Essays on Financial Econometrics

with Applications to Commodity, Equity, and Foreign Exchange Markets

Doctoral Dissertation
in partial fulfillment of the requirements for the degree of
Dr. rer. pol.

by
Thomas Walther, M.Sc.
born June 11, 1986
in Eisenhüttenstadt, Germany

supervised by
Prof. Dr. Hermann Locarek-Junge
and
Prof. Dr. Bernhard Schipp
Faculty of Business and Economics
Technische Universität Dresden

submitted: April 24, 2017
defensed: November 14, 2017

Contents

Abbreviations

AGARCH	Asymmetric Generalised Autoregressive Conditional Heteroscedasticity
AIC	AKAIKE Information Criterion
ARCH	Autoregressive Conditional Heteroscedasticity
ARCH-M	Autoregressive Conditional Heteroscedasticity in Mean
APARCH	Asymmetric Power Autoregressive Conditional Heteroscedasticity
ARMA	Autoregressive Moving Average
BCBS	Basel Committee on Banking Supervision
BIC	Bayesian Information Criterion
EGARCH	Exponential Generalised Autoregressive Conditional Heteroscedasticity
EMU	Economic and Monetary Union
ES	Expected Shortfall
FIGARCH	Fractionally Integrated Generalised Autoregressive Conditional Heteroscedasticity
FIAPARCH	Fractionally Integrated Asymmetric Power Autoregressive Conditional Heteroscedasticity
FIEGARCH	Fractionally Integrated Exponential Generalised Autoregressive Conditional Heteroscedasticity
FX	Foreign Exchange
GARCH	Generalised Autoregressive Conditional Heteroscedasticity
GJR	GLOSTEN, JAGANNATHAN, RUNKLE
HYGARCH	Hyperbolic Generalised Autoregressive Conditional Heteroscedasticity
IGARCH	Integrated Generalised Autoregressive Conditional Heteroscedasticity
i.i.d.	Independent and identically distributed
MAE	Mean Absolute Error
MLE	Maximum-Likelihood Estimation

MMGARCH	Mixture Memory Generalised Autoregressive Conditional Heteroscedasticity
MRS	Markov-Regime-Switching
NGARCH	Nonlinear Generalised Autoregressive Conditional Heteroscedasticity
QMLE	Quasi Maximum-Likelihood Estimation
QGARCH	Quadratic Generalised Autoregressive Conditional Heteroscedasticity
RMSE	Root Mean Squared Error
TGARCH	Threshold Generalised Autoregressive Conditional Heteroscedasticity
VaR	Value-at-Risk
WTI	West Texas Intermediate

List of Figures

List of Tables

Symbols

Roman Letters

a	Value-at-Risk level
AS	ACERBI and SZEKELY (2014) test statistic (direct test)
b	HYGARCH coefficient
B	truncation lag
Chr	CHRISTOFFERSEN (1998) test statistic
$\mathbb{C}$ov	covariance operator
d	fractional integration coefficient
D	Outer-Product matrix
$\mathbb{E}$	expectation operator
F	cumulative distribution function
F^{-1}	quantile function of the distribution F
f	probability density function
g_t	high-frequency/short-term variance
H	Hessian matrix
h_t	conditional variance
I	indicator function
k	number of Spline knots
Kup	KUPIEC (1995) test statistic
ℓ	likelihood
L	lag operator
$\mathcal{L}$	Likelihood function
M	number of out-of-sample observations
n	number of model parameters

N	number of in-sample observations
p	GARCH lag order
$\mathbb{P}$	probability measure
P_{ij}	transition probability of moving from regime i to j
P_{t,S_t}	probability at time t of being in regime S_t
P	transition matrix
q	ARCH lag order
r_t	return series
R	number of regimes
S_t	regime at time t
T	number of observations
$\mathbb{V}$	variance operator
z_t	white noise series

Greek Letters

α_i	ARCH coefficients
β_i	GARCH coefficients
γ_i	leverage coefficients
Γ	Gamma function
δ	Box-Cox power transformation coefficient
ε_t	residuals, innovations
ζ	long-term ARCH coefficient
η	vector of conditional probability density functions
θ	parameter set
Θ	parameter space
κ	autoregressive coefficients
λ_i^{FI}	ARCH(∞) weights for FIGARCH
λ_i^{HY}	ARCH(∞) weights for HYGARCH

μ_t	conditional mean
ν	degrees-of-freedom (Student-t)
ξ	vector of state probabilities
$\rho(k)$	auto-correlation function with lag k
σ	unconditional volatility
τ_t	low-frequency/long-term variance
ϕ	FIGARCH coefficient
φ	standard Normal probability density function
Φ	standard Normal cumulative distribution function
Φ^{-1}	standard Normal quantile function
ψ	long-term GARCH coefficient
Ω_t	information set
ω	constant variance coefficient

Miscellaneous

$\odot$	element-wise multiplication operator

Acknowledgments

I would like to use this part to thank the people who accompanied me on my journey to complete this work. First of all, I thank my supervisor Prof. Hermann Locarek-Junge for giving me the opportunity to work at his department, for advices as well as providing the freedom to work on my own research interests. Also, I would like to thank Prof. Bernhard Schipp for introducing me into time series analysis and for being my second supervisor. I thank Prof. Stefan Huschens for his fruitful seminars on statistical problems. I am thankful to my department colleagues Arite Schrehardt, Denise Erhardt, Ruben Sippel, Sven Loßagk, Thorsten Klug, Leif Hansen, Anne Sumpf, and Nga Nguyen for help, advise, and hints. I am especially grateful to Tony Klein, with whom I started this episode and had always someone to discuss single and broader issues of scientific and not-so-much-scientific nature. I want to express my gratitude to other faculty fellows, such as the department of statistics (especially to Daniel Tillich), