mu - p_t\Delta)^2}{v_t^2} - 1 + 2(X_t - \mu - p_t\Delta)f'(a + bX_{t-1} + cv_t^2)\Delta c\right)\right.\\
+ \ & \frac{1}{2v_t^2}\frac{(X_t - \mu - p_t\Delta)^2}{v_t^4}\\
+ \ & \left.\frac{1}{v_t^2}((-(f'(a + bX_{t-1} + cv_t^2))^2)c^2\Delta^2 + (X_t - \mu - p_t\Delta)f''(a + bX_{t-1} + cv_t^2)c\Delta)\right]\\
& (1, X_{t-1}^2, X_{t-1}^4)
\end{aligned}$$

$$\left(\frac{\partial^2 l_t}{\partial^2 a^2}, \frac{\partial^2 l_t}{\partial a \partial b}, \frac{\partial^2 l_t}{\partial a \partial c}, \frac{\partial^2 l_t}{\partial^2 b^2}, \frac{\partial^2 l_t}{\partial b \partial c}, \frac{\partial^2 l_t}{\partial^2 c^2}\right) \tag{6.36}$$
$$= -1\frac{(-(f'(a+bX_{t-1}+cv_t^2))^2)\Delta^2 + (X_t - \mu - p_t\Delta)f''(a+bX_{t-1}+cv_t^2)\Delta}{v_t^2}$$
$$(1, X_{t-1}, v_t^2, X_{t-1}^2, X_{t-1}v_t^2, v_t^4)$$

$$\frac{\partial^2 l_t}{\partial \alpha \partial c} \tag{6.37}$$
$$= \left(-1\frac{(-(f'(a+bX_{t-1}+cv_t^2))^2)c\Delta^2 + (X_t - \mu - p_t\Delta)f''(a+bX_{t-1}+cv_t^2)c\Delta}{v_t^2}\right.$$
$$+ \left.\frac{(X_t - \mu - p_t\Delta)f'(a+bX_{t-1}+cv_t^2)\Delta}{v_t^4}\right) v_t^2 X_{t-1}^2$$

Now here even after canceling (modulo constants) powers of σ_t or X_{t-1} against such of σ_t^{-1}, there remain powers of X_{t-1}. We use a corollary of Hölder's theorem to deal with that.

Propositition 6.6.7. *Let $i, j > 0$*

1. $$E|X^iY^j| \leq (E|X|^{i+j})^{\frac{i}{i+j}}(E|Y|^{i+j})^{\frac{j}{i+j}}$$

2. *Let $\{X_t\}$ be strongly stationary. If $E|X_t|^{i+j} = K < \infty$ then $E|X_t^iX_{t-1}^j| < \infty$.*

Proof. 1. Hölders theorem with $p = \frac{i+j}{i}$ and $q = \frac{i+j}{j}$ yields

$$[E|X^iY^j| \leq (E|X|^{i\frac{i+j}{i}})^{\frac{i}{i+j}}(E|Y|^{j\frac{i+j}{j}})^{\frac{j}{i+j}}$$

2. Because of the stationarity and that $E|X_{t-1}|^{i+j} = K$ also holds, point (1) of this Lemma yields the result.

□

So this Lemma gives us a method how to establish moment conditions on expressions h which are polynomials in $\mathbb{R}[X_t, X_{t-1}]$. We just have to map X_{t-1} to X_t. This gives a polynomial p in $\mathbb{R}[X_t]$. If X_t^m where m denotes the degree of p satisfies the moment condition so will h.

Propositition 6.6.8. *If $E|X_t|^{3\delta} < \infty$ for a $\delta \geq 1$ then*

$$E\left|\frac{\partial^2 l_t}{\partial^2 \theta_i \theta_j^2}\right|^\delta < \infty$$

Proof. We survey two of the most extreme cases:

$$\left|\frac{\partial^2 l_t}{\partial^2 c^2}\right| \leq c_1 + c_2|X_{t-1}|^2 + c_3|X_t||X_{t-1}|^2$$

for constants c_1, c_2, c_3. And for appropriate constantsc_4 to c_8:

$$\left|\frac{\partial^2 l_t}{\partial^2 \alpha^2}\right| \leq c_4 + c_5|X_t| + c_6|X_t|^2 + c_7|X_{t-1}|^2 + c_8|X_t||X_{t-1}|^2$$

Now Proposition 6.6.7 yields the finiteness of $E\left|\frac{\partial^2 l_t}{\partial^2 \theta_i \theta_j^2}\right|^{\delta}$ if $E|X_t|^{3\delta} < \infty$ □

Theorem 6.6.9. *If* $E|X_t|^{3+\gamma} < \infty$ *for a* $\gamma > 0$ *then*

$$\frac{1}{n}\frac{\partial^2 L_n}{\partial\theta\partial\theta'}|_{\theta^*} \rightarrow_p A(\theta_0) \textit{ for } \theta^* \rightarrow_p \theta_0$$

Proof. According to Theorem 4.1.5 in [Ame85] in order to show that

$$\frac{1}{n}\frac{\partial^2 L_n}{\partial\theta\partial\theta'}|_{\theta^*} \rightarrow_p A(\theta_0) \text{ for } \theta^* \rightarrow_p \theta_0$$

it suffices to show that an uniform law of large numbers applies to the second partial derivatives

$$\frac{\partial^2 l_t}{\partial\theta\partial\theta'}$$

We do that via Theorem 5.2 in [PP97] and we use the notation of Chapter 5. Proposition 6.6.8 shows that to be sure that the δst moments of all second partials exist we must claim the existence of the 3δst moments of $|X_t|$.
All second partials are continuous and the set of second partials is finite. So this set is equicontinuous. The tightness of H_n is assured by the moment condition too. If $E|X_t|^{3+3\gamma} < \infty$ then for all $\theta_{fixed} \in \mathbf{\Theta}$ then

$$\left|\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}||_{\theta_{fixed}}\right|^{1+\gamma} < +p(X_t, X_{t-1}) \tag{6.38}$$

where $p(X_t, X_{t-1})$ is a polynomial in X_t, X_{t-1} with monomials of maximal grade $3 + 3\gamma$. This implies

$$E\sup_{\theta}|\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}| < \infty \tag{6.39}$$

With the stationarity and mixing conditions of X_t we get these properties for

$$\frac{\partial^2 l_t}{\partial\theta\partial\theta'}$$

We need the first to conclude

$$\sup_n \frac{1}{n} \sum_{t=1}^{n} E \sup_\theta |\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}| < \infty$$

The second imply L_0 approximability of $\frac{\partial^2 l_t}{\partial \theta \partial \theta'}$.
We conclude with Theorem 6.13 of [PP97] that local laws of large numbers hold. Setting $K = 1$, $r_{kt} = 1$, $s_{kt} = \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}$ we can conclude with Theorem 5.2 of [PP97] that

$$\sup_{\theta \in \boldsymbol{\Theta}} \frac{1}{n} \sum_{t=1}^{n} |\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} - E \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}| \rightarrow_p 0$$

because then conditions B, C and D are already verified and compactness of the parameter space was assumed. Now taking an open neighborhood $\Theta' \subset \boldsymbol{\Theta}$ we can use Theorem 4.1.5 in Amemiya [Ame85] which gives us the desired result. □

Assumption 6.6.10. *1. The Assumptions of 6.3.5 3–6*

2. Either Assumption 6.3.5 point 1 or the assumption 1' of Theorem 6.6.4

3. f_D and f_ϵ are continuous

4. f is two times continuously differentiable

5. $\alpha_0 > 0$

6. θ_0 is in the interior of $\boldsymbol{\Theta}$.

7. $E|X_t|^{4+\gamma} < \infty$ for a $\gamma > 0$.

Theorem 6.6.11. *If 6.6.10 holds then*

$$\sqrt{N}(\hat{\theta}_n - \theta_0) \rightarrow^{\mathcal{D}} N(0, A^{-1} B A^{-1}) \tag{6.40}$$

where A and B are the matrices arising from Lemma 6.5.1 and Theorem 6.5.2.

Proof. Under the assumptions given here we first match the assumptions of Lemma 6.5.1. Secondly Theorem 6.6.9 established

$$\frac{1}{n} \frac{\partial^2 L_n}{\partial \theta \partial \theta'}|_{\theta^*} \rightarrow_p A(\theta_0) \text{ for } \theta^* \rightarrow_p \theta_0$$

If θ_0 is identifiable unique then $A(\theta_0)$ being the hessian of a local extremum is invertible. So we match the assumptions of Theorem 6.5.2 to conclude. □

Similarly to have third partials with δst moments we must have the 5δst moments of $|X_t|$.

6.7 Asymptotic Normality and alternative Consistency for the Quasi–Maximum Likelihood Estimator in the CARCH-S model

We want to see if the results of the previous section carry over to the CARCH-S model which we have described at the beginning of section 6.1. Again we assume that $g(y)$ has the following parametric form

$$g(y) = f(a + by + c(\omega + \alpha y^2)) \tag{6.41}$$

In order to do calculations on the derivatives state:

Assumption 6.7.1. *f is continuously differentiable*

Theorem 6.7.2. *For every $\delta \geq 1$, if $E|X_t|^{2\delta} < \infty$ then*

$$E \sup_{\theta} \left| \frac{\partial l_t}{\partial \theta_i} \right|^{\delta} < \infty$$

Proof. Switching back to the CARCH-S model we also calculate the most extreme derivatives of the Quasi–Maximum Likelihood Estimator:

$$\begin{aligned}
&\left(\frac{\partial l_t}{\partial \omega}, \frac{\partial l_t}{\partial \alpha}\right) \qquad (6.42)\\
= &\left(\frac{1}{2\sigma_t^2} + \frac{1}{2h_t^2} T f'(a + bX_{t-1} + c\sigma_{t-1}^2)c)\right.\\
+ &\frac{1}{2h_t^2}\left(\frac{(X_t - \sigma_t(\mu_\epsilon + p_t\Delta))^2}{\sigma_t^4}\right.\\
&+2(X - \sigma_t(\mu_\epsilon + p_t\Delta))\left(\frac{(\mu_\epsilon + p_t\Delta)}{\sigma_t^3} + \frac{f'(a + bX_{t-1} + c\sigma_{t-1}^2)c\Delta}{\sigma_t}\right)\Bigg)\\
+ &\left.\frac{1}{2h_t^4} T f'(a + bX_{t-1} + c\sigma_{t-1}^2)c\frac{(X_t - \sigma_t(\mu_\epsilon + p_t\Delta)^2)}{\sigma_t^2}\right)(1, X_{t-1}^2)
\end{aligned}$$

Here T is an abbreviation of the term:

$$\mu_{D^2} - \mu_{\epsilon^2} + 2[\mu_\epsilon + f(a + bX_{t-1} + c\sigma_{t-1}^2)(\mu_D - \mu_\epsilon)](\mu_D - \mu_\epsilon)$$

$$
\begin{aligned}
& \left(\frac{\partial l_t}{\partial a}, \frac{\partial l_t}{\partial b}, \frac{\partial l_t}{\partial c}\right) && (6.43)\\
= & \left(\frac{1}{2h_t^2}(Tf'(a + bX_{t-1} + c\sigma_{t-1}^2))\right.\\
+ & \frac{1}{h_t^2\sigma_t^2}((X_t - \sigma_t(\mu_\epsilon + p_t\Delta))(\sigma_t f'(a + bX_{t-1} + c\sigma_{t-1}^2)\Delta)\\
+ & \left.\frac{1}{2h_t^4}Tf'(a + bX_{t-1} + c\sigma_{t-1}^2)\frac{(X_t - \sigma_t(\mu_\epsilon + p_t\Delta))^2}{\sigma_t^2}\right)\\
& (1, X_{t-1}, \sigma_t^2)
\end{aligned}
$$

We prove that we can bound the derivatives uniformly in θ by expressions just depending on constants and monomials in X_i $i \in \{t, t-1\}$ of grade 2. In analogy to the previous section we get for positive constants $C_1, \dots, C_7$:

$$
\left|\frac{\partial l_t}{\partial \alpha}\right| \leq \frac{1}{2\tilde{\omega}} + \frac{C_2}{C_1} + \frac{1}{C_1}\left(\frac{C_3 + C_4 X_t^2}{\tilde{\omega}} + \frac{C_5 + C_6|X_t|}{\sqrt{\tilde{\omega}}} + C_7|X_t||X_{t-1}|\right)
$$

and

$$
\left|\frac{\partial l_t}{\partial c}\right| \leq K_1 + K_2|X_t| + K_3 X_t^2
$$

for positive constants K_1 to K_3. □

Corollary 6.7.3. *If we replace Assumption 6.1.12 1 and 2 with*

1'. $EX_t^2 < \infty$

2'. f *is continuously differentiable,* f_D *and* f_ϵ *are continuous*

Then the quasi maximum likelihood estimator is consistent.

Proof. Using Theorem 6.7.2 we can copy the proof of Theorem 6.6.4. □

Propositition 6.7.4. *If* $E|X_t|^{4\delta} < \infty$ *for a* $\delta \geq 1$ *then*

$$
E\left|\frac{\partial^2 l_t}{\partial^2 \theta_i \theta_j^2}\right|^\delta < \infty
$$

Proof. In order to gain also asymptotic normality we have to calculate the second derivatives. Again we fix as a general assumption:

Assumption 6.7.5. f *is two times continuously differentiable*

We give here only $\frac{\partial^2 l_t}{\partial^2 c^2}$ because $\frac{\partial^2 l_t}{\partial^2 \alpha^2}$ while not adding complexity to the intrinsic mathematical problem is a formula of even more extreme length. So the argumentation would be blurred by pseudo exactness. Nothing to the complexity is added, because the additional effect of working in a GARCH–M context being present in the α derivatives is dominated by the effects of differentiation with respect to h_t in the product rule.

$$
\begin{aligned}
&\frac{\partial^2 l_t}{\partial^2 c^2} \qquad (6.44)\\
= & \left(\frac{1}{2h_t^4}(Tf'(a+bX_{t-1}+c\sigma_{t-1}^2))^2\right.\\
- & \frac{1}{2h_t^2}[Tf''(a+bX_{t-1}+c\sigma_{t-1}^2)+\frac{\partial T}{\partial c}f'(a+bX_{t-1}+c\sigma_{t-1}^2)]\\
+ & \left(\frac{1}{2h_t^4}(-\sigma_t f'(a+bX_{t-1}+c\sigma_{t-1}^2)\Delta)\frac{2(X_t-\sigma_t(\mu_\epsilon+p_t\Delta))}{\sigma_t^2}\right)\\
- & \frac{1}{2h_t^2}((-\sigma_t f''(a+bX_{t-1}+c\sigma_{t-1}^2)\Delta)\frac{2(X_t-\sigma_t(\mu_\epsilon+p_t\Delta))(\sigma_t f')^2}{\sigma_t^2})\\
+ & \frac{1}{2h_t^8}(Tf'(a+bX_{t-1}+c\sigma_{t-1}^2))^2\frac{(X_t-\sigma_t(\mu_\epsilon+p_t\Delta))^2}{\sigma_t^2}\\
- & \frac{1}{2h_t^4}\left([Tf''(a+bX_{t-1}+c\sigma_{t-1}^2)+\frac{\partial T}{\partial c}f'(a+bX_{t-1}+c\sigma_{t-1}^2)]\frac{(X_t-\sigma_t(\mu_\epsilon+p_t\Delta))^2}{\sigma_t^2}\right)\\
- & \left.\frac{1}{2h_t^4}\left((Tf'(a+bX_{t-1}+c\sigma_{t-1}^2))^2\frac{2(X_t-\sigma_t(\mu_\epsilon+p_t\Delta))}{\sigma_t^2})\right)\right)\sigma_t^4
\end{aligned}
$$

The dominant term here is $\frac{(X_t-\sigma_t(\mu_\epsilon+p_t\Delta))^2}{\sigma_t^2}\sigma_t^4$ which equals

$$X_t^2\sigma_t^2 - const.X_t\sigma_t^3+\sigma_t^4$$

This is a monomial of order 4 in X_t and X_{t-1}. So using Lemma 6.6.7. we get

$$E\left|\frac{\partial^2 l_t}{\partial^2\theta_i\theta_j^2}\right|^\delta < \infty$$

if $E|X_t^{4\delta} < \infty$. ☐

Theorem 6.7.6. *Assume*

1. *The Assumptions of 6.1.11 3–7*
2. *Either Assumption 6.1.11 point 1 and* $\alpha_0 > 0$ *or the assumption 1' of Corollary 6.7.3*
3. f_D *and* f_ϵ *are continuous*

4. *f is two times continuously differentiable*

5. $\alpha_0 > 0$

6. *θ_0 is in the interior of $\mathbf{\Theta}$.*

7. $E|X_t|^{4+\gamma} < \infty$ *for a* $\gamma > 0$.

Then

$$\sqrt{N}(\hat{\theta}_n - \theta_0) \to^{\mathcal{D}} N(0, A^{-1}BA^{-1}) \tag{6.45}$$

where A and B are the matrices arising from Lemma 6.5.1 and Theorem 6.5.2.

Proof. We can actually use the same proof like in Theorem 6.6.11. The slightly stronger moment condition we have here in Proposition 6.7.4 in comparison to the alternative model does not matter because the moment conditions arising from the first order conditions for asymptotic normality require the existence of slightly more than 4th moments anyway in both cases. □

6.8 Asymptotic Normality Results for the Maximum Likelihood Estimator in the CARCH-S Model

We consider the same model as in section 6.7. Rather than the arguments above we will use again the methodology of [PP97] to establish asymptotic normality results for the estimators containing the density functions of ϵ_t and D_t. So we work in the setting of section 6.2 and use the notation from this section. We further will use the following notation.

Definition 6.8.1. 1. $S_n := \frac{1}{n}\sum_{t=1}^{n} q_t$

2. $C_n = E\nabla^2 S_n$

3. $D_n = (nE(\nabla S_n' \nabla S_n))^{\frac{1}{2}}$

Where ∇ is the gradient and ∇^2 is the Hessian with respect to θ.

Theorem 6.8.2. *Let Assumptions 6.2.1 number 1 to 3, 5 and the assumptions of Proposition 6.2.5 hold. If additionally*

1. *f_ϵ, f_D and f are two times continously differentiable,*

2. $|uf_*'(u)| \leq const.u^2 f_*(u)$ *for* $|u| \to \infty$ *for* $* \in \{\epsilon, D\}$,

3. $|u^2 f_*''(u)| \leq const.u^4 f_*(u)$ *for* $|u| \to \infty$ *for* $* \in \{\epsilon, D\}$,

4. $E|X_t|^{4+4\gamma} < \infty$ *for a* $\gamma > 0$,

then $\hat{\theta}_n$ is asymptotically normal i.e.:

$$\sqrt{n}(\hat{\theta}_n - \bar{\theta}_0) = C_n^{-1} D_n \zeta_n + o_p(1) \tag{6.46}$$

with $\zeta_n \to^D N(0, I)$, I denoting the identity, and

$$\sqrt{n} D_n^{-1} C_n (\hat{\theta}_n - \bar{\theta}_0) \to^D N(0, I) \tag{6.47}$$

Furthermore the norms of C_n,D_n and their inverses are $O(1)$.
So $\hat{\theta}_n$ is $n^{\frac{1}{2}}$ consistent for $\bar{\theta}_n$

Beside statements marked as remarks the rest of this section is the proof of the above theorem.
In order to establish asymptotic normality of the maximum likelihood estimator we first establish a condition to fulfill assumption 11.5 of [PP97].
This is done by the following Proposition.

Propositition 6.8.3. *For $r \geq 2$ let*

1. $|uf'_\epsilon(u)| \leq const.u^2 f_\epsilon(u)$ *for* $|u| \to \infty$
2. $|uf'_D(u)| \leq const.u^2 f_D(u)$ *for* $|u| \to \infty$
3. $E|X_t|^{2r} < \infty$

then

$$E\left|\frac{\partial q_t}{\partial \theta_i}\right|^r < \infty$$

Proof.

$$\begin{aligned} &\frac{\partial q_t}{\partial \alpha} \\ &= -\frac{1}{2}\frac{1}{\sigma_t^2}X_{t-1}^2 \\ &\quad + \frac{1}{2}\frac{(1-p_t)\, f'_\epsilon\left(\frac{X_t}{\sigma_t}\right)\frac{X_t X_{t-1}^2}{\sigma_t^3} + p_t f'_D\left(\frac{X_t}{\sigma_t}\right)\frac{X_t X_{t-1}^2}{\sigma_t^3}}{(1-p_t)\, f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)} \\ &\quad + \frac{\frac{\partial p_t}{\partial \alpha}\left(-f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + f_D\left(\frac{X_t}{\sigma_t}\right)\right)}{(1-p_t)\, f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)} \end{aligned} \tag{6.48}$$

$E|X_t|^{2r} < \infty$ implies the statement if we can somehow use the terms in the numerator to neutralize the asymptotical tendency of the nominator towards zero such that the whole fraction asymptotically does not grow faster than $|X_t|^{2r}$.

Because for $\delta > 1$

$$\left(\sum_{t=1}^{n} T_i\right)^{\delta} \leq 2^{\delta(n-1)} \sum_{t=0}^{n} \left(\sum_{t=0}^{n} T_i^{\delta}\right)$$

we can treat each summand separately.
The last summand can be treated by:

$$\left|\frac{\frac{\partial p_t}{\partial \alpha}\left(-f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + f_D\left(\frac{X_t}{\sigma_t}\right)\right)}{(1-p_t)\, f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)}\right| \leq \left|\frac{\frac{\partial p_t}{\partial \alpha}}{1-p_t} + \frac{\frac{\partial p_t}{\partial \alpha}}{p_t}\right|$$

Because p_t is bounded away from zero and one the denominator does not play any role. Again using the parametric form $p_t = f(a + bX_{t-1} + c\sigma_t^2)$.

$$\frac{\partial p_t}{\partial \alpha} = f'(a + bX_{t-1} + c\sigma_t)cX_{t-1}^2$$

The outer derivatives is bound by constants, due to the continuous differentiability of f. The inner derivative cX_{t-1}^2 allows rth moments if X_t^2 does.
In fact this argumentation hold also holds for $p_t = f(a + bX_{t-1} + cl(\sigma_t))$ where l is a continuous concave function which is positive on the positive real line.
Now we want to handle

$$\frac{(1-p_t)\, f_\epsilon'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t X_{t-1}^2}{\sigma_t^3} + p_t f_D'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t X_{t-1}^2}{\sigma_t^3}}{(1-p_t)\, f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)}$$

Firstly, as

$$\frac{X_{t-1}^2}{\sigma_t^2} = \frac{X_{t-1}^2}{\omega + \alpha X_{t-1}^2} \leq \frac{1}{\alpha}$$

and due to our assumptions on $\mathbf{\Theta}$

$$\begin{aligned}
&\left|\frac{(1-p_t)\, f_\epsilon'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t X_{t-1}^2}{\sigma_t^3} + p_t f_D'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t X_{t-1}^2}{\sigma_t^3}}{(1-p_t)\, f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)}\right| \\
\leq\ & c\frac{\left|(1-p_t)\, f_\epsilon'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t} + p_t f_D'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t}\right|}{(1-p_t)\, f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)}
\end{aligned}$$

for a positive constant c.

Secondly,

$$\frac{\left|(1-p_t)\, f_\epsilon'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t}+p_t f_D'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t}\right|}{(1-p_t)\, f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_t f_D\left(\frac{X_t}{\sigma_t}\right)} \tag{6.49}$$

$$\leq \frac{\left|(1-p_t)\, f_\epsilon'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t}\right|}{(1-p_t)\, f_\epsilon\left(\frac{X_t}{\sigma_t}\right)}+\frac{\left|p_t f_D'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t}\right|}{p_t f_D\left(\frac{X_t}{\sigma_t}\right)}$$

$$= \frac{\left|f_\epsilon'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t}\right|}{f_\epsilon\left(\frac{X_t}{\sigma_t}\right)}+\frac{\left|f_D'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t}\right|}{f_D\left(\frac{X_t}{\sigma_t}\right)}.$$

So it suffices to investigate

$$\frac{f_\epsilon'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t}}{f_\epsilon\left(\frac{X_t}{\sigma_t}\right)}$$

and

$$\frac{f_D'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t}}{f_D\left(\frac{X_t}{\sigma_t}\right)}.$$

If the support of both densities is the entire space $\mathbb{R}$ we just have to investigate the behavior for $|\frac{X_t}{\sigma_t}| \to \infty$.
We set as usual $\eta_t := \frac{X_t}{\sigma_t}$.
If the conditions 1 and 2 are fulfilled we have for $*$ being D or ϵ.

$$\frac{|f_*'(\eta_t)\,\eta_t|}{f_*(\eta_t)} \leq const.|\eta_t|^2$$

for $|\eta_t| \geq c$ we get

$$\frac{|f_*'(\eta_t)\,\eta_t|^r}{f_*(\eta_t)} \leq const.|\eta_t|^{2r}$$

Because of the definition of η_t and the fact that $\frac{1}{\sigma_t^2}$ is bounded by $\tilde{\omega}^{-1}$ the finiteness of $E|X_t|^{2r}$ implies the finiteness of $E|\eta_t|^{2r}$. This implies

$$E|f_*'(\eta_t)\,\eta_t|^{2r} < \infty.$$

This implies using 6.49

$$E\frac{\left|(1-p_t)\, f_\epsilon'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t}+p_t f_D'\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t}\right|^{2r}}{(1-p_t)\, f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_t f_D\left(\frac{X_t}{\sigma_t}\right)} < \infty$$

which with the argumentation before yields the result for $\theta_i = \alpha$. To handle

$$\frac{\partial q_t}{\partial \omega}$$

we see that

$$\begin{aligned}
&\frac{\partial q_t}{\partial \omega} \\
&= -\frac{1}{2}\frac{1}{\sigma_t^2} \\
&\quad + \frac{(1-p_t)\, f'_\epsilon\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t^3} + p_t f'_D\left(\frac{X_t}{\sigma_t}\right)\frac{X_t}{\sigma_t^3}}{(1-p_t)\, f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)} \\
&\quad + \frac{p'_t\left(-f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + f_D\left(\frac{X_t}{\sigma_t}\right)\right)}{(1-p_t)\, f_\epsilon\left(\frac{X_t}{\sigma_t}\right) + p_t f_D\left(\frac{X_t}{\sigma_t}\right)}
\end{aligned} \tag{6.50}$$

Now $\frac{1}{\sigma_t^2}$ is bounded by $\tilde{\omega}^{-1}$. This handles the first summand. But having the calculation in equation 6.49 in mind and the fact that

$$\frac{X_t}{\sigma_t^3} = \frac{X_t}{\sigma_t}\frac{1}{\sigma_t^2}$$

this argument also handles the second summand. The third one again can be treated like in the case of the quasi maximum likelihood. Derivatives in direction of the parametrisation of p_t also can be done using the last argument, because we don't have to deal with derivatives of the densities in this case. ☐

Corollary 6.8.4. *Under the assumptions of the preceeding Proposition Assumption 11 of [PP97] is fulfilled.*

Proof. Because q_t does not depend on t

$$\sup_t E\left|\frac{\partial q_t}{\partial \theta_i}\right|^r = E\left|\frac{\partial q_t}{\partial \theta_i}\right|^r$$

Because of this argument and the continuity of q_t the concept of identifiable uniqueness collapses to the existence of a unique minimizer $\bar{\theta}_0$. So to ensure assumption 11.5 of [PP97] it suffices to show that:

$$E\left|\frac{\partial q_t}{\partial \theta_i}(\bar{\theta}_0)\right|^r < \infty$$

We are working with the geometrically α mixing version of X_t. So X_t is near epoch dependent on itself of rate 1 and moreover fulfills the conditions on the mixing rates. ☐

Remark 6.8.5. $|uf_*'(u)| \leq const.u^2 f_*(u)$ is satisfied for

$$f_* \propto \exp(-x^\gamma)\,,\ \gamma \leq 2$$

and for

$$f_* \propto \frac{1}{x^\gamma}$$

In the latter case to apply Proposition 6.8.3 we cannot choose γ arbitrarily small, because on the other hand we have to fulfill $E|X_t|^{2r} < \infty$.

Proof. 1.

$$f_*(u) \propto \exp(-u^\gamma) \Rightarrow f_*'(u) \propto \gamma u^{\gamma-1} \exp(-u^\gamma)$$

So asymptotically holds

$$\frac{uf_*'(u)}{f_*} \propto u^\gamma \leq u^2$$

2.

$$f_*(u) \propto \frac{1}{u^\gamma} \Rightarrow f_*'(u) \propto -\gamma \frac{1}{u^{\gamma+1}} \propto \frac{1}{u} f_*(u)$$

□

In the next step we translate the moment condition in assumption 11.2 of [PP97] in an concept comparable to the one above.

Propositition 6.8.6. *Suppose we work in a compact parameter space* $\mathbf{\Theta}$*. If additionally to the requirements of Proposition 6.8.3*

$$|f_*(u)''| \leq const.u^2 f_*(u)$$

is satisfied then

$$E \sup_\theta \left| \frac{\partial q_t}{\partial \theta_i} \right|^{1+\gamma} < \infty$$

Proof. We will look at the most extreme second derivative:

$$
\begin{aligned}
& \frac{\partial^2 q_t}{\partial^2 \alpha^2} \qquad (6.51)\\
= \quad & -\frac{1}{2}\frac{1}{\sigma_t^4}X_{t-1}^4 \\
& -\frac{1}{4}\frac{\left((1-p_t)f'_\epsilon\left(\frac{X_t}{\sigma_t}\right)\frac{X_tX_{t-1}^2}{\sigma_t^3}+p_tf'_D\left(\frac{X_t}{\sigma_t}\right)\frac{X_tX_{t-1}^2}{\sigma_t^3}\right)^2}{\left((1-p_t)\,f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_tf_D\left(\frac{X_t}{\sigma_t}\right)\right)^2} \\
& +\frac{1}{4}\frac{[(1-p_t)\,f''_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_tf''_D\left(\frac{X_t}{\sigma_t}\right)]\frac{X_t^2X_{t-1}^4}{\sigma_t^6}}{(1-p_t)\,f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_tf_D\left(\frac{X_t}{\sigma_t}\right)} \\
& -\frac{3}{4}\frac{[(1-p_t)\,f'_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_tf'_D\left(\frac{X_t}{\sigma_t}\right)]\frac{X_tX_{t-1}^2}{\sigma_t^5}}{(1-p_t)\,f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_tf_D\left(\frac{X_t}{\sigma_t}\right)} \\
& -\frac{\left(\frac{\partial p_t}{\partial\alpha}\left(-f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right)^2}{\left((1-p_t)\,f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_tf_D\left(\frac{X_t}{\sigma_t}\right)\right)^2} \\
& -\frac{1}{2}\frac{\left(\frac{\partial p_t}{\partial\alpha}\left(-f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right)\left((1-p_t)f'_\epsilon\left(\frac{X_t}{\sigma_t}\right)\frac{X_tX_{t-1}^2}{\sigma_t^3}+p_tf'_D\left(\frac{X_t}{\sigma_t}\right)\frac{X_tX_{t-1}^2}{\sigma_t^3}\right)}{\left((1-p_t)\,f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_tf_D\left(\frac{X_t}{\sigma_t}\right)\right)^2} \\
& +\frac{\left(\frac{\partial^2 p_t}{\partial\alpha^2}\left(-f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+f_D\left(\frac{X_t}{\sigma_t}\right)\right)\right)}{(1-p_t)\,f_\epsilon\left(\frac{X_t}{\sigma_t}\right)+p_tf_D\left(\frac{X_t}{\sigma_t}\right)}
\end{aligned}
$$

Again we can treat each summand separately. As general moment condition we fixed $E|X_t|^{4(1+\gamma)} < \infty$ for a $\gamma > 0$ which is the same as $E|X_t|^{2r} < \infty$ for a$r > 2$. This handles the first term and will also be used implicitely for the following terms.
The requirement that the $1+\gamma$th moment of the second term exists is just a redundancy of the requirement above. The third term also leads to similar tail conditions, namely.

$$|u^2 f''_*(u)| \leq const.u^4 f_*(u) \qquad (6.52)$$

Which boils down to

$$|f''_*(u)| \leq const.u^2 f_*(u) \qquad (6.53)$$

because of the even grade polynomials on the real numbers take only positive values.

To deal with the fourth term we can use the arguments for the derivation in ω direction of the proof of the previous proposition. The fifth and the sixth term are bounded in expectation under our conditions by the arguments of the proof of Proposition 6.8.3. The same argument and Hölder's inequality yield the result for the seventh term. All second derivatives are continuous and there are only finitely many. Hence this family is equicontinuous. So all of Assumption 11.2 of [PP97] is implied.

$\square$

Corollary 6.8.7. *Under the assumptions of Proposition 6.8.6 Assumption 11.2 of [PP97] is fulfilled.*

Proof. If we write $q_t = q_t(X_t, X_{t-1}, \theta)$ then first $q_t(x, y, \theta)$ is the same function for all t, so are $\frac{\partial q_t(x,y\theta)}{\partial \theta_i}$ for each fixed θ_i. Hence the following formula holds:

$$\frac{1}{n}\sum_{t=1}^{n} E \sup_{\theta\in\mathbf{\Theta}} \left|\frac{\partial q_t(X_t, X_{t-1}, \theta)}{\partial \theta_i}\right|^{1+\gamma} = E \sup_{\theta\in\mathbf{\Theta}} \left|\frac{\partial q_t(X_t, X_{t-1}, \theta)}{\partial \theta_i}\right|^{1+\gamma}$$

Therefore, we just have to check the latter moment condition. In fact we will get comparable moment and tail conditions ensuring

$$\frac{1}{n}\nabla^2 q_t \rightarrow_p B$$

for a matrix B. $\square$

If we have a unique minimizer in the interior also 11.3 is fulfilled. To get assumption 11.1 we just have to consider besides the assumptions for consistency that the unique minimizer is in the interior of $\mathbf{\Theta}$ to imply (d) and continuous differentiability of f, f_ϵ and f_D.

6.9 Asymptotic Normality of the Maximum Likelihood Estimator in the ACARCH-V model

Now we consider again the model described in the beginning of section 6.3. We first realize that we can use the main idea of the previous section that is to split the derivations in terms depending on f_D or f_ϵ exclusively. Then it suffices to impose conditions on the asymptotic relation between these functions and their derivatives. We fix as a general assumption:

Assumption 6.9.1. *f_ϵ, f_D and f are two times continuously differentiable*

The trick we used in the CARCH-S model must be modified when we try to apply it to the ACARCH-V model. The reason is that the terms inside the densities are to complex. We use again the notation f_* if we mean f_D and f_ϵ simultaneously. Trying to derive reasonable conditions to impose on the tails of f_ϵ and f_D we come up to the following most complicated derivative:

$$\begin{aligned}
&\frac{\partial}{\partial\alpha} f_*([X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]v_t^{-1}h_t) \qquad (6.54)\\
&= \quad f_*'([X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]v_t^{-1}h_t)\\
&\qquad \{c\Delta f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))v_t^{-1}h_t\\
&\qquad +[X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]\\
&\qquad ([f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))c\Delta(\sigma_D - \sigma_\epsilon)v_t^{-1}]\\
&\qquad + \frac{-1}{2}[h_t v_t^{-3}])\Big\}
\end{aligned}$$

We will use this in order to get condition 11.5 of [PP97].

Propositition 6.9.2. *Let $r > 2$. If*

1. $E|X_t|^{2r} < \infty$
2. $|uf_\epsilon'(u)| \leq const.u^2 f_\epsilon(u)$ *for* $|u| \to \infty$
3. $|uf_D'(u)| \leq const.u^2 f_D(u)$ *for* $|u| \to \infty$ *Then*

$$E\left|\frac{\partial q_t}{\partial \theta_i}\right|^r < \infty$$

Proof. We can use argumentation in analogy to the CARCH-S model to handle the term:

$$\begin{aligned}
&f_*'([X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]v_t^{-1}h_t)\\
&[X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]\\
&\frac{-1}{2}h_t X_{t-1}^2 v_t^{-3}
\end{aligned}$$

and come up with the condition $|uf_*'(u)| \leq const.u^2 f_*(u)$. Here u was substituted for $[X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]$ which moments exist if the ones of X_t do.
To handle

$$\begin{aligned}
&f_*'([X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]v_t^{-1}h_t)\\
&[X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]\\
&([f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))c\Delta(\sigma_D - \sigma_\epsilon)]v_t^{-1}X_{t-1}^2
\end{aligned}$$

we can argue that $f'(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))c\Delta(\sigma_D-\sigma_\epsilon)$ and h_t are bounded by constants and therefore require $|f'_*(u)| \leq const.uf_*(u)$.
If this holds $E|f'_*(u)X_{t-1}|^r < \infty$ holds if $E|uX_{t-1}|^r < \infty$ which is implied by $E|X_t|^{2r} < \infty$.
To handle

$$\begin{aligned}&f'_*([X_t-\mu+f(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))\Delta]v_t^{-1}h_t)\\&c\Delta X_{t-1}^2 f'(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))v_t^{-1}h_t\end{aligned}$$

we can argue that $f'(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))h_t c\Delta$ is bounded by constants and the expectation of $(X_{t-1}^2 v_t^{-1})^r$ exists if the expectation of $|X_{t-1}|^r$ exists. Assuming the latter we can again require $|f'_*(u)| \leq const.uf_*(u)$. If this holds again $E|f'_*(u)X_{t-1}|^r < \infty|$ holds under the assumption $E|uX_{t-1}|^r < \infty$ which is implied by $E|X_t|^{2r} < \infty$. But $|uf_*(u)| \leq const.u^2 f_*(u)$ implies $|f_*(u)| \leq const.uf_*(u)$ asymptotically. so we don't get a new condition. □

Propositition 6.9.3. *Let $|uf''_*(u)| \leq u^2 f_*(u)$ hold asymptotically in addition to the Assumptions of Proposition 6.9.2. Then Assumption 11.2 of [PP97] is fulfilled.*

Proof. To investigate the second derivatives if suffices to explore the most extreme term of the most extreme derivative. The most extreme derivative is $\frac{\partial^2}{\partial\alpha^2}$. The summands of the nominator of the most extreme term are:

$$\begin{aligned}&f''_*([X_t-\mu+f(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))\Delta]v_t^{-1}h_t)\\&\quad\{c\Delta X_{t-1}^2 f'(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))v_t^{-1}h_t\\&\quad+[X_t-\mu+f(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))\Delta]\\&\quad([f'(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))c\Delta(\sigma_D-\sigma_\epsilon)v_t^{-1}]\\&\quad+[\frac{-1}{2}h_t X_{t-1}^2 v_t^{-3}])\}\\+\;&f'_*([X_t-\mu+f(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))\Delta]v_t^{-1}h_t)\\&\quad\frac{\partial}{\partial\alpha}\left\{c\Delta X_{t-1}^2 f'(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))v_t^{-1}h_t\right.\\&\quad+[X_t-\mu+f(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))\Delta]\\&\quad\left([f'(a+bX_{t-1}+c(\omega+\alpha X_{t-1}))c\Delta(\sigma_D-\sigma_\epsilon)X_{t-1}^2 v_t^{-1}]\right.\\&\quad\left.\left.+\left[\frac{-1}{2}h_t X_{t-1}^2 v_t^{-3}\right]\right)\right\}\end{aligned} \tag{6.55}$$

The first part of this term can be handled like in the proof of Proposition 6.9.2, f''_* taking the role of f'_*.

It remains to handle:

$$
\begin{aligned}
& f'_*([X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]v_t^{-1}h_t) \\
& \quad \frac{\partial}{\partial\alpha}\{c\Delta X_{t-1}^2 f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))v_t^{-1}h_t \\
& \quad +[X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta] \\
& \quad ([f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))c\Delta(\sigma_D - \sigma_\epsilon)X_{t-1}^2 v_t^{-1}] \\
& \quad + \left[\frac{-1}{2}h_t X_{t-1}^2 v_t^{-3}\right]\Big)\Big\}
\end{aligned} \tag{6.56}
$$

Now f and its first and second derivatives are bounded by constants. So this term can be maximally of order of X_{t-1}^3 because v_t^{-1} cancels modulo constants with X_t^4 to X_{t-1}^3. And this extreme case is the fact. Splitting into summands we get:

$$
\begin{aligned}
& \frac{\partial}{\partial\alpha}(c\Delta X_{t-1}^2 f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))v_t^{-1}h_t) \\
= \; & X_{t-1}t^4 v_t^{-1} \\
& [c^2\Delta^2(f''(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))h_t + f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))(\sigma_D - \sigma_\epsilon)) \\
& + \frac{-1}{2}c\Delta f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))h_t v_t^{-2}]
\end{aligned} \tag{6.57}
$$

$$
\begin{aligned}
& \frac{\partial}{\partial\alpha}\{[X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta] \\
& [f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))c\Delta(\sigma_D - \sigma_\epsilon)X_{t-1}^2 v_t^{-1}]\} \\
= \; & X_{t-1}^4 v_t^{-1} \\
& \{(f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))^2 c^2\Delta^2(\sigma_D - \sigma_\epsilon) \\
& +[X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta] \\
& [c^2\Delta^2(f''(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))h_t + f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))(\sigma_D - \sigma_\epsilon))] \\
& + \frac{-1}{2}[X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]f'(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))c\Delta v_t^{-2}\}
\end{aligned} \tag{6.58}
$$

$$
\begin{aligned}
& \frac{\partial}{\partial\alpha}\left(\left[\frac{-1}{2}h_t X_{t-1}^2 v_t^{-3}\right][X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]\right) \\
= \; & \frac{-1}{2}X_{t-1}^4 v_t^{-3} \\
& [c\Delta f'(\omega + \alpha X_{t-1})h_t \\
& +c\Delta f'(\omega + \alpha X_{t-1})(\sigma_D - \sigma_\epsilon)[X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta] \\
& + \frac{-3}{2}v_t^{-2}h_t[X_t - \mu + f(a + bX_{t-1} + c(\omega + \alpha X_{t-1}))\Delta]]
\end{aligned} \tag{6.59}
$$

So in fact we need $|u^3 f_*'(u)| \leq u^4 f_*(u)$ because e.g. the first terms of (6.57) and (6.58) are of the size of $X_{t-1}^4 v_t^{-1}$. But $|u^3 f_*'(u)| \leq u^4 f_*(u)$ follows from Assumption 2 of Proposition 6.9.2. □

Having established these facts we can state:

Theorem 6.9.4. *Let Assumptions 6.4.1 number 1 to 3, 5 and the assumptions of Proposition 6.4.4 hold. if additionally*

1. *f_ϵ, f_D and f are two times continuously differentiable,*
2. *$|uf_*'(u)| \leq const.u^2 f_*(u)$ for $|u| \to \infty$ for $* \in \{\epsilon, D\}$,*
3. *$|uf_*''(u)| \leq const.u^2 f_*(u)$ for $|u| \to \infty$ for $* \in \{\epsilon, D\}$*
4. *$E|X_t|^{4+4\gamma} < \infty$ for a $\gamma > 0$,*

then $\hat{\theta}_n$ is asymptotically normal that means:

$$\sqrt{n}(\hat{\theta}_n - \bar{\theta}_0) = C_n^{-1} D_n \zeta_n + o_p(1) \tag{6.60}$$

with $\zeta_n \to^D N(0, I)$, where I denotes the identity and

$$\sqrt{n} D_n^{-1} C_n (\hat{\theta}_n - \bar{\theta}_0) \to^D N(0, I) \tag{6.61}$$

Furthermore the norms of C_n,D_n and their inverses are $O(1)$.
So $\hat{\theta}_n$ is $n^{\frac{1}{2}}$ consistent for $\bar{\theta}_n$

Proof. Using Propositions 6.9.2 and 6.9.3 we can conclude in analogy to the proof of Theorem 6.8.2. □

6.10 Chiastic complexity

Having investigated as well the pseudo-normal quasi maximum likelihood estimators as well as the mixed distribution ones, it is interesting that the complexity of proof and in a certain sense also the harshness of conditions we have to impose have a kind of chiastic manner: When using the mixed distribution maximum likelihood, the proofs are more straightforward and some conditions weaker if we look at the CARCH-S model rather than in the case of the ACGARCH-V model. When using the pseudo-normal quasi maximum likelihood estimators the conditions are easier to verify in the ACARCH-V model. These differences are to be attributed as well to the difference in modeling σ_t or v_t via the ARCH equation as well as the differences in modeling the mean. Whereas the impact of the former is stronger than the latter.

Chapter 7

Some Approximation Properties

In this short chapter, we give a short discussion on L_1-approximability and near epoch dependence of general GARCH–processes without assuming the innovations to be iid random variables. These approximation concepts are crucial for applying some asymptotic results of Pötscher and Prucha in [PP97].

7.1 Moment Properties

Propositition 7.1.1. *Consider the following model:* $X_t = v_t Z_t$, $E(Z_t|\mathfrak{F}_{t-1}) = 0$, $Var(Z_t|\mathfrak{F}_{t-1}) = 1$, $v_t^2 = \omega + \alpha X_{t-1}^2 + \beta v_{t-1}^2$, *assume* $Ev_0^2 < \infty$ *and that* Z_0 *is given.*
If $\alpha + \beta < 1$ *then:*

1.
$$EX_t = 0$$

2.
$$EX_t^2 = Ev_t^2 = \frac{\omega}{1-(\alpha+\beta)} + (\alpha+\beta)^t \left[Ev_0^2 - \frac{\omega}{1-(\alpha+\beta)} \right]$$

3.
$$EX_t^2 \rightarrow_{t\rightarrow\infty} \frac{\omega}{1-(\alpha+\beta)}$$

Proof. 1. $E(X_t) = E[v_t E(Z_t|\mathfrak{F}_{t-1})] = 0$ as v_t is $\mathfrak{F}_{t-1}$ measurable.

2. a) If Y is $\mathfrak{F}_{t-1}$measurable then $E(Z_t^2 Y) = E[Y E(Z_t^2|\mathfrak{F}_{t-1})] = EY$.
b)$EX_t^2 = Ev_t^2 Z_t^2 = Ev_t^2$ by a).
c)$Ev_t^2 = \omega + \alpha E X_{t-1}^2 + \beta E v_{t-1}^2 = \omega + (\alpha+\beta) E v_{t-1}^2$ by b).

Iterating the argument we get

$$
\begin{aligned}
& Ev_t^2 && (7.1) \\
= \; & \omega + (\alpha+\beta)(\omega + (\alpha+\beta)(\omega + (\alpha+\beta)(\omega + \ldots (\alpha+\beta)Ev_0^2)\ldots) \\
= \; & \omega \sum_{k=0}^{t-1} (\alpha+\beta)^k + (\alpha+\beta)^t Ev_0^2 \\
= \; & \omega \frac{1-(\alpha+\beta)^t}{1-(\alpha+\beta)} + (\alpha+\beta)^t Ev_0^2 \\
= \; & \frac{\omega}{1-(\alpha+\beta)} + (\alpha+\beta)^t \left[Ev_0^2 - \frac{\omega}{1-(\alpha+\beta)} \right]
\end{aligned}
\tag{7.2}
$$

3. Using the form of EX_t^2 shown in 2. and that $(\alpha+\beta)^t \to_{t\to\infty} 0$ yields the result.

□

Corollary 7.1.2. *Let $X_t = \mu_t + v_t Z_t$ and μ_t be $\mathfrak{F}_{t-1}$–measurable, further $v_t^2 = \omega + \alpha(X_{t-1} - \mu_{t-1})^2 + \beta v_{t-1}^2$, $E(Z_t|\mathfrak{F}_{t-1}) = 0$, $Var(Z_t|\mathfrak{F}_{t-1}) = 1$,$Ev_0^2 < \infty$. Let Z_0 be given. Then*

1.

$$Var(X_t) = Ev_t^2 = \frac{\omega}{1-(\alpha+\beta)} + (\alpha+\beta)^t (Ev_0^2 - \frac{\omega}{1-(\alpha+\beta)})$$

2.

$$Var(X_t) \to_{t\to\infty} \frac{\omega}{1-(\alpha+\beta)}$$

Proof. We apply Proposition 7.1.1 to $\tilde{X}_t := X_t - \mu_t$. □

7.2 L_1-Approximability

Propositition 7.2.1. *Consider the model used in Proposition 7.1.1. Then*

1.

$$v_t^2 = \omega \sum_{k=0}^{m} \prod_{i=1}^{k} (\alpha Z_{t-i}^2 + \beta) + \prod_{i=1}^{m+1} (\alpha Z_{t-i}^2 + \beta) v_{t-m-1}^2$$

where $m \le t$ and the empty product is defined as 1.

2. If $\alpha + \beta < 1$ and $Ev_0^2 < \infty$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E \left| v_t^2 - \sum_{k=0}^{m} \prod_{i=0}^{k} (\alpha Z_{t-i}^2 + \beta) \right| \to 0 \text{ as } m \to \infty \tag{7.3}$$

Proof. 1. Iterating we get

$$\begin{aligned}
v_t^2 &= \omega + \alpha X_{t-1}^2 + \beta v_{t-1}^2 \\
&= \omega + (\alpha Z_{t-1}^2 + \beta) v_{t-1}^2 \\
&= \omega + (\alpha Z_{t-1}^2 + \beta)[\omega + (\alpha Z_{t-2}^2 + \beta) v_{t-2}^2] \\
&= \dots \\
&= \omega \sum_{k=0}^{m} \prod_{i=1}^{k} (\alpha Z_{t-i}^2 + \beta) + \prod_{i=1}^{m+1} (\alpha Z_{t-i}^2 + \beta) v_{t-m-1}^2
\end{aligned}$$

2. a) We realize:

$$E\left[\prod_{i=1}^{m+1} (\alpha Z_{t-i}^2 + \beta) v_{t-m-1}^2\right] = E\left[(\alpha Z_{t-1}^2 + \beta) \prod_{i=2}^{m+1} (\alpha Z_{t-i}^2 + \beta) v_{t-m-1}^2\right]$$

$$= (\alpha + \beta) E\left[\prod_{i=2}^{m+1} (\alpha Z_{t-i}^2 + \beta) v_{t-m-1}^2\right]$$

The last equation follows by the argument of the proof of 2.a) of Proposition 7.1.1.

Iterating we get

$$E\left[\prod_{i=1}^{m+1} (\alpha Z_{t-i}^2 + \beta) v_{t-m-1}^2\right] = (\alpha + \beta)^{m+1} E v_{t-m-1}^2.$$

By point 1. of this proposition we get:

$$E\left|v_t^2 - \sum_{k=0}^{m} \prod_{i=0}^{k} (\alpha Z_{t-i}^2 + \beta)\right| = (\alpha + \beta)^{m+1} E v_{t-m-1}^2. \tag{7.4}$$

b) We get

$$\frac{1}{n} \sum_{t=1}^{n} E\left|v_t^2 - \sum_{k=0}^{m} \prod_{i=0}^{k} (\alpha Z_{t-i}^2 + \beta)\right| = (\alpha + \beta)^{m+1} \frac{1}{n} \sum_{t=1}^{n} E v_{t-m-1}^2.$$

Now by the proof of Proposition 7.1.1 the following holds:

$$E v_{t-m-1}^2 = \frac{\omega}{1 - (\alpha + \beta)} + (\alpha + \beta)^{t-1} \left[E v_{-m}^2 - \frac{\omega}{1 - (\alpha + \beta)}\right]$$

$$\Rightarrow \frac{1}{n} \sum_{t=1}^{n} E v_{t-m-1}^2 = \frac{\omega}{1 - (\alpha + \beta)} + \frac{1}{n} \left[E v_{-m}^2 - \frac{\omega}{1 - (\alpha + \beta)}\right] \frac{1 - (\alpha + \beta)^n}{1 - (\alpha + \beta)}$$

It remains to show that $Ev_{-m}^2 < \infty$ for all $m \geq 0$. By assumption $Ev_0^2 < \infty$. Then

$$v_0^2 = \omega + (\alpha Z_{-1}^2 + \beta)v_{-1}^2 \Rightarrow Ev_0^2 = \omega + (\alpha + \beta)Ev_{-1}^2.$$

By part 2a) of the proof of Proposition 7.1.1 which also holds if $EY = \infty$ we can conclude $Ev_{-1}^2 < \infty$. By iterating this argument we get $Ev_{-m}^2 < \infty$ for all $m \geq 0$.

□

Remark 7.2.2. Point 2. of the previous Proposition is the L_1–approximability of v_t^2 by the basis process $\{Z_t\}$ – compare the general definition in section 5.2.

Remark 7.2.3. The analogue of 7.2.1 holds for

$$X_t = \mu_t + v_t Z_t,\ v_t^2 = \omega + \alpha(X_{t-1} - \mu_{t-1})^2 + \beta v_{t-1}^2$$

Just consider like in the corollary of Proposition 7.1.1 $\tilde{X}_t := X_t - \mu_t$.

Remark 7.2.4. Let $X_t = v_t Z_t$, $v_t > 0$ and v_t^2 L_1–approximable by $\{Z_t\}$, then X_t is also L_1–approximable by $\{Z_t\}$.

7.3 Near Epoch Dependence

Definition 7.3.1. *Let $\{X_t\}$ and $\{Z_t\}$ be stochastic processes on $(\Omega, \mathfrak{A}, P)$ then X_t is called L_r near epoch dependent on Z_t if there exist constants $\{\nu_m | m \in N\}$, $\{d_t | t \geq 1\}$ such that $\sup_t ||X_t - E(X_t | Z_{t+m}, \ldots, Z_{t-m})||_r \leq d_t \nu_m$ with $\nu_m \downarrow 0$ for $m \to \infty$.*

The following notation is chosen to coincide with the notation of the previous section although in point 2. of Theorem 7.3.4 the notation σ_t instead of v_t might be more consistent.

Remark 7.3.2. If we want to get L_r near epoch dependence of (v_t, X_t) on Z_t instead of the weaker property of L_1–approximability in the setting of the previous section we can use Theorem 1 of [Han91a], which requires

$$E[(\beta + \alpha Z_t^2)^r | \mathfrak{F}_{t-1})] \leq c^5 < 1 \text{ a.s.}$$

In section 3 of this article it is shown that this condition is implied by $E(Z_t^{2r} | \mathfrak{F}_{t-1}) < (1+\delta)^5$ almost surely for a $\delta > 0$ and additionally $\beta + \alpha(1+\delta) < 1$. By definition of Z_t in this chapter this holds for any $\delta > 0$ for $r = 1$. For $r > 1$ we get a stronger condition on α and β which is also dependent on Z_t.

Assumption 7.3.3. *Let X_t arise from imposing of a GARCH(1,1) dynamic on a double infinite α–mixing process Z_t, i.e.*

$$X_t = v_t Z_t,\ v_t^2 = \omega + \alpha X_{t-1}^2 + \beta v_{t-1}^2.$$

Assume that

$$E[(\beta + \alpha Z_t^2)^r | \mathfrak{F}_{t-1}] \leq c^5 < 1$$

almost surely for all t.

Namely

Theorem 7.3.4. *Let $v_t < \infty$ hold almost surely.*

1. *Suppose 7.3.3 holds for an Z_t with conditional mean 0 and and conditional variance 1, then X_t is L_r–near epoch dependent of size $d_t = \frac{2\omega c}{1-c}$*
2. *Suppose 7.3.3 holds for Z_t without the moment restrictions above, then X_t is L_r–near epoch dependent of size $d_t = \frac{2\omega c}{1-c}$*

Proof. 1. Here we are directly in the situation to apply Theorem 1 of [Han91a].

2. We can also use Theorem 1 of [Han91a], because in the proof the fact that in the setup there Z_t is a martingale difference with unit variance is never used:

$$v_t^2 = \omega \sum_{k=0}^{\infty} \Pi_{i=1}^k (\beta + \alpha Z_{t-i}^2)$$

also holds in our context. To see this note that the calculation

$$\begin{aligned} v_t^2 &= \omega + \alpha X_{t-1}^2 + \beta v_{t-1}^2 \\ &= \omega + (\alpha Z_{t-1}^2 + \beta) v_{t-1}^2 \\ &= \omega \sum_{k=0}^{m} \Pi_{i=1}^k (\beta + \alpha Z_{t-i}^2) + v_{t-m-1}^2 \Pi_{i=1}^{m+1} (\beta + \alpha Z_{t-i}^2) \\ \rightarrow_{m\rightarrow\infty} & \ \omega \sum_{k=0}^{\infty} \Pi_{i=1}^k (\beta + \alpha Z_{t+1-i}^2) \end{aligned}$$

makes no use of the two above mentioned condition. Minkowski's theorem and Blackwell's theorem are also of general nature. The only other theorems used in this proof, namely Theorem 4.2 and Corollary 4.3 b of [GW88] make neither use of any of the above mentioned conditions. So we are free to apply Theorem 1 of [Han91a].

□

The first part of this theorem can be used as an strengthening of the L_1-approximability result above. The second part allows us also to handle the original model. If we can assure in the general models that the innovation process η_t is mixing we also can get NED for them. Trying to prove this property when $p_t = f(X_{t-1}, v_t)$ we can get stuck in a circular argument. The general setup is too dynamic in a self referential way. One way out is the dependence of p_t on external variables only, another will be introduced in the next chapter.

Chapter 8

Some results concerning asymptotics and inference in a restricted model with real GARCH dynamic

8.1 A restricted model and its approximability by mixing processes

We are working with the alternative model, that is the ACGARCH context, using the notation of sections 7.1 and 7.2. That means we consider:

$$X_t = v_t Z_t$$

with $Z_t = \frac{\eta_t}{h_t}$, $\eta_t = (1 - B_t)\epsilon_t + B_t D_t$, $E\epsilon_t = ED_t = 0$, $\mathcal{L}(B_t|\mathfrak{F}_{t-1}) = \mathfrak{B}(1, p_t)$ and further $h_t^2 := E(\eta_t^2|\mathfrak{F}_{t-1})$.

Remark 8.1.1. In the previous models we had $p_t = f(X_{t-1}, v_t)$ but in the case of a true GARCH dynamic, we run even in problems when using $p_t = f(X_{t-1})$, because $\{X_t\}$ is not automatically mixing but only L_1–approximable by $\{Z_t\}$. But $\{Z_t\}$ is dependent on $\{X_t\}$ via p_t. If we take $p_t = f(Z_{t-1})$ we run also in problems: Because of

$$Z_t = \frac{\eta_t}{h_t} = \frac{\eta_t}{((1 - p_t)\sigma_\epsilon^2 + p_t\sigma_D^2)^{\frac{1}{2}}}$$

Z_{t-1} depends on p_{t-1} so in fact then $p_t = f(\eta_{t-1}, p_{t-1})$. So we have an ARMA like structure again and the standard theorems are not applicable. So we restrict ourselves to a simple case:

$$p_t = f(\eta_{t-1}) \tag{8.1}$$

Then $\{\eta_t\}$ is a Markov chain with transition density

$$p(x|y) = (1 - f(y))f_\epsilon(x) + f(y)f_D(x) \tag{8.2}$$

Theorem 8.1.2. *Let f, f_ϵ, f_D be continuous. Suppose further one of the following conditions holds:*

1. *The support of f_ϵ and the support of f_D is $\mathbb{R}$.*
2. *The support of f_ϵ is $\mathbb{R}$ and $\forall t : p_t \le 1 - \delta < 1$.*
3. *$supp(f_\epsilon) \cup supp(f_D) = \mathbb{R}$ and $\forall t : 0 < \gamma \le p_t \le 1 - \delta < 1$.*

Then $\{\eta_t\}$ is geometric ergodic.

Proof. Analogous to the Lemma 4.5.5 the continuity of f, f_ϵ, f_D imply the Feller property for $\{\eta_t\}$. Like in the discussion of models with pure ARCH dynamic in chapter 4 either of conditions 1.–3. imply that $\{\eta_t\}$ is irreducible and aperiodic. Further for $\gamma(y) = 1 + y^2$ the following holds:

$$\frac{E(\gamma(\eta_t)|X_{t-1} = y) - \gamma(y)}{\gamma(y)} = \frac{(1 - f(y))\sigma_\epsilon^2 + f(y)\sigma_D^2 - y^2}{1 + y^2} \to_{y^2 \to \infty} -1$$

as f is bounded. Therefore condition 4.2.8 is satisfied. $\square$

Corollary 8.1.3. *$\{p_t\}$ and $\{Z_t\}$ are geometric ergodic.*

Proof. By 8.1 $p_t = f(\eta_{t-1})$ is a function of η_{t-1}. And

$$Z_t = \frac{\eta_t}{(1 - p_t)\sigma_\epsilon + p_t\sigma_D}$$

is a function of η_t and p_t. But functions of finitely many elements of a time series preserve the mixing properties of this time series. $\square$

Remark 8.1.4. By the previous chapter X_t and v_t are L_1 approximable by $\{Z_t\}$ which is α–mixing. So we try to use the methods of [PP97] to gain asymptotics of the (quasi)log-likelihood–estimator.

Corollary 8.1.5. *Under the conditions of Proposition 7.2.1 and $p_t = f(\eta_{t-1})$ the following statements hold:*

1. $$\frac{1}{N}\sum_{t=1}^{N} v_t^2 \to_p \frac{\omega}{1 - (\alpha + \beta)}$$

2. $$\frac{1}{N}\sum_{t=1}^{N} X_t^2 \to_p \frac{\omega}{1 - (\alpha + \beta)}$$

3.
$$\frac{1}{N}\sum_{t=1}^{N} X_t \to_p 0$$

Proof. a) As Z_t is geometric ergodic so is $S_t^m := \sum_{k=0}^{m} \Pi_{i=1}^{k}(\alpha Z_{t-i}^2 + \beta)$ for every fixed m. So S_t^m satisfies a law of large numbers for all m.
b) Now by Proposition 7.2.1 v_t^2 is L_1–approximable by $\{Z_t\}$ and therefore also $X_t = v_t Z_t$. Using $EX_t^2 = Ev_t^2 \to \frac{\omega}{1-(\alpha+\beta)}$ for $t \to \infty$ as provided by Proposition 7.1.1 and $EX_t = 0$ for all t, then the assertion follows from a) and Theorem 6.2.a) of [PP97]. □

The arguments here should be generalizable to $p_t = f(\eta_{t-1}, B_{t-1})$,where the Markov process would be $(\eta_t, B_t)^T$, or to $p_t = f(\eta_{t-1}, \dots, \eta_{t-d})$ with Markov process $(\eta_t, \dots, \eta_{t-d+1})^T$.

Then following theorem gives some stationarity results. Because we can get results for the alternative model from section 3.4 immediately by the same technique we give a short definition to let it fit into the setting here.

Definition 8.1.6. *Let the general setting of this section hold we define $\mu_t = \mu + p_t\Delta$ for constants μ and Δ and further*
$$Y_t := \mu_t + X_t = \mu + p_t\Delta + v_t Z_t$$

Theorem 8.1.7. *Let additionally to the assumptions of Theorem 8.1.2*
$$E\log(\alpha Z_t^2 + \beta) < 0$$
hold.

1. *there exists a strictly stationary and ergodic version of v_t*
2. *there exist strictly stationary and ergodic versions of X_t and Y_t.*

Proof.
1. Theorem 8.1.2 yields a strongly stationary and ergodic version of $\{\eta_t\}$. This yields strongly stationary and ergodic versions of Z_t. Like in the proof of Lemma 2 in [LH94] we can use Theorem 3.5.7 of [Sto74] in order to gain the result of Theorem 2 of [Nel90] in the case of strongly stationary and ergodic Z_t. This yields a strongly stationary and ergodic version of v_t.
2. Having established point 1. we use Proposition 4.3 of [Kre85] with $f(v_t, Z_t) = v_t Z_t$ which yields strong stationarity and ergodicity of $X_t = v_t Z_t$. A strongly stationary and ergodic version of p_t yields strong stationarity and ergodicity of μ_t. Again by Proposition 4.3 of [Kre85] applied to $f(\mu_t, v_t Z_t) = \mu_t + v_t Z_t$ shows that $Y_t = \mu_t + v_t Z_t$ has a strongly stationary and ergodic version.

. □

8.2 Steps towards asymptotic theory of estimation

Lemma 8.2.1. *If $v_t^2 = \omega + \alpha X_{t-1}^2 + \beta v_{t-1}^2$, then v_t^2 is a function of ω,α, β given the observations $\{X_t\}$ and the starting value v_0^2. Explicitly*

$$v_t^2 = \sum_{i=0}^{t-1} \beta^i(\omega + \alpha X_{t-i-1}^2) + \beta^t v_0^2 \tag{8.3}$$

Proof. Induction:
If we define the empty sum being zero, the lemma holds for $t = 0$.
$t \to t+1$

$$\begin{aligned}
& v_{t+1}^2 \\
= \;& \omega + \alpha X_t^2 + \beta v_t^2 \\
= \;& \omega + \alpha X_t^2 + \beta \left[\sum_{i=0}^{t-1} \beta^i(\omega + \alpha X_{t-i-1}^2) + \beta^t v_0^2\right] \text{ by Induction} \\
= \;& \omega + \alpha X_t^2 + \sum_{i=1}^{t} \beta^i(\omega + \alpha X_{t-i-1}^2) + \beta^{t+1} v_0^2 \\
= \;& \sum_{i=0}^{t} \beta^i(\omega + \alpha X_{t-i-1}^2) + \beta^{t+1} v_0^2
\end{aligned}$$

□

Lemma 8.2.2. *If $v_t^2 = \omega + \alpha X_{t-1}^2 + \beta v_{t-1}^2$ and $p_t = f(\eta_{t-1}, \overline{\gamma})$ where $\overline{\gamma}$ are the parameters of f.Then η_t is a function of ω, α, β and $\overline{\gamma}$ given the observations $\{X_t\}$ and the starting values v_0^2 and η_0*

Proof. Induction. For η_0 the statement holds trivially.
$t-1 \to t$

$$\begin{aligned}
Z_t = \frac{X_t}{v_t} \Rightarrow \eta_t = \frac{X_t}{v_t} h_t &= \frac{X_t}{v_t}\sqrt{(1-p_t)\sigma_\epsilon^2 + p_t\sigma_D^2} \\
&= \frac{X_t}{v_t}\sqrt{\sigma_\epsilon + f(\eta_{t-1}, \overline{\gamma})(\sigma_D^2 - \sigma_\epsilon^2)}
\end{aligned} \tag{8.4}$$

By Lemma 8.2.1 v_t is a function of ω,α, β given the observations $\{X_t\}$ and the starting value v_0^2, and by induction η_{t-1} is a function of η_t is a function of ω, α, β and $\overline{\gamma}$ given the observations $\{X_t\}$ and the starting values v_0^2 and η_0 σ_ϵ^2 and σ_D^2 are either given constants or we have to treat them as nuisance parameters.

□

Having established these features we can write the log-likelihood in terms of the data and parameters.

Remark 8.2.3. We could now proceed to use the methodology of [PP97] to establish consistency, if we would know that we can establish a nice metric on the image of $(X_i|i \in \mathbb{Z}_{<t})$, where $\mathbb{Z}_{<t}$ denotes the whole numbers smaller than t. Then we could use the methodology for likelihood functions depending on data of NON–fixed lag-length propose on page 75 of [PP97]. The problem is to find an appropriate metric with respect to which σ_t and p_t are continuous functions of the infinite past. If we assumed that the projections of the image of $(X_i|i \in \mathbb{Z}_{<t})$ are bounded we could handle at least σ_t. But due to the highly nonlinear nature of p_t as a function of the data that wouldn't help much.
If we would get the continuity of σ_t and p_t then the summands of the log likelihood function would be continuous. Then we could proceed like in the ARCH case to establish the rest of the conditions.

Chapter 9

Estimation and simulation in the CGARCH-S and the CGARCH-V model

Now we want to switch to the question, how do the models work in practice.

9.1 The models we used and corresponding notation

We want to specify the models and notation we use in this and the following chapters. The GARCH dynamic is always GARCH(1,1) except in section 11.2. Generally we will use the notation from appendix A. In tables we will not state te prefix CGARCH. So e.g. CGARCH-SP will be denoted just SP. For all models using

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2$$

we will speak of models of CGARCH-S type. In fact the models abbreviated with S, SB, SP, SX are of CGARCH-S type. When using

$$v_t^2 = \omega + \alpha X_{t-1}^2 + \beta v_{t-1}^2$$

we will speak of models of CGARCH-V type. In all models we assumed that $\{\epsilon_t\}$ and $\{D_t\}$ are both iid and independent from each other. Furthermore we always use Normal distributed $\{\epsilon_t\}$ and Lognormal distributed $\{D_t\}$. Unless otherwise stated the distribution of $\{\epsilon_t\}$ is standard normal. For D_t we use the lognormal(0,1) and the lognormal(1,1) distributions and always state which we use. Further we always use a standard logistic function to model p_t. In the original model we use

$$p_t = \frac{1}{1 + \exp(-[a + bX_{t-1} + c\sigma_t])}. \tag{9.1}$$

In the original model with signum induced crash probabilities we use

$$p_t = \frac{1}{1 + \exp(-[a + b\,\mathrm{sign}(X_{t-1})])}.$$

When using additional external variables

$$p_t = \frac{1}{1 + \exp(-[a + bX_{t-1} + c\sigma_t + dMA_{t-1} + e\sqrt{WA_{t-1}}])} \tag{9.2}$$

was used, where the exogenous processes $\{MA_t\}$ and $\{WA_t\}$ are defined in the beginning of Chapter 10. We use $\sqrt{WA_{t-1}}$ because this made the numerical calculation faster and more stable than using WA_{t-1}. In the model with crash probability depending just on external variables

$$p_t = \frac{1}{1 + \exp(-[a + dMA_{t-1} + e\sqrt{WA_{t-1}}])}$$

was modeled. The modeling of the crash probabilities in the volamodel with and without additional external variables was done by substituting σ_t by v_t in equations (9.1) and (9.2). When using standard normal ϵ_ts we use in tables the abbreviations without the prefix CGARCH followed by the parameters of the lognormal distribution we use to model the crashes. So SX(0,1) means a CGARCH-SX model with standard normal ϵ_t and lognormal(0,1) distributed D_t. When using a normal distribution with mean 0.0001 we write e.g. S(0,1)0.0001. When using an additional constraint we use cS(0,1)0.00001. Talking of the underlying variables of the dynamic of p_t we will e.g. write "purely external induced crashes". If D_t has the distribution NN we will also use the notation NN based crashes.
In all cases we use Matlab routines to minimize the conditional log-likelihood function directly. A trial to develop a kind of EM algorithm yielded the practical result of not converging.

9.2 Estimation and simulation of models without external variables

We fitted several of the models we introduced to BASF data from 1990 to 1992. We present here the financial data used in most of the practical part. The raw material are the DAX values and the BASF prices of the time period 1990–1992. These were transformed to log returns. Due to the fact that we derive from the DAX some variables, which we want to interpret as external variables in some of the models we investigate, we focus on the BASF as the data that we try to estimate. In the following we fit various models to these data and use the fitted models to generate artificial data. The parameters of these fitted models are given in Table 10.1

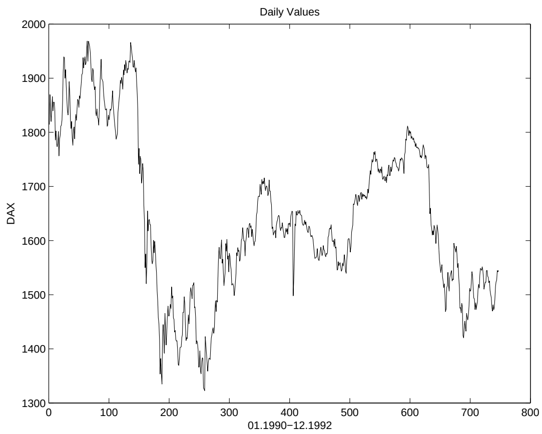


Figure 9.1: The DAX

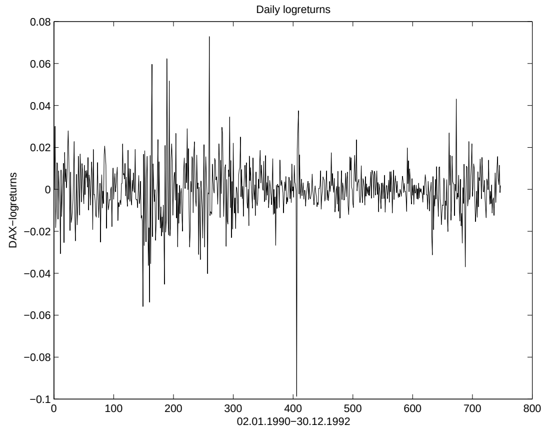


Figure 9.2: DAX log–returns

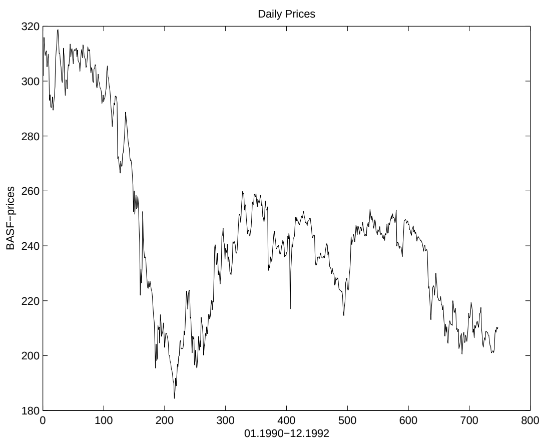


Figure 9.3: BASF Prices

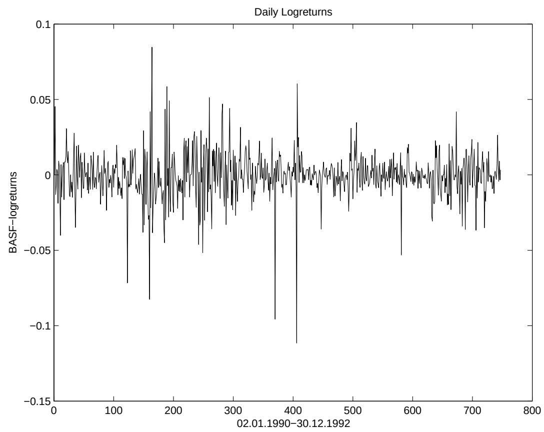


Figure 9.4: BASF log–returns

9.3 A first study

We want to see, what happens when we estimate the BASF–log returns, then fit a GARCH model, a CGARCH-S model and a CGARCH-V model and then simulate. Due to the fact that nothing like a crash probability exists in the pure GARCH case our benchmark will be GARCH estimates of the volatility of the original series compared to the volatilities of the simulations of the different crash

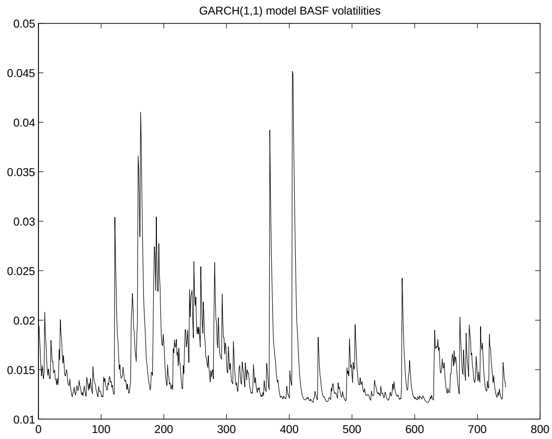


Figure 9.5: The sigma=volatility of the fitted GARCH(1,1) model

models. The question being: Does the simulated volatility expose similar features in range and shape as the volatility estimate of the input series. Between the two models with a probability of a crash we also ask the question if the simulation does show some behavior, that could be interpreted as a crash.

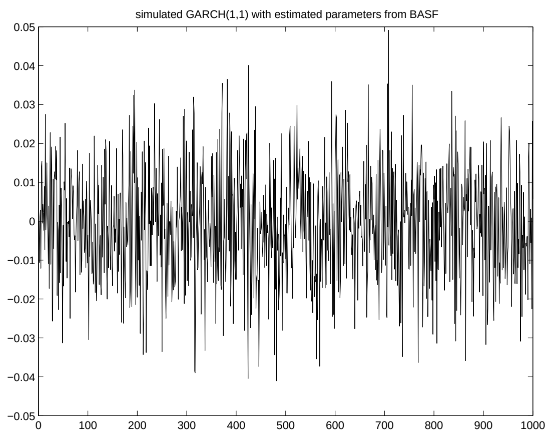


Figure 9.6: GARCH(1,1) model simulation

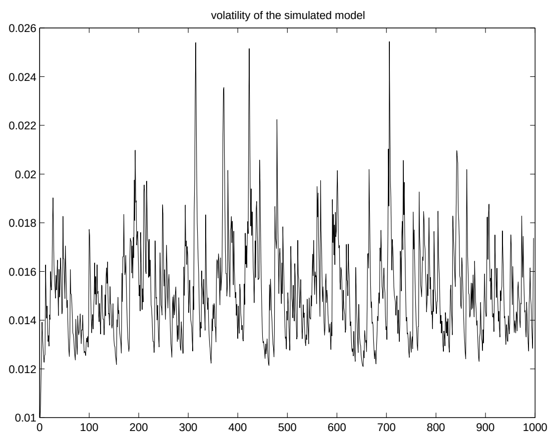


Figure 9.7: The estimated volatility in the simulated GARCH(1,1) model

Now we look at the simulation of a estimated CGARCH-S model with crashes which are based on a lognormal(0,1) distribution. Actually the path of the simulation doesn't seem to expose a more extreme nature than the simulation of the GARCH model. When looking at the volatilities we see that the peaks of the volatility in this model are even smaller than the ones in the GARCH models. The model is far from exposing such extreme volatility peaks like the example in Figure 2.4. Like we will see later choosing a more extreme crash distribution, namely lognormal(1,1), will make simulated models more extreme, but in a way that isn't realistic, because it is too extreme.

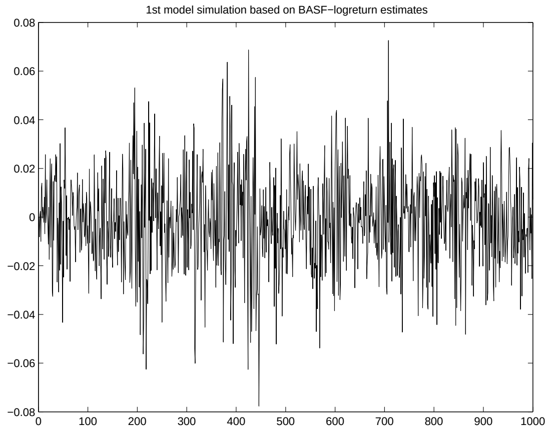


Figure 9.8: The CGARCH-S model simulation

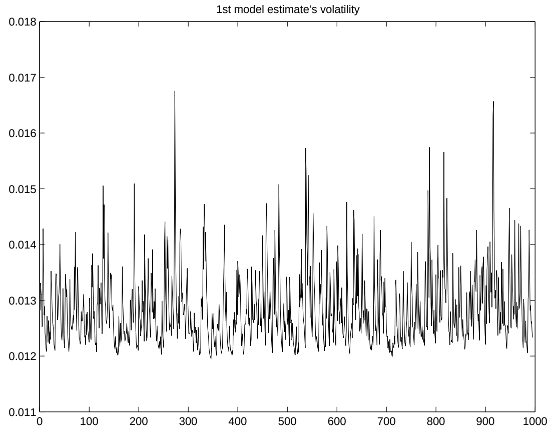


Figure 9.9: The volatility of this simulated model

When we choose a CGARCH-V model with lognormal(1,1) based crashes, we get really events which look like crashes, and which are still in an realistic range. So from the simulation side this model seems to be preferable. Also the volatility has peaks reaching into the regions of the volatility of the GARCH model of the original time series. But still, the overall picture of the volatility seems to be more quiet than the estimate of the original series. As we will see later this model will not be necessarily better than the CGARCH-S model, when it comes to model the quantiles.

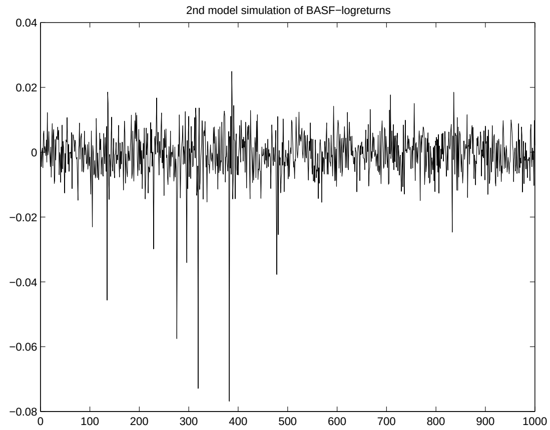


Figure 9.10: The CGARCH-V model simulation

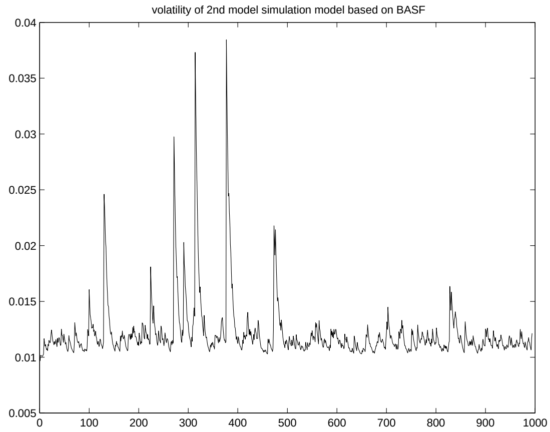


Figure 9.11: The volatility of the data in Figure 9.10

9.4 Some models of CGARCH-S type

We now investigate the different outcome of the CGARCH-S model, especially concerning the question of using constraints or not. Besides that we look what happens to the crash probabilities when altering the mean of the non-crash probability.

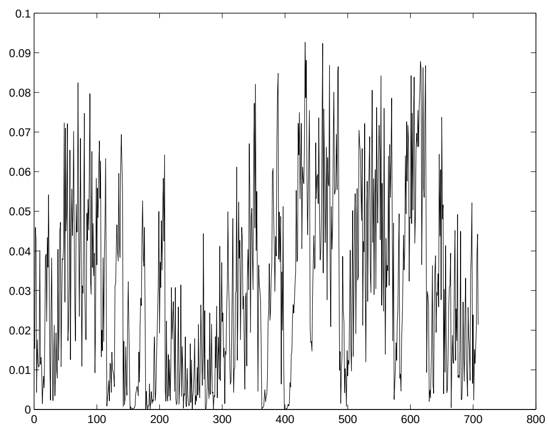


Figure 9.12: CGARCH-S models crash probability with zero mean ϵ_t

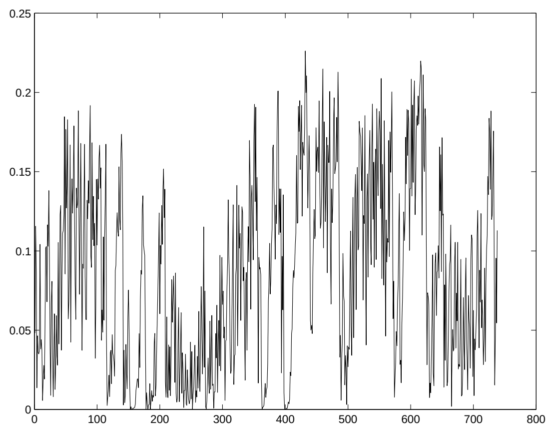


Figure 9.13: CGARCH-S models crash probability with mean 0.0001 for ϵ_t

Here we used instead of a $N(0,1)$ distribution for ϵ_t a $N(0.0001,1)$ distribution. In Figures 9.12 and 9.13, we see that this leads to a very similar shaped graph

of the crash probability. But the difference in the range of both is strong enough that, if simulating, there will be considerably more crashes. In fact we chose 0.0001 as a mean for the non-crash distribution, as a constant, which adding to the paths of simulations of the N(0,1) non-crash distribution model made most of them have a slightly positive mean. But using the new distribution in estimating, led to a negative mean in all simulated model paths. It seems that it might be, even in the model depending purely on external variables in its crash probability, hard if not impossible to find distributions for ϵ_t and D_t such that the model really describes something that satisfy modeling a crash in the sense of “big losses in relatively short time” and doesn’t predict certain bankruptcy in the near future. While in the context of the CGARCH-S model we will stick in this chapter to the non-crash probability having a $N(0.0001, 1)$ basis. This is due to the fact that firstly we already discussed aspects of the zero mean model before and that concerning the question of using a constraint or not is treated here in a setting with higher crash probabilities. As there is no considerable difference to the zero mean model when applied to the real world data, the comparability to models where we used zero mean non-crash distributions is retained.
First we look at the σ_t and the volatility. Although using the model with higher crash probabilities, we see that the difference between the two is rather small, distinguishable by pure eyesight only as a more noisy nature of the volatility. The overall shape and range is quite similar to the volatility of the pure GARCH process in 9.5.

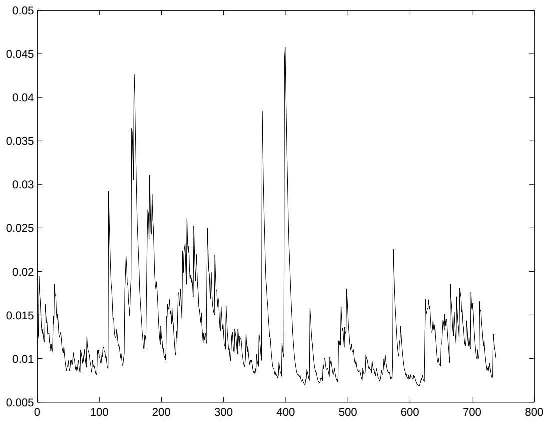


Figure 9.14: CGARCH-S model’s sigma

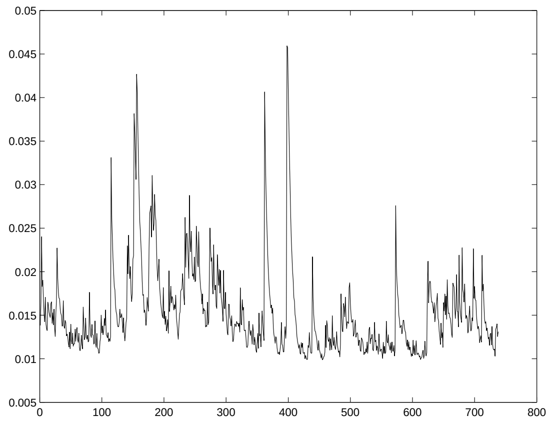


Figure 9.15: CGARCH-S model’s volatility

If we use now a constrained model in the sense of Proposition 2.2.7, we end up not only having smaller crash probabilities, but also with an altered shape of the process p_t – compare figure 9.13 with figure 9.16.

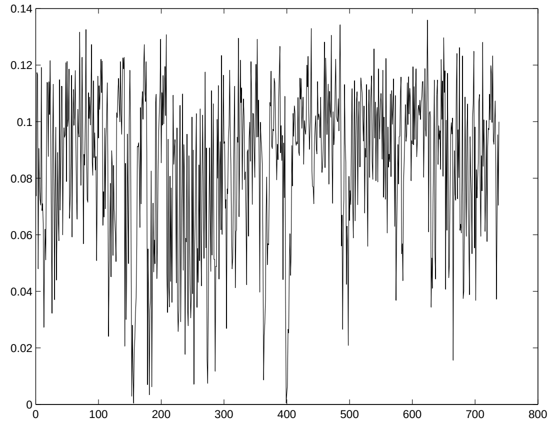


Figure 9.16: The constrained models crash probability

There is now an obvious difference between the σ_t and the volatility, the latter being extremely noisy. In comparison to the pure GARCH model the estimated σ_t and volatility are rather small. But this is no wonder: The parameters governing the dynamic of σ_t is where the constraint is imposed. The strong constraint gives us a model really underestimating $ARCH$–effects.

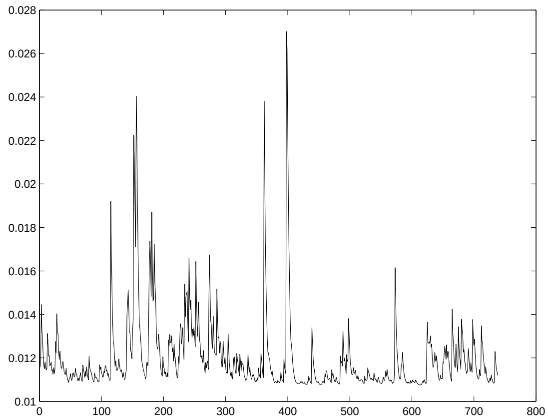


Figure 9.17: The constrained model's sigma

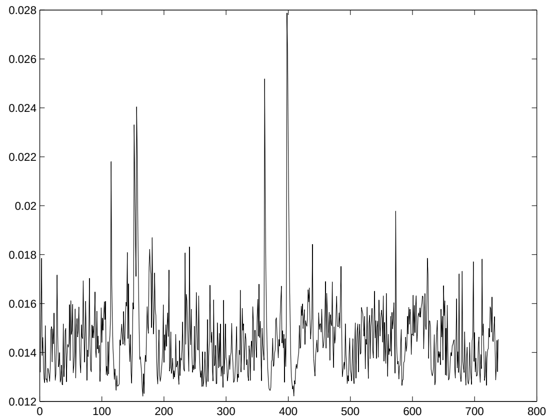


Figure 9.18: The constrained model's volatility

The question is: Do we end up in an explosive situation, not using the strong constraint? This fear is not supported by simulation evidence. In Figure 9.19 we see one typical path of the simulated non constrained model.
Having the discussion in section 9.3 in mind we see that altering the mean of the non-crash distribution leads to an situation, where there are more extreme losses in the simulation, which could be interpreted as crashes.

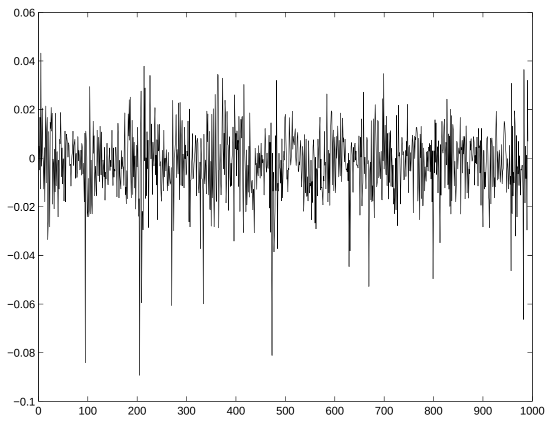


Figure 9.19: The simulated CGARCH-S model

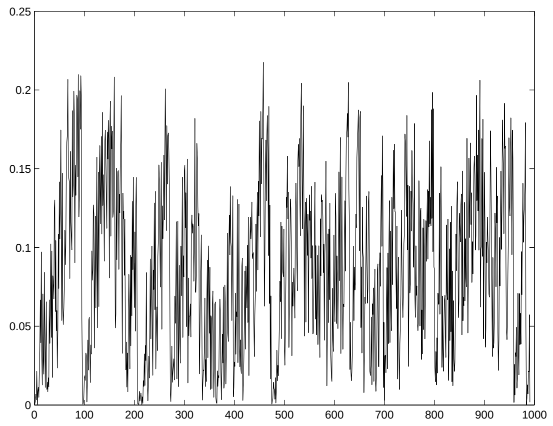


Figure 9.20: The simulated CGARCH-S models crash probability

Again the shape of the process $\{\sigma_t\}$ and the volatility have a similar shape. There is no evidence that we were in danger of ending in an explosive situation. The peaks of the volatility are still smaller than the peaks of the pure GARCH estimate of the volatility of the real data.

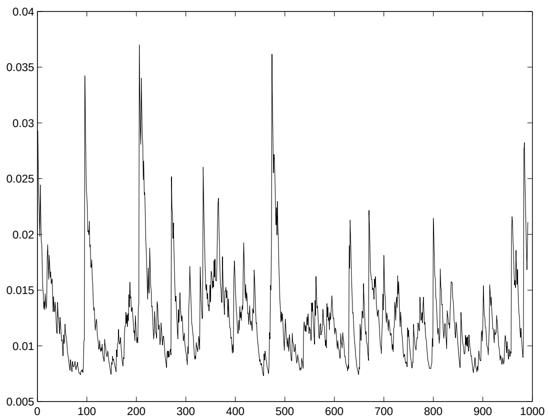


Figure 9.21: The simulated model's sigma

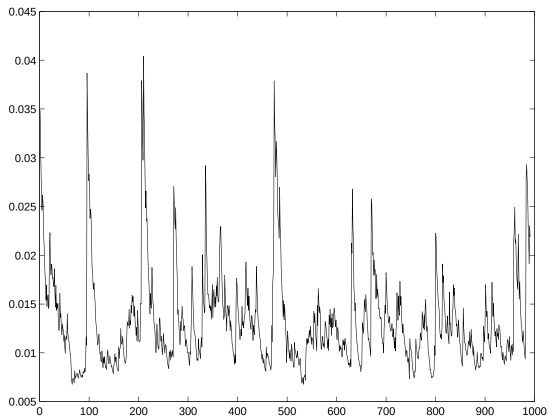


Figure 9.22: The simulated model's volatility

Now we switch to a model with more extreme crashes: We used the lognormal distribution with parameters (1,1) as the distribution for D_t. Due to the fact that the constraint from Proposition 2.2.7 would make vanish the GARCH effect nearly completely we didn't use it, trading in the possibility of an explosive model. Simulation shows, that this seems not to be a problem. As we can see in Figure

9.23 in comparison to Figure 9.12 assuming more extreme crashes brings a lower crash probability.

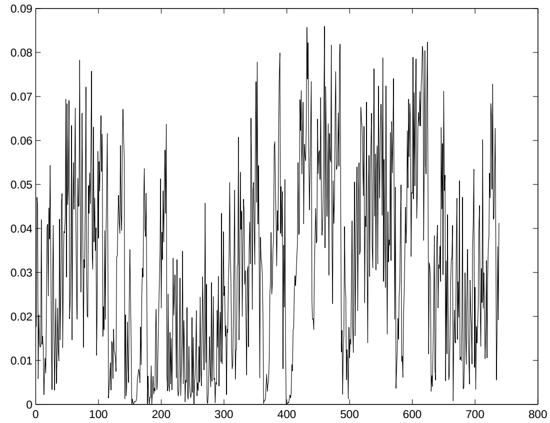


Figure 9.23: CGARCH-S model's crash probability with lognormal(1,1) crashes

But Figures 9.14 and 9.24 show that the differences for σ_t are negligible. This is something not so surprising, but comparing Figures 9.15 and 9.25 show that the volatilities of both models are also very close. So the choice of the crash distribution doesn't seem to affect the volatility related part of the model too much while using the real world data.

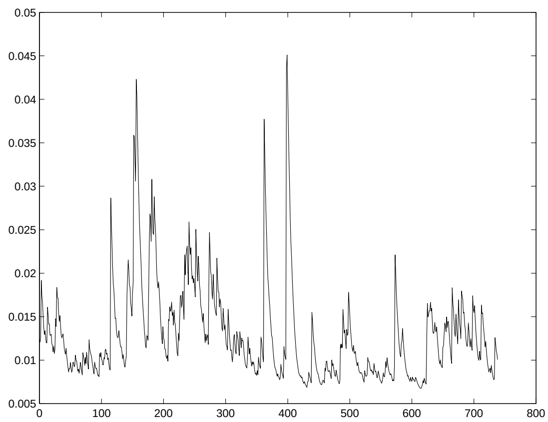


Figure 9.24: This model's sigma

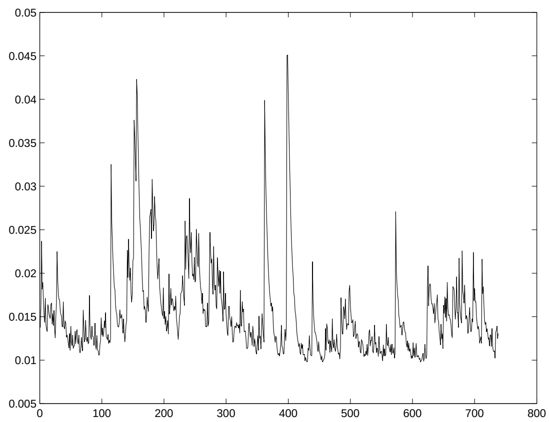


Figure 9.25: This model's volatility

Switching to simulation we state again the question, whether we have ended in an explosive situation. Firstly the simulated time-series doesn't seem to be very realistic for financial returns, having exordinary high and low "returns". But we are far from being in an explosive situation: The simulated returns are not explosive. And the crash probabilities are even in the range of the crash probabilities in the real data fit, although the simulated “return” time-series is of a far more extreme nature. The fact that we avoided explosivity can be explained by the parameters b and c both being negative. So the effect of extremely negative returns on the crash probability is self neutralizing.

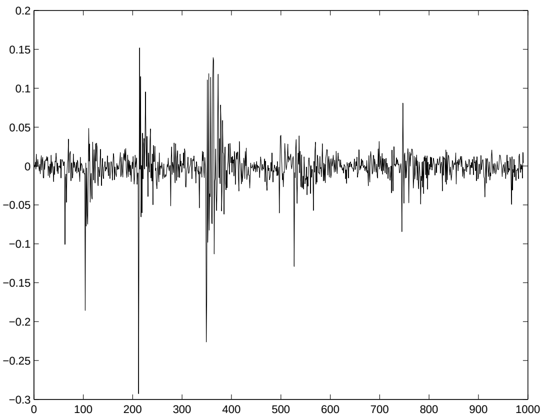


Figure 9.26: The simulated CGARCH-S model wit logn(1,1) crashes

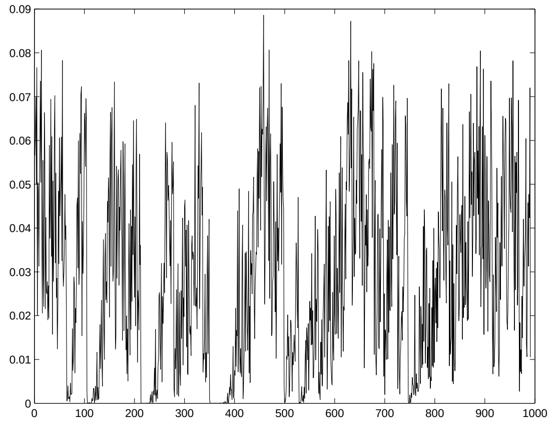


Figure 9.27: The simulated model’s crash probability

Secondly we have exordinary high volatilities, but they come down rather quickly.

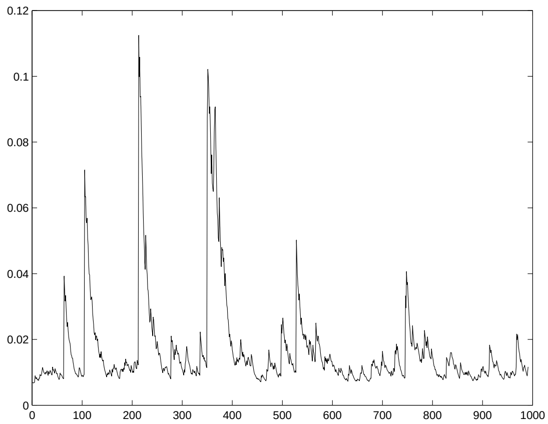


Figure 9.28: The simulated model’s sigma

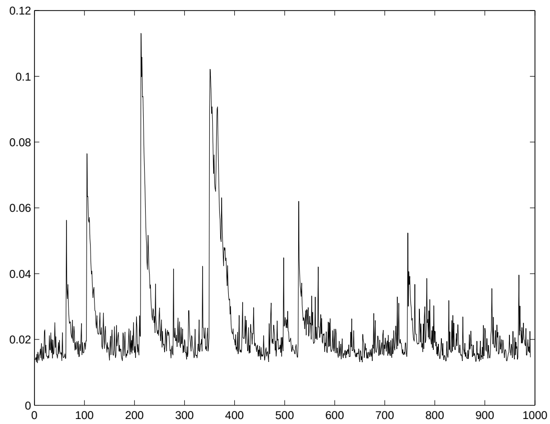


Figure 9.29: The simulated model’s volatility

So the simulation evidence is, that we don’t need a constraint stronger than $\alpha + \beta < 1$ to avoid estimating a model which is explosive, when using real financial data.

The findings in the models with non zero mean non-crash distributions are similar to the corresponding results in the zero mean case. Both the crash probabilities and the σ_t's are smaller there. So the evidence of the need to use constraints is also smaller. Again the crash probabilities in the simulations have a similar range, shape and mean as the real data fits and more extreme crashes yield a model with smaller crash probability, as we should demand from a model to make sense at all.

9.5 The CGARCH-V model

Now we switch to the basic CGARCH-V model with crashes based on the log-normal distribution with parameters (1,1). Looking at the crash probabilities the most obvious feature in comparison to the original model is that there are periods, e.g. around time step 250, with very low crash probabilities an no considerable "noise" for a small time interval.

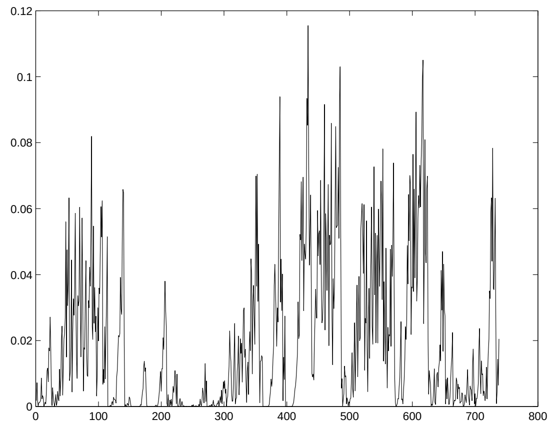


Figure 9.30: CGARCH-V model's crash probability with lognormal(1,1) crashes

We observe that the mechanism which produces σ_t from the volatilities is such that that the difference between the two time series is considerable. But the location and size of peaks of the volatility and the sigma are nearly identical under this mechanism. The overall shape of the graph of the volatility process is similar to the one in the pure GARCH case, however the peaks are considerably smaller. For simulation results concerning the volamodel we refer to section 9.3.

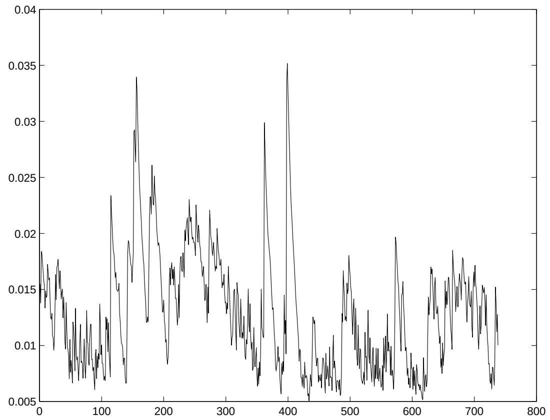


Figure 9.31: The CGARCH-V model's σ

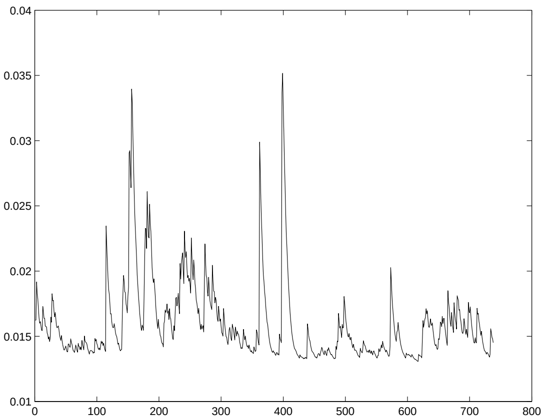


Figure 9.32: CGARCH-V model's volatility

Chapter 10

Estimation in models with additional external parameters in the crash probability

In addition to the data provided above, we used a DAX–log return moving average and a weighted average of the squared log returns of the DAX as external parameters to investigate BASF returns. The first time series is a measure of local market trend, the latter of squared market volatility.

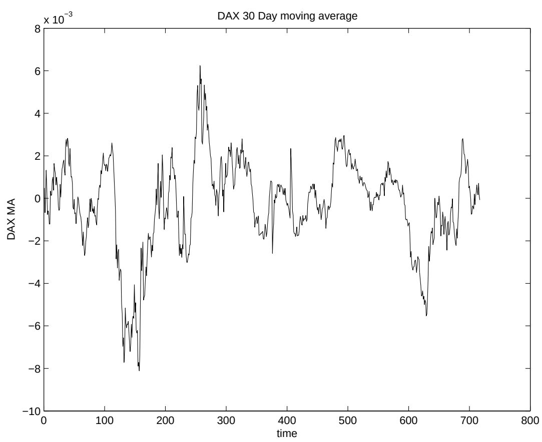


Figure 10.1: The DAX moving average

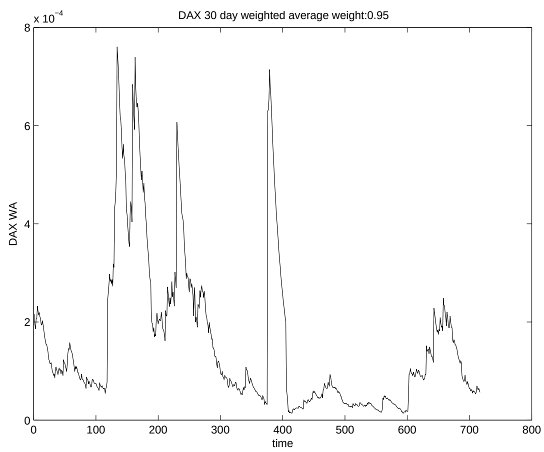


Figure 10.2: The DAX weighted average

10.1 A threshold model

This is a kind of model we didn't treat before, because we could not find any good theoretical treatment. But in practice it seems to give at least some qualitative insights and we have an easy way to control the meaning of a crash. This model was inspired by the use of quantitatively not specified crashes in [Kor01] and [KW02]. As an alternative we model the behavior of the process in quiet times as an GARCH–process and define the down crossing of a threshold as a crash. If we take such a threshold e.g. being -0.1 we get:

$$X_t \begin{cases} = \sigma_t \epsilon_t \text{ with probability } 1 - p_t \\ < -0.1 \text{ with probability } p_t \end{cases}$$

This model is incomplete. It is sufficient to calculate σ_t and p_t in a backwards oriented setup. However if we want to simulate in the exact sense or use the value at risk as a risk measure, we would have to make assumptions on what the distribution is, when the threshold is exceeded. When coming to estimation a similar problem is present. As we don't define any density when a crash occurs, we can only use the qualitative modeling of the crashes in a function which is a likelihood only in the non crash case and just an indicator in the crash case. We will study the outcome of such a model in 10.1.
Now we give a short survey of an estimated threshold model. We use internal and external variables for the model. The model with additional external variables was chosen, because it yielded the seemingly best results, while the core problem of the threshold models are still present. We took any return smaller or equal to 0.09 to be a crash. As an objective function for estimation we used

$$\sum_{t=1}^{n} \log \left[(1 - p_t) \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{X_t^2}{2\sigma_t^2} \right) \mathbb{I}(x \geq 0.09) + p_t \mathbb{I}(x < 0.09) \right]$$

Due to the fact that no crash-distribution is specified the objective function maximized is not to be called a likelihood function. While we cannot compare this model with the other models presented here on a Value at Risk based benchmark, we can still investigate the properties of its modeled volatilities. Whereas the volatilities are very alike the pure GARCH one (no surprise: When no crash occurs we basically fit a GARCH(1,1) model and we chose the threshold such that only two crashes occurred) the crash-probabilities are fairly low except for jumps up in the two cases where actually the crashes occur.
Because the threshold model is underspecified it is not clear how to simulate according to this model. One way out is to use the value of the threshold as "crash value." It might occur that the GARCH part of the model yields values smaller than the threshold. But interpreting our model so that we state that when a crash occurs it doesn't matter from which distribution it is coming, this does not have to be interpreted as a mistake of the model. The high peaks near the occurence of the defined crashes might suggest that this model is able to predict crashes. But the peaks occur in both cases after the crashes occured. Having in mind, that due to definition we can see from the data that a crash occurs, just at the time it does occur, this is not a new information. So this model isn't good

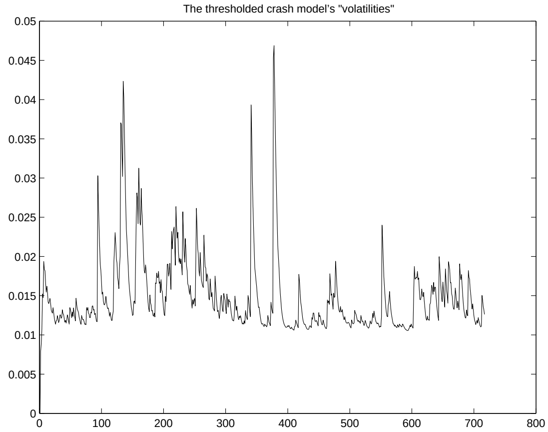


Figure 10.3: Volatility in the threshold model

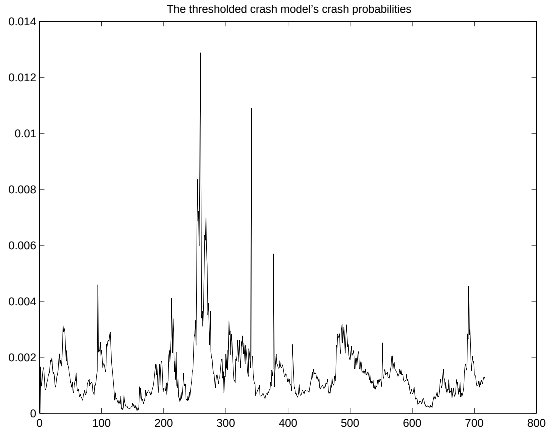


Figure 10.4: Crash-probability in the threshold model

for crash prediction. On the other hand we are not able to define quantitave risk measures like the Value at Risk in a natural way in this model. So it seems that the use of crashes without an assumed law is not very helpful for practical matters in this context.

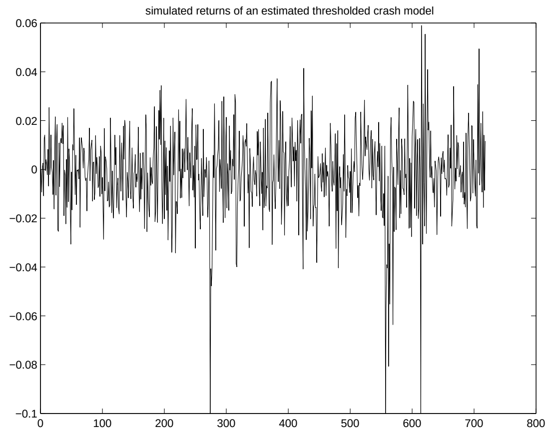


Figure 10.5: Simulated threshold model with estimated coeefficents

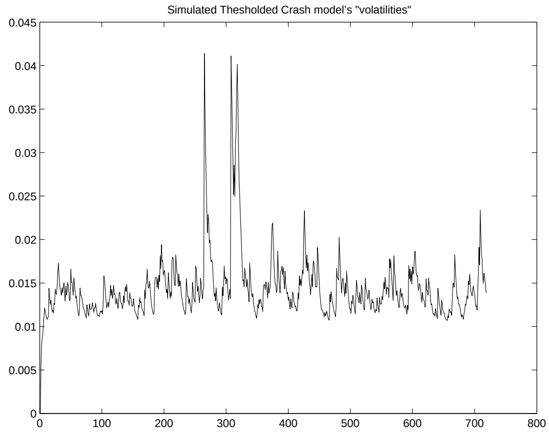


Figure 10.6: The volatility of the simulated threshold model

The crash-probabilities of the "simulation" are much smaller than the one arising from real world data. While this might indicate a model which isn't very good, there still is the relatively big difference between low and high crash probabilities.

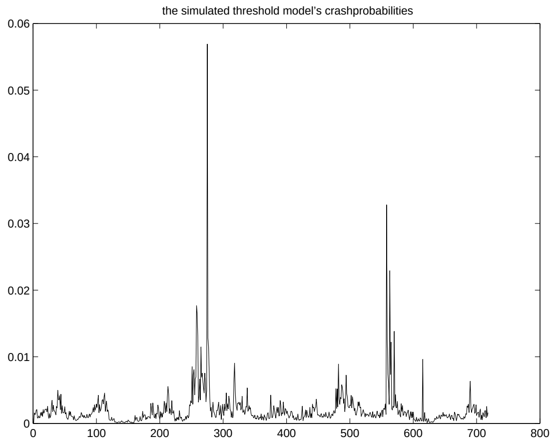


Figure 10.7: The crash probability of the simulated threshold model

In fact I used this model mainly tho check if in the models with specified crash distributions these distributions did not lead to overly high crash probabilities in quiet time periods.

10.2 Models with additional external variables

In the following we investigate the outcome of the introduction of additional external variables in the CGARCH-S and the CGARCH-V model. For the CGARCH-S model we switch back to the zero mean non-crash distribution. In the CGARCH-S model with additional external variables the introduction of the latter leads to a less blurred distinction between time intervals with relatively high and relatively low crash probabilities. In addition there appear to be three short periods where the crash probability reaches a magnitude not present in the model without the external variables.

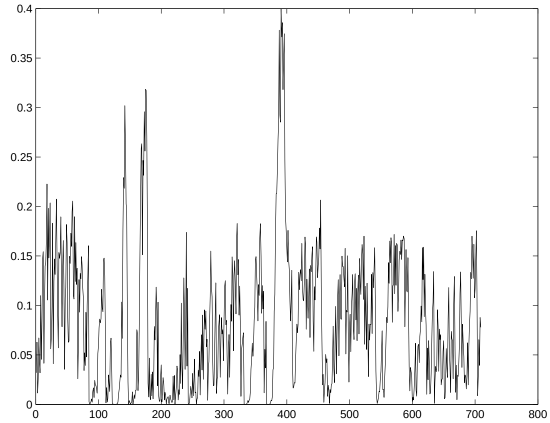


Figure 10.8: CGARCH-SX model's crash probability with lognormal(0,1) crashes

We see in Figure 9.30 that adding the dependence on external variables in the crash probability doesn't lead to a strong smoothing or trend effect in the estimated model's crash probability . It isn't a big improvement to the model without the additional external variables. The only obvious effect is the slight deformation (first going down more quickly and then staying higher for some time) of the highest peak (indicating the attempted military coup in Russia).

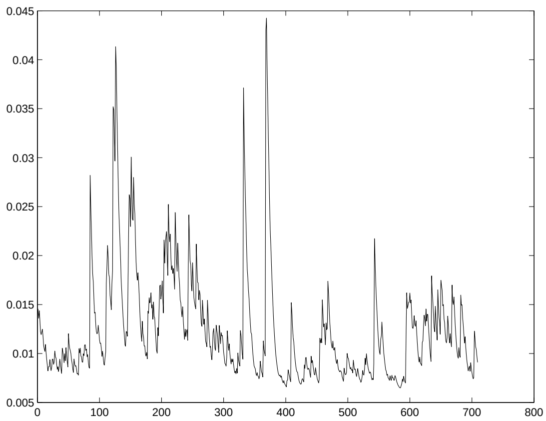


Figure 10.9: OX model's sigma

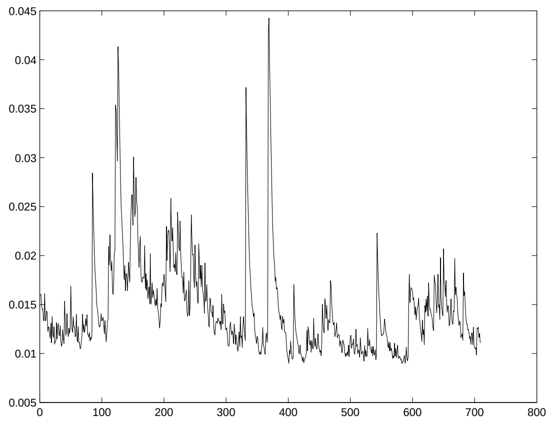


Figure 10.10: OX model's volatility

Switching to the CGARCH-V model with additional external variables leads to a completely different picture. The graph of the crash probabilities looks completely different from the one without external variables. Beside isolated extreme peaks the graph exposes an even more rigid distinction between periods with near zero and higher crash probabilities than the model without external variables. In Figure 10.11 we see the astonishing effect that the introduction of the external variables leads to periods where the crash probability is almost 0 on the one hand side, but with extremely high peaks on the other hand. At least we get by the introduction of external variables something like a trend shape in the crash probabilities. Interesting is that in both models containing self reference and external dependence in the crash probability we get different signs in the estimate of c and f the first referring to the processes and the second referring to the markets volatility. It might be naught to contain both dynamics in one model, because it seems that the estimation leads to an effect of both at least partially neutralizing each other.

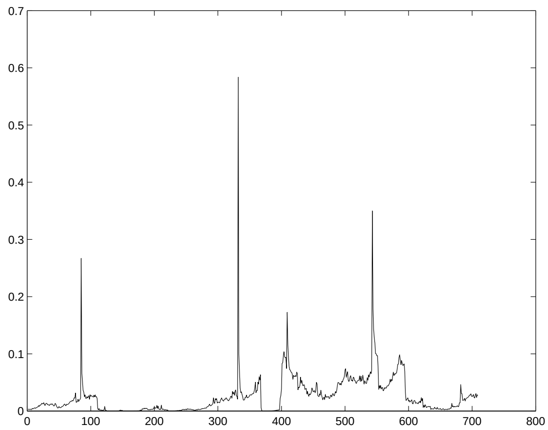


Figure 10.11: CGARCH-VX model's crash probability with lognormal(1,1) crashes

We see in 10.12 that the extreme differences in p_t leads to a strong deformation of the graph of σ_t arising from the volatility. Due to the highest peak of the crash probability, the response of σ_t to the BASF specific extreme event around time step 340 is very weak

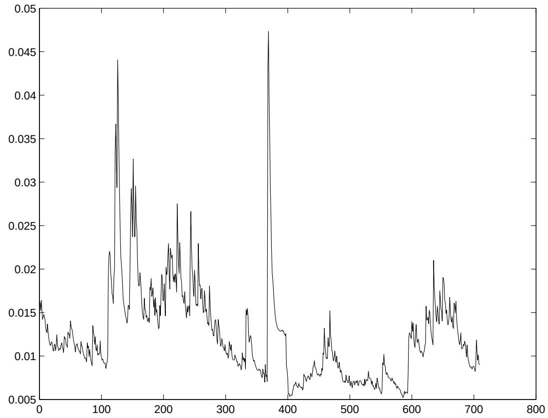


Figure 10.12: VolatiltityX model's sigma

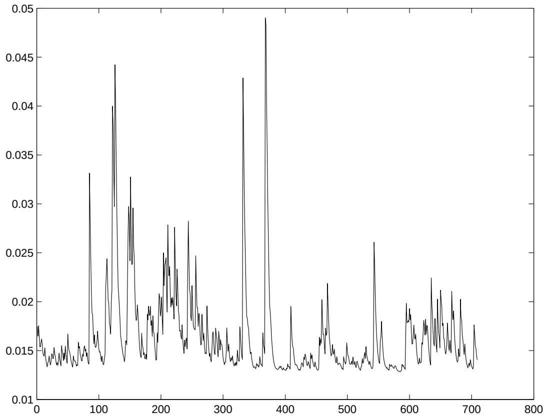


Figure 10.13: CGARCH-VX model's volatility

10.3 Using just external variables

Here we fitted to the BASF data a model with crash probabilities depending on the external variables MA and WA exclusively. We modeled: $\mathcal{L}(\epsilon_t) =_D N(0,1)$, $\mathcal{L}(D_t) =_D LOGN(1,1)$, $p_t = \frac{1}{1+\exp(-[a+dMA_t+e\sqrt{WA_t}])}$. As can be seen in Figure 10.14 and Table 10.1 we got the result that c is negative. That means, that our crash probability is bigger, if the volatility of the market, described by WA is small. On the other hand we get the effect we expect from the estimated value of d: If the market is on a downwards trend, the probability of big losses is bigger. It is not clear, if the negativity of e is due to the fact that small values of WA often occur when σ_t is small and so the conditional crash distribution is shifted towards a region where the true density of X_t is bigger or if it is a real economic effect. By using just the external variables for modeling the crash probabilities, we get a much smoother path for $\{p_t\}$ than in the models using internal variables. Using the chosen quite smooth external variables exclusively is the reason for this effect.

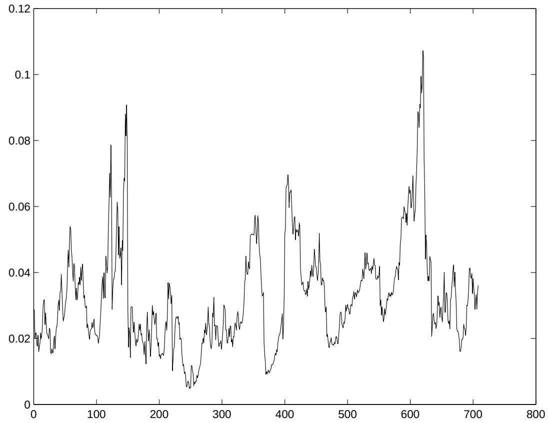


Figure 10.14: CGARCH-SP model's crash probability with lognormal(1,1) crashes

We see the modeled $\{\sigma_t\}$ in Figure 10.15 and the calculated volatility of the model in Figure 10.16. Surprisingly the graph of σ_t is more similar to the graph of the CGARCH-S model with lognormal(0,1) crashes than the lognormal (1,1) crashes. This model shows the most extreme values of the conditional volatility of all models investigated here. And this maximum lies in the gulf war period rather than being induced by the russian event which happened later.

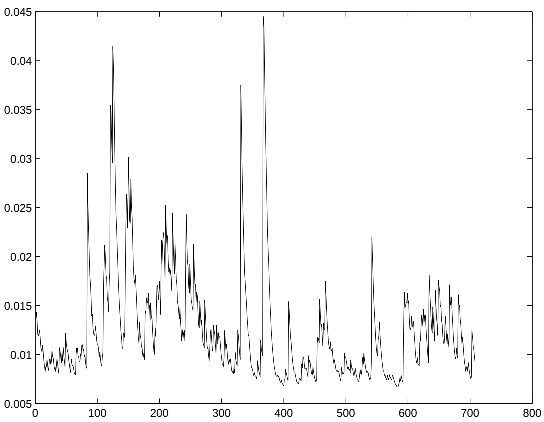


Figure 10.15: CGARCH-SP model's sigma

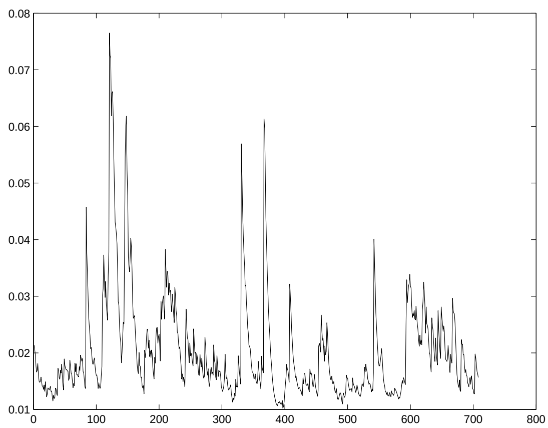


Figure 10.16: CGARCH-SP model's volatility

The same model with lognormal(0,1) based crashes shows a graph of the crash probability, which is almost a linear transformation than the graph shown above. It' maximum is 0.199. The graph of its volatility and σ_t gives no new insights.

Where the graph of σ_t is very similar to the one of the model above the volatility is comparable to Figure 9.15. So we don't present these graphs here.

Table 10.1: Estimates of parameters of the BASF returns 1990–1992 in several models

Model	ω	α	β	a	b	c	d	e
S(0,1)0.0001	9.4934×10^{-6}	0.1521	0.77183	0.91445	−49.925	−229.75	–	–
S(0,1)	9.4770×10^{-6}	0.1527	0.77237	0.08385	−50.61	−231.00	–	–
cS(0,1)0.0001	4.145×10^{-5}	0.048777	0.63958	0.46912	−26.486	−246.14	–	–
cS(0,1)	4.1689×10^{-5}	0.048764	0.63968	1.3303	−33.134	−323.57	–	–
S(1,1)0.0001	8.6489×10^{-6}	0.14612	0.78328	−1.3147	−51.96	−199.22	–	–
S(1,1)	4.635×10^{-6}	0.099959	0.88079	−1.4284	−8.7013	−243.44	–	–
cS(1,1)0.0001	2.2918×10^{-4}	0.018316	0	−90	−49.431	−231.18	–	–
cS(1,1)	1.797×10^{-4}	0.011351	0.38028	−74.984	−45.378	−241.00	–	–
SB(0,1)	1.0786×10^{-5}	0.14460	0.76838	−2.5805	−0.53182	–	–	–
SB(1,1)	1.0088×10^{-5}	0.14494	0.77338	−3.6126	−0.52116	–	–	–
SX(0,1)	8.3613×10^{-6}	0.14205	0.7746	0.25089	−49.881	−379.11	−281.5	118.16
SP(0,1)	1.0218×10^{-5}	0.14602	0.75875	−2.0484	–	–	−191.50	−43.243
SP(1,1)	9.3191×10^{-6}	0.14488	0.76644	−2.7543	–	–	−242.52	−73.047
V(1,1)	3.1136×10^{-5}	0.077453	0.81351	11.592	−66.653	−1080.8	–	–
V(0,1)	1.0153×10^{-5}	0.14772	0.75786	−1.9225	−34.682	−75.514	–	–
VX(1,1)	5.5966×10^{-5}	0.18384	0.65736	−2.2625	−11.319	104.11	91.435	−360.69

Table 10.2: Estimates of parameters of the BASF returns 1990–1992 in genuine GARCH models

Model	ω	α	β	mean
GARCH	4.0590×10^{-5}	0.17703	0.66828	−0.0010531
GARCHt4	3.7457×10^{-5}	0.16713	0.69666	−0.0010175

10.4 Estimation of simulated models

Due to the fact that the conditions for consistency of the CMLE–estimators for the models with genuine GARCH dynamic are unknown we made a numerical experiment, estimating the known parameters of a simulation of this model. Therefore we created 500 paths of length 717 following such a model, all with the same parameters. The parameters were chosen to be the ones we estimated from the BASF–returns. Then we imposed our estimation procedure on these series. First we look at the outcome for a model depending only on external variables. The estimates for the GARCH parameters ω,α and β showed a little leptokurtic distribution. Their sample mean was near the real parameters. The parameters belonging to p_t showed a different behavior. Firstly the majority of the estimates showed nearly no spread, such that the median of the estimates were nearly the true parameters. Secondly there were also estimate clusters at two points bigger and smaller than the real parameters. And thirdly there were extreme outliers. So if these effects are not due to a bad numerical implementation, or to short time-series there is some doubt, that consistency properties for the particular model are met. Secondly the limiting distribution, provided it exists, seems to be not normal. But that could be due to short time-series.

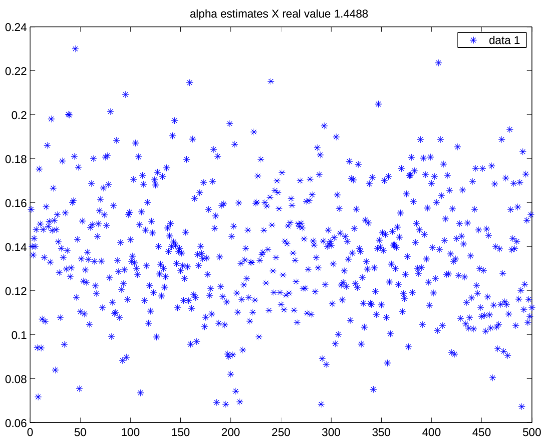


Figure 10.17: Estimates of the parameter α (X_{t-1}^2) of the CGARCH-SP model

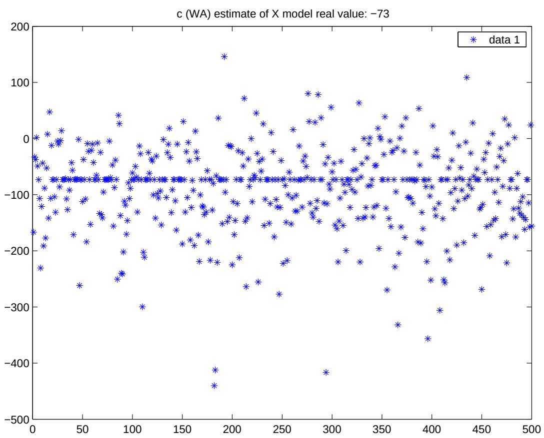


Figure 10.18: Estimates of the parameter e ($\sqrt{WA_t}$) of the CGARCH-SP model

We took the same approach also for the CGARCH-S model with and without the strong constraint 2.2.7 on the parameter space and for the CGARCH-V model,all three without external variables. For the CGARCH-V model the outcome was quite similar to the outcome in the CGARCH-SP model. In both cases of the genuine CGARCH-S model the GARCH parameters were also biased. In the strongly constrained model there were estimates at the border of the parameter space.

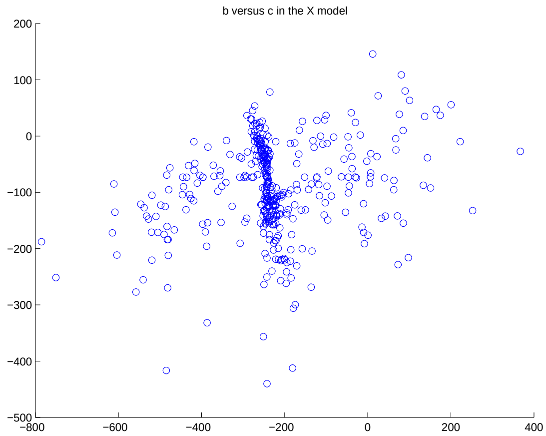


Figure 10.19: Scattering c versus d estimates of the CGARCH-SP model

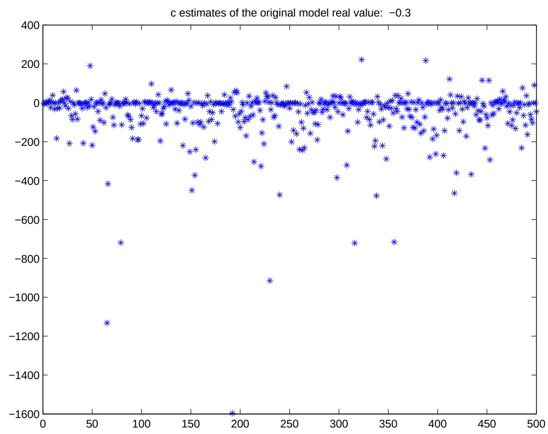


Figure 10.20: Estimates of the parameter c (σ_t) of the original model

Chapter 11

Practical results concerning the Value at Risk of real world data

For the notation used in this chapter we refer to section 9.1.

11.1 Comparison of the different models via Value at Risk

The standard risk measure used in the GARCH context is the conditional "Value at Risk" for the level γ.

$$VaR_t(\gamma) := \inf\{x | F_t(x) \geq \gamma\}$$

Here, F_t is the conditional distribution function of X_t given the past up to time $t-1$. In the standard GARCH model you get $VaR_t(\gamma)$ by multiplying the γ–quantile of the standard normal distribution with σ_t. In the models with crashes drawn from specified distributions we have to calculate the quantiles of the calculated "mixed distribution" at every time step . In the threshold model a VaR is not calculable. Because we did not find any natural extension of this concept to the threshold model, it is skipped in the following survey. We calculated the 1% and 5% conditional value at risk of some models in the classes we defined fitted to the BASF returns from 1990-1992. We chose the non–crash distribution to be $N(0,1)$ except in one case where we chose it to be $N(0.0001, 1)$. The crash distribution was chosen to be lognormal, either with parameters (0,1) or (1,1). In order to get the γ%–VaR accurately we have to consider that inverting the mixed cumulative density function is a nonlinear problem. The linear attempt using

$$(1-p_t)\sigma_t q_\epsilon^\gamma + p_t\sigma_t q_D^\gamma \tag{11.1}$$

with q_ϵ^γ and q_D^γ denoting the $(1-\gamma)$-quantiles of ϵ_t, D_t respectively, will lead to systematic inaccuracies when p_t is not near 0 or 1. We used the theoretically more accurate way of minimizing

$$\left(\left((1-p_t)\int_{-\infty}^{x}\frac{1}{\sigma_t}f_\epsilon\left(\frac{y}{\sigma_t}\right)dy + p_t\int_{-\infty}^{x}\frac{1}{\sigma_t}f_D\left(\frac{y}{\sigma_t}\right)\right)-\gamma\right)^2 \tag{11.2}$$

With the distributions we chose we get:

$$\begin{aligned}\Bigg(\Bigg((1-p_t)\int_{-\infty}^{x}\frac{1}{\sigma_t}\phi\left(\frac{y-\mu}{\sigma_t}\right)dy \\ +p_t\int_{-\infty}^{x}\frac{1}{ys\sqrt{2\pi}}\exp\left(\frac{1}{2s^2}(\log(-y)-m-\log(\sigma_t))^2\right)dy\Bigg)-\gamma\Bigg)^2.\end{aligned} \tag{11.3}$$

In comparison to the linear attempt this led to no big improvements on the 5% level in all cases and the 1% level when using a not too extreme crash distribution. But at the 1% level in models with extreme crash distributions the nonlinear effects really seemed to be of importance, which we were able to capture. We show the number of exeedances in table 11.1 and the corresponding ratios in 11.2. The constrained cases showed an almost constant VaR. Besides these models, all crash models seem to be more accurate than the pure GARCH model. The question is if these "good" results are due to good modeling. As we have seen in the theoretical part of this investigations the theoretical properties of all these models are not clear. Moreover the practical investigations of the paths of the simulated models suggests that at least for the models with lognormal(1,1) based crashes the sizes of crashes are overestimated. So the good VaR performance is due to integration effects like in the models where stable distributions are directly fitted to time series [RSK01].

The models with lognormal(0,1) induced crashes may mirror the real tail behavior better, having more mass in the upper tails, but it doesn't match the goal to model something we would interpret as crashes. In all cases, even if we take models with positive mean for the non-crash part, seem to predict certain bankruptcy meaning the models seeming to have negative mean. In the case with purely negative crashes and zero mean non-crashes this is clear, In the case with positive mean crashes we can theoretically check this fact only in the model with the crash probability just depending on some transform of the signum function. But the evidence in simulation suggests this fact for all models.

Here we give the 1% and 5% Value at risk of a GARCH(1,1) model with normal innovations as a benchmark for the Value at Risk for the different new models we investigated.

Model	5% VaR ex	1 % VaR ex
GARCH	28	16
GARCHt_4	31	9
S(0,1)	32	11
cS(0,1)	40	8
S(1,1)	29	12
cS(1,1)	27	8
S(0,1)0.0001	33	10
SB(0,1)	31	11
SB(1,1)	31	7
SX(0,1)	32	10
SP(0,1)	32	11
SP(1,1)	32	7
V(1,1)	38	14
V(0,1)	34	12
VX(1,1)	31	11

Table 11.1: VaR exceedances of BASF 1990–1992

Model	5%	1%
GARCH	0.0391	0.0223
GARCHt_4	0.0432	0.0126
S(0,1)	0.0446	0.0153
cS(0,1)	0.0558	0.0112
S(1,1)	0.0404	0.0167
cS(1,1)	0.0377	0.0112
S(0,1)0.0001	0.0460	0.0134
SB(0,1)	0.0418	0.0153
SB(1,1)	0.0418	0.0098
SX(0,1)	0.0446	0.0134
SP(0,1)	0.0446	0.0153
SP(1,1)	0.0446	0.0084
V(1,1)	0.0509	0.0187
V(0,1)	0.0475	0.0167
VX(1,1)	0.0432	0.0153

Table 11.2: exceedance ratios of BASF 1990–1992

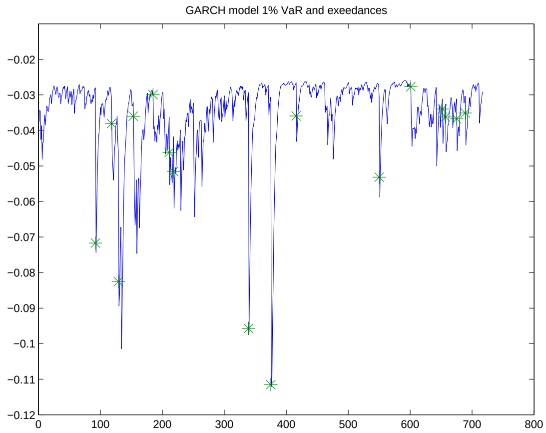


Figure 11.1: 1 % Value at Risk and exceedances of a pure GARCH model

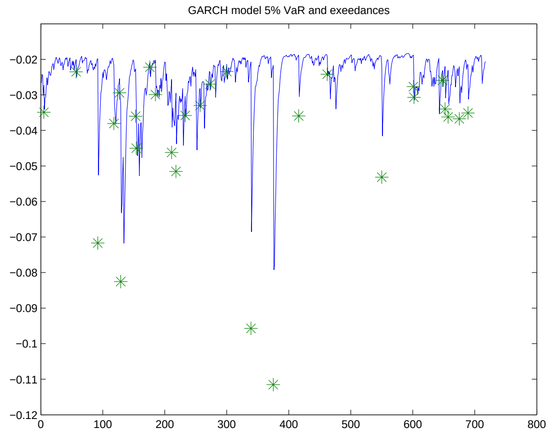


Figure 11.2: 5 % Value at Risk and exceedances of a pure GARCH model

We now investigate the results concerning the Value at Risk of the CGARCH-SP model where the crash probabilities depend just on external variables. First of all we see that the crash coming from the coup d'etat in Russia is an exceedance in all models. This fact holds as we will see for all investigated models. In the pure optical impression the VaR's for different levels in the model with crashes which are based on the negative of a lognormal distribution with parameters zero and one don't seem to differ very much from the GARCH VaRs. But the GARCH effect of this model is slightly smaller than the one of the pure GARCH model, so on the 5% level, where the non-crash distribution already has a dominant role, we get more exceedances like in the pure GARCH case. The fact that the opposite is true in the case of the 1% level is that in comparison the quantiles of the crash distribution are so big that they lever the quantiles of the overall distribution.

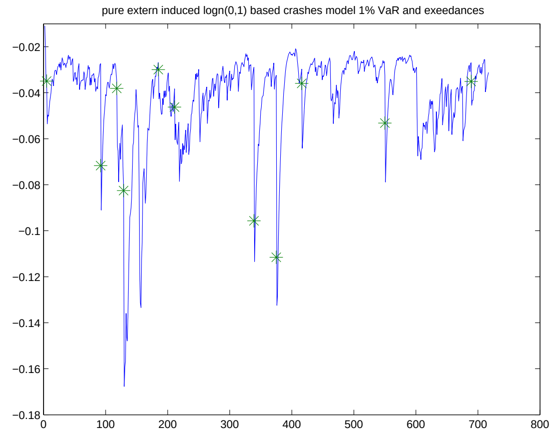


Figure 11.3: 1 % Value at Risk and exceedances of a CGARCH-SP model with lognormal(0,1) based crashes

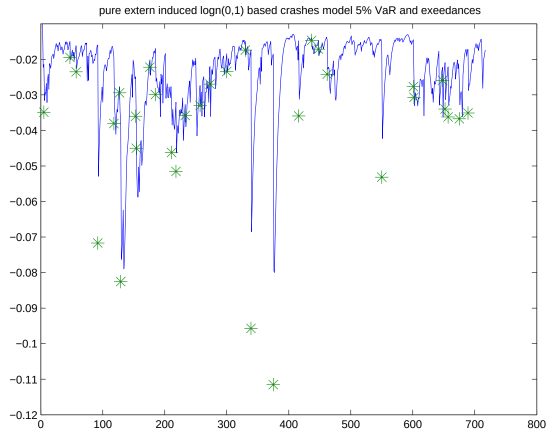


Figure 11.4: 5 % Value at Risk and exceedances of a CGARCH-SP model with lognormal(0,1) based crashes

The following pictures show the result if we base the crashes of a CGARCH-SP model on a lognormal variable with parameters $(1, 1)$. The most obvious feature of this model is that the 1% VaR of this model becomes very negative in the time period between 100 and 200 being the period of the first gulf war and being a time period of a very volatile market. My explanation for this is, the external variables being derived from the DAX capture the market insecurities and the model was able to mirror this via a high crash probability in this period. So using a crash distribution with very extreme quantiles mirrors this fact in the overall model, such that we get only one exceedance in this period. Besides that this model seems to be the best performing on the 1% level, the differences to the preceeding model on the 5% level aren't very big.

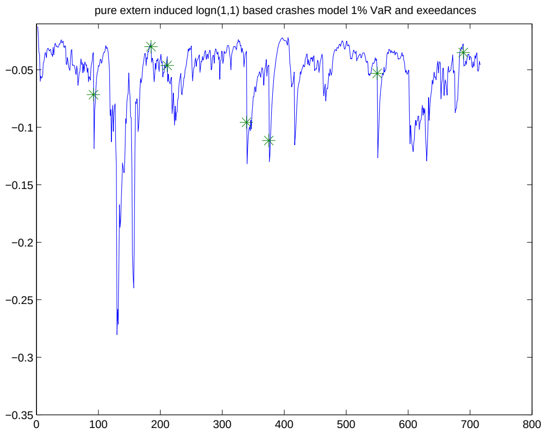


Figure 11.5: 1 % Value at Risk and exceedances of a CGARCH-SP model with lognormal(1,1) based crashes

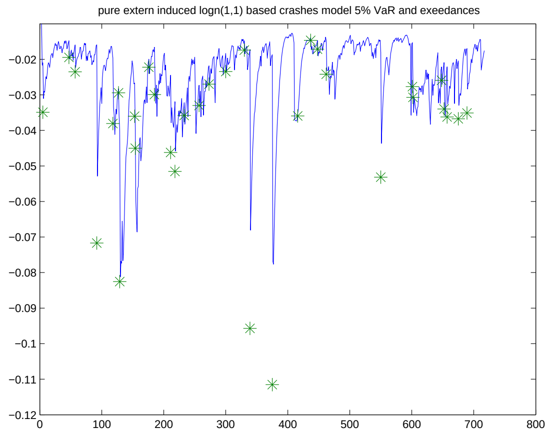
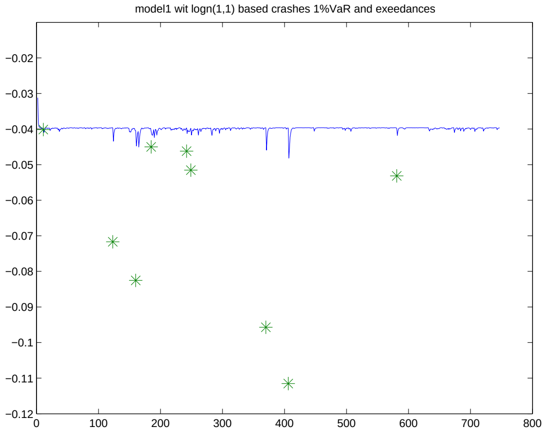


Figure 11.6: 5 % Value at Risk and exceedances of a CGARCH-SP model with lognormal(1,1) based crashes

Now we switch to different incarnations of the CGARCH-S model. First the question whether to introduce the easy to implement but in fact very strong constraints from 2.2.7 is settled from a point of view concerning a meaningful VaR in the way that this constraint should be dropped. The reason mentioned at the beginning of this chapter is illustrated below. Actually the dynamic in the model with lognormal (1,1) based crashes is very poor and irrelevant concerning the exceedances. The dynamic in the model with lognormal (0,1) is stronger, but still very small setting the GARCH(1,1) model as a benchmark. We see in Figure 11.8 that even on the 5% level the dynamic is nearly irrelevant.



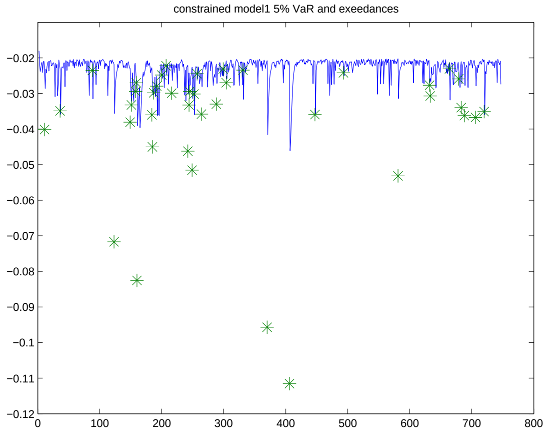
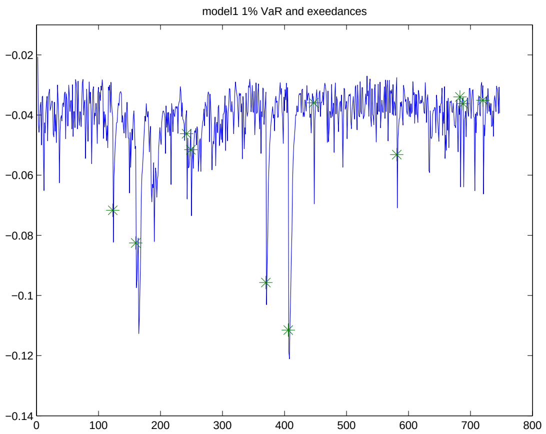


Figure 11.7: 1 % Value at Risk and exceedances of a constrained CGARCH-S model with lognormal(1,1) based crashes

Figure 11.8: 5 % Value at Risk and exceedances of a constrained CGARCH-S model with lognormal(0,1) based crashes

Paying the prize of using a model, which might be explosive we get a more dynamic model without the constraint, and it is doing quite well. The model with a non-crash distribution (for which we give no picture here) with a positive mean looks similar and has a better performance on both levels.



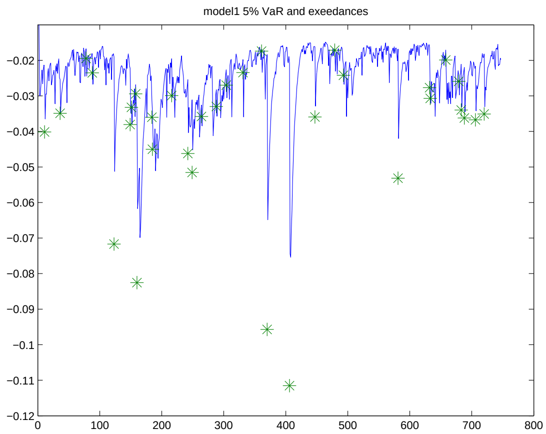


Figure 11.9: 1 % Value at Risk and exceedances of an unconstrained CGARCH-S model with lognormal(0,1) based crashes model

Figure 11.10: 5 % Value at Risk and exceedances of an unconstrained CGARCH-S model with lognormal(0,1) based crashes model

Considering more extreme crashes (lognormal(1,1) based) doesn't improve the model in terms of exceedances on the 1% level. This is due to the fact that the more extreme crashes lead to a smaller crash probability. On the 5% level we have fewer exceedances, but that is bad, because we had already an conservative model in the case with the lognormal(0,1) based crashes.

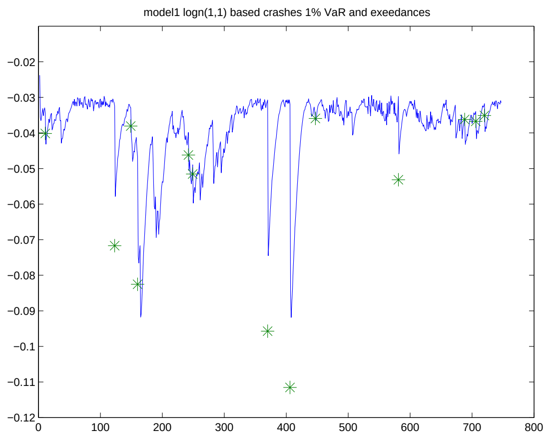


Figure 11.11: 1 % Value at Risk and exceedances of an unconstrained CGARCH-S model with lognormal(1,1) based crashes model

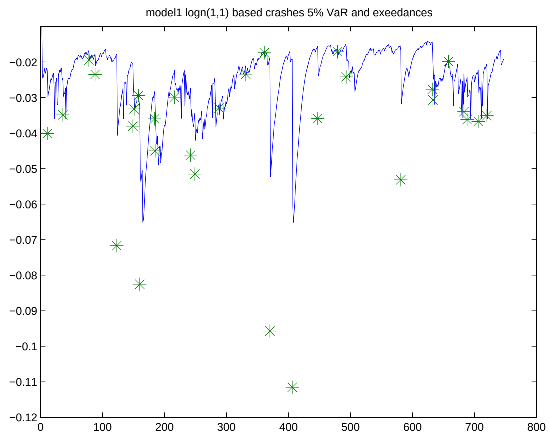


Figure 11.12: 5 % Value at Risk and exceedances of an unconstrained CGARCH-S model with lognormal(1,1) based crashes model

With added external variables modeling $\{p_t\}$ in the model with lognormal(0,1) based crashes, we get barely an improvement in terms of VaR exceedances.

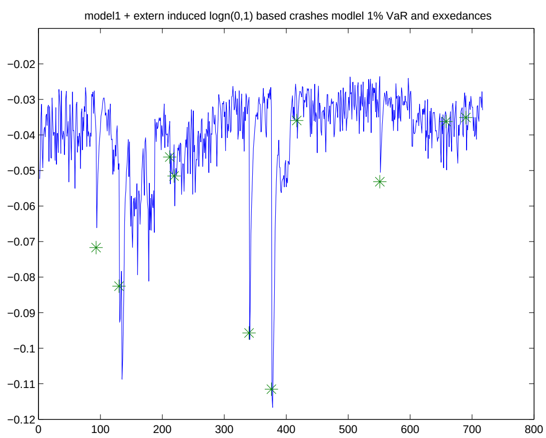


Figure 11.13: 1 % Value at Risk and exceedances of a CGARCH-SX model with lognormal(0,1) based crashes model

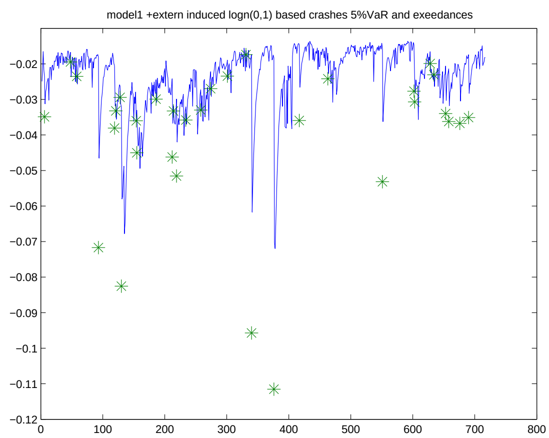


Figure 11.14: 5 % Value at Risk and exceedances of a CGARCH-SX model with lognormal(0,1) based crashes model

Now we investigate the model with the signum function applied to X_{t-1} transformed being the crash probability. That means where we used

$$p_t = \frac{1}{1 + \exp(-(a + b\,\mathrm{sign}(X_{t-1})))}.$$

Beside having the theory of this model more developed than the other models with an self reference in the crash probability, it works in comparison very well as a model for the Value at Risk.

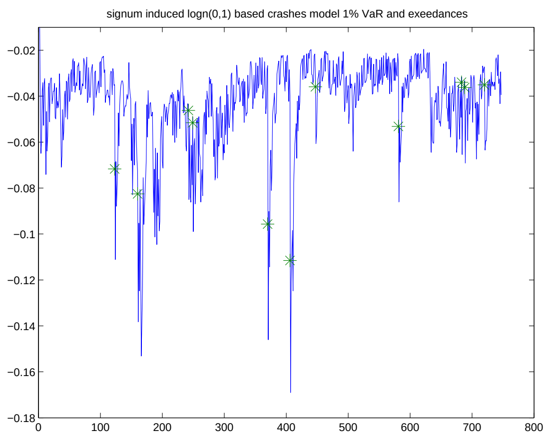


Figure 11.15: 1 % Value at Risk and exceedances of a CGARCH-SB model with lognormal(0,1) based crashes

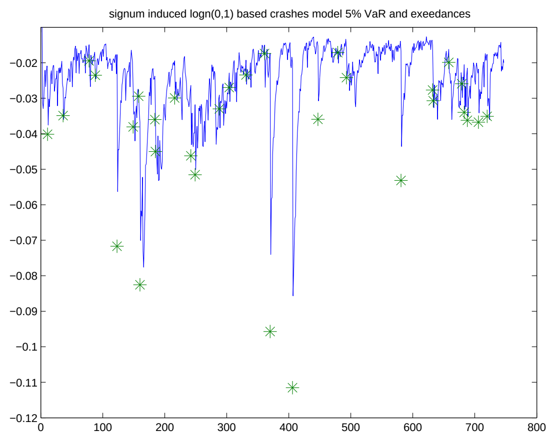


Figure 11.16: 5 % Value at Risk and exceedances of a CGARCH-SB with model lognormal(0,1) based crashes

If we take the negative of the lognormal distribution with parameter (1,1) as the basis for the crashes, we get the only model beside the model with crash probabilities depending purely on external variables being conservative with respect to the 1% VaR exceedances. And this fact holds instead of the considerable noisiness of the 1% VaR, which arises from the jump nature of p_t in cooperation with the extremely different tail behavior of the crash and the non-crash distribution.

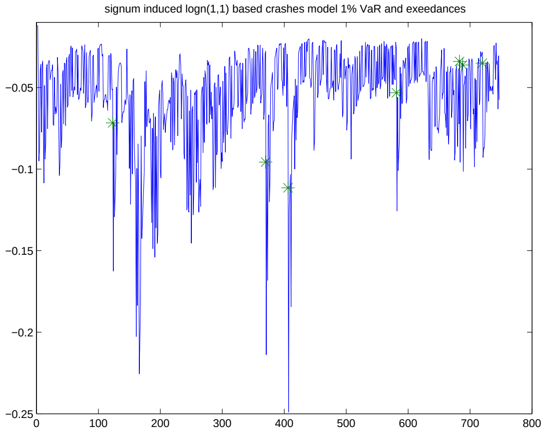


Figure 11.17: 1 % Value at Risk and exceedances of a CGARCH-SB model with lognormal(1,1) based crashes

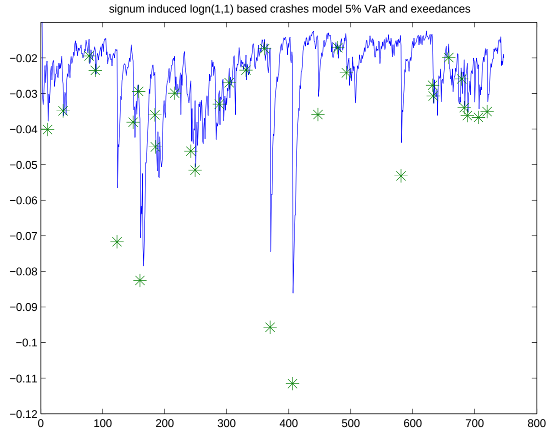
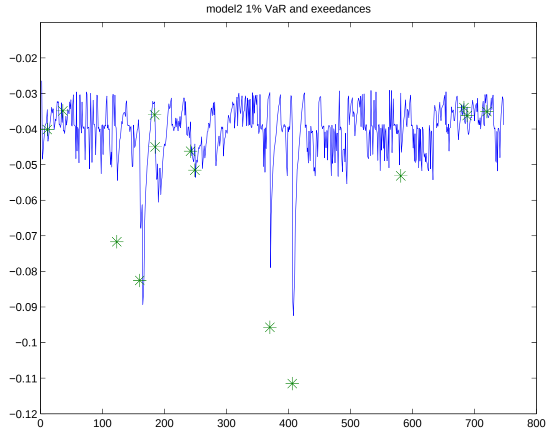


Figure 11.18: 5 % Value at Risk and exceedances of a CGARCH-SB model with lognormal(1,1) based crashes

We now switch to the CGARCH-V model. Just having simulation results in mind this occurred to be a model more realistic than the models above. However, on the VaR side, the exceedances show that to this purpose this model seems not to be better than the CGARCH-S model. In fact from some point of view it seems to perform quite poorly. The reason is that the same mechanism which prevented the crashes in simulations to become too extreme, namely getting σ_t by dividing the conditional variance by an term which gets big if p_t does, certainly also makes the quantiles in time points with high crash probability smaller. So we get a relatively high number of exceedances of the VaR on extreme levels. Also less extreme levels are affected. Using lognonormal(0,1) based crashes (no picture) leads to lower numbers of exceedances on both levels, mocking the reasons why we introduced this model.



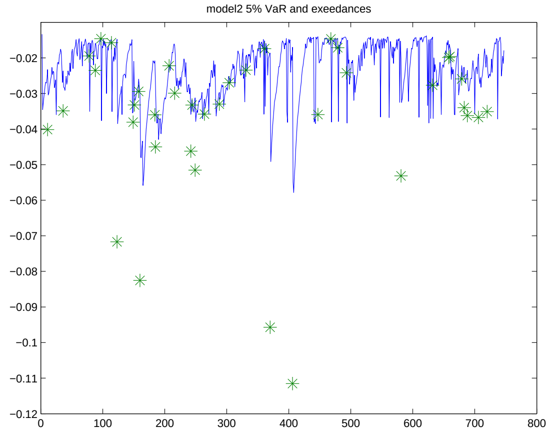
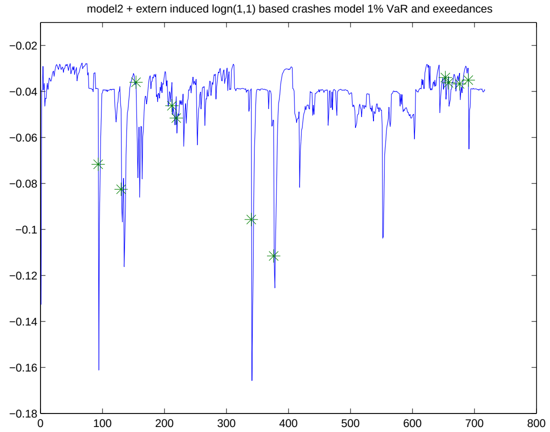


Figure 11.19: 1 % Value at Risk and exceedances of a CGARCH-V model with lognormal(1,1) based crashes

Figure 11.20: 5 % Value at Risk and exceedances of a CGARCH-V model with logn(1,1) based crashes

If we take additional external variables into account, that means working in a CGARCH-VX model, the situation gets better. But far from showing the best performance the awkward theory of this model class and the strange estimation outcome of this model's crash probabilities should be reason enough to treat this model with caution.



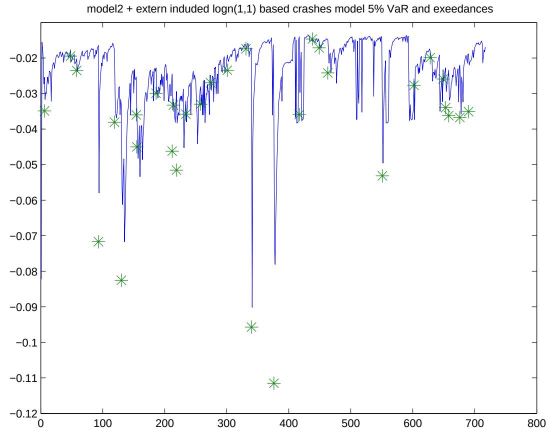


Figure 11.21: 1 % Value at Risk and exceedances of a CGARCH-VX model with lognormal(1,1) based crashes

Figure 11.22: 5 % Value at Risk and exceedances of a CGARCH-VX model with logn(1,1) based crashes

The pure GARCH model with t_4 innovations seems to work quite well here. It is better on the 1% level than all other models save the $X(1,1)$ and the $S(1,1)$

Model	5% ex	1% ex
GARCH	29	13
GARCHt_4	45	8
S(0,1)	31	10
S(1,1)	27	10
S(0,1)0.0001	32	10
SB(0,1)	36	11
SB(1,1)	34	10
SX(0,1)	32	12
SP(0,1)	31	12
SP(1,1)	30	8
V(1,1)	31	11
V(0,1)	35	12
VX(1,1)	37	10

Table 11.3: Value at Risk exceedances of Deutsche Bank

model, and it is not bad at the 5% level. But as we will see, when we investigate other time series of stock returns this result seems to be only representative for the 1% level. In the case of the other estimated time series the 5% Value at Risk is underestimated.

In order to check that the previous findings don't depend on the particular stock we fitted the models discussed above to the log-return time series of the Deutsche Bank in the same time period. A bank was chosen, because banking being a non-industrial business seemed a market segment as far from chemistry as possible. Besides that judging by pure eyesight the BASF log-returns seem to be among the "wildest" data available for the observed period and the Deutsche Bank seems to be much smoother. We omit the constrained models, because the problems with them are due to the strong constraint rather than due to the time series they are used to model.
Talking about models of CGARCH-S type our intuition of the Deutsche Bank to be smoother than the BASF data seems to be confirmed: Choosing a more extreme crash distribution gives us a 5% VaR exceedance ratio which is far below the desired result. On the 1% level the different genuine CGARCH-S models seem to be equally good. The positive mean non crash distribution yields a better result at the 5% level than the zero mean. Using additional external variables doesn't seem to make the model better: The improvement on the 5% level trades in a worse 1% level behavior.
Using just a signum induced crash probability is not worse than using the information of the volatility and the value. When applying the not so extreme crash distribution the 1% level is a little worse, but the 5% level is nearly nominal

Model	5%	1%
GARCH	0.0404	0.0181
GARCHt_4	0.0628	0.0112
S(0,1)	0.0432	0.0139
S(1,1)	0.0377	0.0139
S(0,1)0.0001	0.0446	0.0139
SB(0,1)	0.0502	0.0153
SB(1,1)	0.0474	0.0139
SX(0,1)	0.0446	0.0167
SP(0,1)	0.0432	0.0167
SP(1,1)	0.0418	0.0112
V(1,1)	0.0432	0.0153
V(0,1)	0.0488	0.0167
VX(1,1)	0.0511	0.0139

Table 11.4: exceedance ratios

(though not conservative). The model with more extreme crash distribution gets the same 1% level performance than the variations of the original model and a slightly better 5% performance.
When we stick to purely externally induced crashes, we can see that choosing a lognormal(0,1) based crashes model is performing worse than the original or the signum model. Using lognormal(1,1) based crashes gives us a model, which seems to be the best for establishing extreme quantiles. But on the 5% level the performance is not good compared to the other new models.
The CGARCH-V model with extreme crashes is as good as the original model on the 5% level but worse on 1%. Using less extreme crashes improves 5% performance on the cost of the 1% level performance. Adding external variables yields the same 1% performance as in the CGARCH-SX model. Looking at the 5% level we are not conservative anymore, though nominally closer to the desired result than most other models.
The only obvious case where one of the new models is worse than the pure GARCH model is the 5% case of the CGARCH-S model with extreme crashes. The signum01 and the volatility model with external variables might be judged as worse than the GARCH if we just accept conservative modeled VaR estimates. The pure GARCH model with t_4 innovations, which seemed to be a very good model for the BASF data is good again on the 1% level, but it is the worst on the 5% level. It isn't just far from nominal, it obviously underestimates the risk here.

We also tried to show that the findings still hold, when looking to a differ-

Modell	5% VaR ex	1 % VaR ex
GARCH	189	74
GARCHt_4	241	51
S(0,1)	194	69
S(1,1)	196	61
S(0,1)0.0001	198	69
SB(0,1)	193	67
SB(1,1)	194	63
SX(0,1)	195	59
SP(0,1)	196	63
SP(1,1)	191	50
V(1,1)	205	73
V(0,1)	200	71
VX(1,1)	184	56

Table 11.5: VaR exceedances in the models of the long BASF data

ent time period. We used daily BASF stock returns again. The time period of data is from May 1987 to October 2004. The BASF data and the DAX data used to calculate the external variables are from two different sources. (`http://corporate.basf.com/de/investor/aktie/kurs.htm` for the BASF stocks and `http://www.markt-daten.de/daten/DAX.txt` for the DAX) The given BASF prices of the time period observed above differ from the closing price used in the previous survey. The source gives no explanation what prices they used. We actually used 4349 observations. As to be expected the fitted models differ from the ones we got for the relatively short time series. But at least in tendency we get similar results concerning the ratio of the number of VaR exceedances to the length of the time series.
All models aren't conservative on the 1% level but the crash models all are nearer to the nominal level than the pure GARCH model with normal innovations. On the 5 % level on the other hand all models are conservative. The only exception is the pure GARCH model with t_4 innovations. Like in the case of the Deutsche Bank time series it fails on this level. This might suggest, that its good performance in the short BASF time series was just coincidental. The crash models except the CGARCH-VX model are again nearer to the nominal 5% than the pure GARCH model with normal innovations. Taking both levels into account the model with the signum induced crash probabilities is at least as good as the original model, although it is using less "information". In the CGARCH-S model the use of non-crash distribution with positive mean increased the the performance on the 5% level slightly. The use of the more extreme crash distribution increased the performance on both levels. This difference to the short time survey might be explained by the fact, that all models (including the pure GARCH)

Model	5%	1%
GARCH	0.0434	0.0170
GARCHt_4	0.0536	0.0117
S(0,1)1	0.0446	0.0159
S(1,1)	0.0451	0.0140
S(0,1)0.0001	0.455	0.0159
SB(0,1)	0.0444	0.0154
SB(1,1)	0.0446	0.0145
SX(0,1)	0.0448	0.0136
SB(0,1)	0.0451	0.0145
SB(1,1)	0.0439	0.0115
V(1,1)	0.0471	0.0168
V(0,1)	0.460	0.0163
VX(1,1)	0.0423	0.0129

Table 11.6: exceedance ratios

yielded in the long case a very small α and so are not very dynamic. So my guess is that a more extreme crash distribution grabbed the remaining dynamic better than the less extreme one. The use of external parameters increased the performance.
Coming to the models using external variables exclusively for modeling the crash probabilities we see that the grading of the use of a more or less extreme crash distribution is not clear in this long term series. The less extreme yields a better 5% performance the more extreme a better 1% performance. Seeing that the latter is the best of all models used here and that firstly the pure GARCH is beaten on the 5% level and secondly the 5% performance can be increased by altering the crash probability function by using an estimate of the DAX conditional variance instead of the volatility as external variables without changing the 1% level, I still would prefer the more extreme crash distribution in this context.
The CGARCH-V model with lognormal(1,1) based crashes performed on the 5% level better than all other models, but is the worst of the new models on the 1% level, beating pure GARCH just by one exceedance. Adding external variable's dynamic to the crash probability trades in a quite good 1% exceedance for the worst 5% performance of all models we checked here save the GARCH model with t_4 innovations. Using the less extreme crash distribution lowers the number of exceedances on both levels, not changing the relation to the CGARCH-S type models.

So taking all three numerical surveys into account we can say that CGARCH-V model in all its incarnations seems to be not very suitable for questions concerning

extreme quantiles. It exposes weaknesses at least on the one percent level in all three studies. On the 5% level it seemed quite suitable. But here the less extreme crash distribution seems to be better in the overall view, contradicting the reasons we had for introducing this model. Adding external variables gives more reasonable 1% exceedance ratios, but the picture on the 5% level is not so clear. Taking into account that this model is the one which takes most time to estimate and that we have no clear idea about its theoretical properties I think models of GARCH-S type should be preferred.
In the CGARCH-S model the use of a non-crash distribution with positive mean had a positive effect on the performance on at least one level, leaving the other level unaffected. Using a more extreme crash distribution only had a positive effect in the long time survey. Its use yielded poorer results in both short series. The use of additional external variables yielded an overall improvement of the VaR performance only in the long time case. In the short BASF case the situation is less clear: There is only one exceedance less in the 1% level and the picture in the Deutsche Bank case the picture gets more positive on the 5% level whereas the performance on the 1% level is worse. In the framework of models of CGARCH-S type there seems to be evidence that the signum models are at least as good as the models using the past values and current volatility for modeling the crash probabilities. The choice of a good crash distribution in the signum case might depend on the particular time series.
The exclusive use of external variables with an extreme crash distribution seems to be the model of choice if we are interested in extreme quantiles. On the 5% level this model isn't so good but still better than the classic GARCH. If fitting short series we can get quite poor results when using a crash distribution which is not extreme.

11.2 The ACGARCH-V models

We wanted to investigate also the alternative models from section 3.4. Working in a more general context, like we did in the case of the CGARCH-S model and CGARCH-V models proved to be numerically instable. To make a statement we chose a situation, where the theory from section 6.4 works. So we investigated a pure ARCH(1) dynamic on distributions defined on the whole real line. Namely we chose both to be normally distributed with different means and variances. This yielded stable numerical results, but these were not very satisfactory. As we will see in the following table there isn't any difference in the outcome of Value at Risk if we compare a pure ARCH model with an ACARCH-V model with very small variance in the distribution of D_t. This is due to the fact that this model is showing only a little dynamic in the crash probability which is jittering around 10^{-12} as seen in figure 11.23.

Model	5 %Var ex	5 % ratio	1 % Var ex	1% ratio
ARCH(1)	18	0.0241	7	0.0094
ACARCH(1)-V $N(10^{-}5, 1), N(-0.1, 2)$	14	0.0187	5	0.0067
ACARCH(1)-V $N(10^{-}5, 1), N(-0.1, 0.2)$	18	0.0241	7	0.0094

Table 11.7: ARCH Model and ACARCH-V models Value at Risk performance

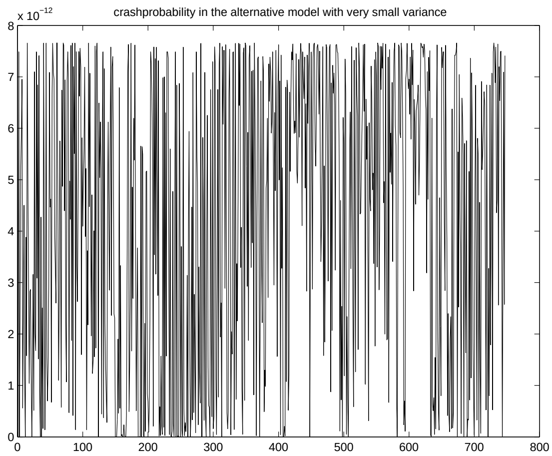


Figure 11.23: The crash probabilities when using small crash variance are too small to show effects

Choosing a fairly large variance of the Crash process leads to a real dynamic but the outcome is even worse than the pure ARCH–model, being by far too pessimistic.
Besides that the very goal we wanted to achieve by introducing the alternative model, namely the managing of positive means fails. Taking the long BASF data and estimating we get an conditional mean process with a negative sample mean. This is preserved when simulating. The effect is seen nicely in figure 11.24 plotting the real stock prices versus simulated ones.

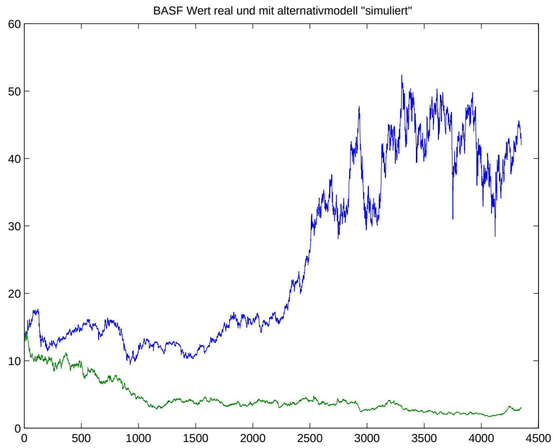


Figure 11.24: Effect of the failure to model a positive mean in the ACARCH-V model with big variance in the Crash–distribution, upper curve: BASF data, lower curve: Simulation

So it seems that even in the pure ARCH case, where we can handle the ACARCH-V model theoretically we are let down by it for practical purposes. The reason for that might be that it is very hard to pick a decent crash mean and crash distribution in this context.

Appendix A

List of the models we introduced

Here we want to resume the notation for all the models we introduced. Some of them were not really investigated here, but are introduced implicitly. Due to the fact we intend to model crashes we call the models we introduced. Crash GARCH models, or in short form CGARCH.
But we have to distinguish between the two kinds of modeling the mean we used, so we split notationally:

- In models with
$$X_t = \sigma_t[(1 - B_t)\epsilon_t + B_t D_t]$$
with no further constraint on the moments of ϵ_t and D_t we just use the notation CGARCH.
- In models defined via
$$X_t = (1 - p_t)\mu + p_t\delta + \sigma_t[(1 - B_t)\epsilon_t + B_t D_t]$$
where $E(D_t) = E(\epsilon_t) = 0$ and δ and μ are constants. we add the adjective "alternative" before the model name and abbreviate it with A. So we speak of Alternative Crash GARCH or ACGARCH models.

Now we also have do distinguish between the different modeling approaches:

1. whether we use a genuine GARCH dynamic or just an ARCH dynamic
2. whether we model σ_t^2 or v_t^2 via a GARCH equation
3. the way we model p_t

We deal with point 1 using the notation CGARCH for genuine GARCH dynamic and CARCH for pure ARCH dynamic. If we model σ_t^2 via a GARCH equation we speak of CGARCH-S, if we model v_t^2 via a GARCH equation we speak of CGARCH-V. If we deviate from modeling p_t as a function of X_{t-1} and σ_t, respectively v_t we express this by an extra character:

- X if we use additional external variables.
- P if we we use just external variables, where we use P, because p_t is a stochastic process in its own right in this context
- B if the function defining p_t is just a transform of $\text{sign}(X_{t-1})$, where B stands for binary, because p_t only takes two values with positive probability.

So we have introduced following models with genuine GARCH dynamic in the CGARCH context:

- CGARCH-S with GARCH dynamic in σ_t^2
- CGARCH-SX like CGARCH-S with additional dependence of p_t on external variables
- CGARCH-SB like CGARCH-S, p_t being a transform of the signum function
- CGARCH-SP like CGARCH-S, p_t depending exclusively on external variables
- CGARCH-V GARCH dynamic in v_t^2
- CGARCH-VX like CGARCH-V with additional dependence of p_t on external variables
- CGARCH-VB like CGARCH-V, p_t being a transform of the signum function
- CGARCH-VP like CGARCH-S, p_t depending exclusively on external variables

and corresponding the models using just an ARCH dynamic:

- CARCH-S with ARCH dynamic in σ_t^2
- CARCH-SX like CARCH-S with additional dependence of p_t on external variables
- CARCH-SB like CARCH-S, p_t being a transform of the signum function
- CARCH-SP like CARCH-S, p_t depending exclusively on external variables
- CARCH-V ARCH dynamic in v_t^2
- CARCH-VX like CARCH-V with additional dependence of p_t on external variables
- CARCH-VB like CARCH-V, p_t being a transform of the signum function

- CARCH-VP like CARCH-S, p_t depending exclusively on external variables

Now all this notation can be done for the ACGARCH case too:

- ACGARCH-S with GARCH dynamic in σ_t^2
- ACGARCH-SX like CGARCH-S with additional dependence of p_t on external variables
- ACGARCH-SB like CGARCH-S, p_t being a transform of the signum function
- ACGARCH-SP like CGARCH-S, p_t depending exclusively on external variables
- ACGARCH-V GARCH dynamic in v_t^2
- ACGARCH-VX like CGARCH-V with additional dependence of p_t on external variables
- ACGARCH-VB like CGARCH-V, p_t being a transform of the signum function
- ACGARCH-VP like CGARCH-S, p_t depending exclusively on external variables

and corresponding the models using just an ARCH dynamic:

- ACARCH-S with ARCH dynamic in σ_t^2
- ACARCH-SX like CARCH-S with additional dependence of p_t on external variables
- ACARCH-SB like CARCH-S, p_t being a transform of the signum function
- ACARCH-SP like CARCH-S, p_t depending exclusively on external variables
- ACARCH-V ARCH dynamic in v_t^2
- ACARCH-VX like CARCH-V with additional dependence of p_t on external variables
- ACARCH-VB like CARCH-V, p_t being a transform of the signum function
- ACARCH-VP like CARCH-S, p_t depending exclusively on external variables

We mainly made only use of the ACARCH-V model in ACGARCH context.

For smaller variations in the models, like using in the the GARCH regression X_{t-1} or $X_{t-1} - E(X_{t-1}|\mathfrak{F}_{t-2})$, we didn't introduce an extra notation. We just state in loco which model we used. We also did not introduce an extra notation for the restricted models with GARCH dynamic from section 8.1. Diciples of short forms might like ACGARCH-RV here where R stands for restricted.

Bibliography

[Ame85] Takeshi Amemiya, *Advanced Econometrics*, Harvard University Press, 1985.

[And87] Donald W.K. Andrews, *Consistency in Nonlinear Econometric Models: A Generic Uniform Law Of Large Numbers*, Econometrica **55** (1987), no. 6, 1465–1471.

[BHK03] Istvan Berkes, Lajos Horvath, and Piotr Kokoszka, *GARCH Processes: Structure and Estimation*, Bernoulli **9** (2003), no. 2, 201–227.

[Bil61] Patrick Billingsley, *Statistical Inference for Markov Processes*, Statistical Reasearch Monographs, vol. II, The University of Chicago Press, 1961.

[Bol86] T. Bollerslev, *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics **31** (1986), 307–327.

[Bos96] Denis Bosq, *Nonparametric Statistics for Stochastic Processes*, Lecture Notes in Statistics, no. 110, Springer, 1996.

[BP92a] Philippe Bougerol and Nico Picard, *Stationarity of GARCH Processes and of some Nonnegative Time Series*, Journal of Econometrics **52** (1992), 115–127.

[BP92b] Philippe Bougerol and Nico Picard, *Strict Stationarity of Generalized Autoregresive Processes*, The Annals of Probability **20** (1992), no. 4, 1714–1730.

[CC02] Marine Carrasco and Xiahong Chen, *Mixing and Moment Properties of various GARCH and Stochastic Volatility Models*, Econometric Theory **18** (2002), 17–39.

[Cha00] Chris Chatfield, *Time-Series Forecasting*, Chapman & Hall/CRC, 2000.

[CL01] C.S.Wong and W.K. Li, *On a Logistic Mixture Autoregressive Model*, Biometrica **88** (2001), no. 3, 833–846.

[Dav73] Y. Davydov, *Mixing Conditions for Markov Chains*, Theory of Probability and its Applications **78** (1973), 312–328.

[DLW94] Francis X. Diebold, Joon-Haeng Lee, and Gretchen C. Weinbach, *Regime Switching with Time–varying Transition Probabilities*, Nonstationary Time Series Analysis and Cointegration (C.P. Hargreaves, ed.), Oxford University Press, 1994, pp. 283–302.

[DM01] Randal Douc and Catherine Matias, *Asymptotics of the Maximum Likelihood Estimator for General Hidden Markov Models*, Bernoulli **7** (2001), no. 3, 381–420.

[Dou94] Paul Doukhan, *Mixing: Properties and Examples*, Lecture Notes in Statistics, no. 85, Springer, 1994.

[ELR87] Robert F. Engle, David M. Lilien, and Russel P. Robins, *Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model*, Econometrica **55** (1987), no. 2, 391–407.

[Eng82] Robert F. Engle, *Autoregressive Conditional Heterocedasticity with Estimates of the Variance of United Kingdom Inflation*, Econometrica **50** (1982), no. 4, 987–1007.

[FHH01] J. Franke, W. Härdle, and C. Hafner, *Einführung in die Statistik der Finazmärkte*, Springer, 2001.

[FRZ01] Chistian Francq, Michel Roussignol, and Jean-Michel Zakoian, *Conditional Hetererokedasticity driven by Hidden Markov Chains*, Journal of Time Series Analysis **22** (2001), 197–219.

[FS93] J. Franke and T. Seligmann, *Conditional Maximum Likelihood Estimates for INAR(1) Processes and their Application to Modelling Epileptic Seizure Counts*, Developments in Time Series Analysis: In Honour of Maurice B. Priestley (T. Subba Rao, ed.), Chapman & Hall, 1993, pp. 310–330.

[FZ04] Chistian Francq and Jean-Michel Zakoian, *Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes*, Bernoulli **10** (2004), no. 4, 605–637.

[Gou97] Christian Gourieroux, *ARCH Models and Financial Applications*, Springer Series in Statistics, Springer, 1997.

[GW88] A.Ronald Gallant and Halbert White, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, 1988.

[Han91a] Bruce E. Hansen, *GARCH(1,1) Processes are Near Epoch Dependent*, Economics Letters **36** (1991), 181–186.

[Han91b] Bruce E. Hansen, *Strong Laws for Dependent Heterogenous Processes*, Econometric Theory **7** (1991), 213–221.

[Hon91] Eun Pyo Hong, *The Autocorrelation Structure for the GARCH-M Process*, Economics Letters **37** (1991), 129–132.

[HW99] Philip Hua and Paul Wilmott, *Crash Modelling, Value at Risk and Optimal Hedging*, OFRC Working Paper Series (1999), `www.finance.ox.ac.uk/file_links/mf_papers/1999mf06.pdf`.

[Kor01] Ralf Korn, *Crash and Earn*, WILMOTT **1** (2001).

[Kre85] Ulrich Krengel, *Ergodic Theorems*, de Gruyter Studies in Mathematics, no. 6, Walter de Gruyter, 1985.

[KW02] Ralf Korn and Paul Wilmott, *Optimal Portfolios under the Threat of a Crash*, International Journal of Theoretical and Applied Finance **5** (2002), no. 2, 171–187.

[LH94] Sang-Won Lee and Bruce E. Hansen, *Asymptotic Theory for the GARCH(1,1) Quasi–Maximum Likelihood Estimator*, Econometric Theory **10** (1994), 29–52.

[Lum96] Robin L. Lumsdaine, *Consitency and Asymptotic Normality of the Quasi–Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models*, Econometrica **64** (1996), no. 3, 575–596.

[McL75] D.L. McLeigh, *A Maximal Inequality and Dependent Strong Laws*, The Annals of Probability **3** (1975), no. 5, 829–839.

[Mil93] Terence C. Mills, *The Econometric Modelling of Financial Time Series*, Cambridge University Press, 1993.

[Mok88] Abdelkader Mokkadem, *Mixing Properties of ARMA Processes*, Strict Stationarity of Generalized Autoregresive Processes **29** (1988), 309–315.

[MT93] S.P. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability*, Springer, 1993.

[Nel90] Daniel B. Nelson, *Stationarity and Persitence in the GARCH(1,1) Model*, Econometric Theory **6** (1990), 318–334.

[Phi86] Walter Philipp, *Invariance Principles for Independent and Weakly Dependent Random Variables*, Dependence in Probability and Statistics (Ernst Eberlein and Murad S. Taqqu, eds.), Progress in Probability and Statistics, vol. 11, Birkhäuser, 1986.

[PP97] B.M. Pötscher and I.R. Prucha, *Dynamic Nonlinear Econometric Models - Asymptotic Theory*, Springer, 1997.

[RSK01] Svetlozar Rachev, Eduardo Schwartz, and Irina Khindanova, *Stable Modeling of Credit Risk*, Proceedings of American Statistical Association Conference, 2001.

[She96] Neil Shephard, *Statistical Aspects of ARCH and Stochastic Volatility*, Time Series Models in Econometrics, Finance and other Fields (D.R. Cox, D.V. Hinkley, and O.E. Barndorff-Nielsen, eds.), Monographs on Statistics and Applied Probability, no. 65, Chapman & Hall, 1996, pp. 1–67.

[Sto74] W.F. Stout, *Almost Sure Convergence*, Academic Press, 1974.

[TK05] Joseph Tadjuidje Kamgaing, *Competing Neural Networks as Models for Non Stationary Financial Time series -Change Point Analysis-*, Ph.D. thesis, Technische Universität Kaiserslautern, `kluedo.ub.uni-kl.de/volltexte/2005/1818`, 2005.

[Ton90] Howell Tong, *Non–linear Time Series: A Dynamical System Approach*, Oxford statistical science series, Oxford University Press, 1990.

Lebenslauf

Geboren	13. Februar 1973 in Bonn
Schulen	1979 bis 1983 Katholische Grundschule “Am Hagelkreuz” in Bad Münstereifel–Nitterscheid 1983 bis Juni 1992 Erzbischöfliches St.Angela Gymnasium in Bad Münstereifel Abschluß: Abitur Note: 1.7
Wehrdienst	1992 bis 1993 in Koblenz Dienstgrad bei Ausscheiden: Hauptgefreiter
Studium	Oktober 1993 bis April 2001 Mathematik und Wirtschaftswissenschaften (Nebenfach) an der Albert–Ludwigs–Universität Freiburg
Auslandssemester	September 1996 bis Juni 1997 Mc Master University, Hamilton, (Canada) In diesem Rahmen Oktober 1996 bis November 1996 am Fields Institute, Toronto (Canada)
Examen	5. April 2001 Diplom–Mathematiker Note: 2
Promotionsstudium	Oktober 2001 bis 2005 an der Technischen Universität Kaiserslautern davon Oktober 2001 bis Oktober 2004 Stipendiat im Graduiertenkolleg Mathematik und Praxis

CV

Born	February 13th 1973 in Bonn
Civil Status	Single
Schools	1979 – 1983 Elementary School "Am Hagelkreuz" in Bad Münstereifel–Nitterscheid 1983 – 1992 High School St.Angela Gymnasium in Bad Münstereifel Degree: Abitur Mark: 1.7
Military Servive	1992 – 1993 in Koblenz
Studies	Oktober 1993 – April 2001 Mathematics and Economics (minor) at Albert–Ludwigs–Universität Freiburg
Studies Abroad	September 1996 – June 1997 Mc Master University, Hamilton, Ontario, (Canada) including a stay at the Fields Institute, Toronto (Canada) Oktober 1996 – November 1996
Gradutation	Degree: Diplom–Mathematiker April 5th 2001 Mark: 2
Graduate Studies	Oktober 2001 – September 2005 Mathematics at Technische Universität Kaiserslautern thereof Oktober 2001 – Oktober 2004 fellow of the Graduiertenkolleg Mathematik und Praxis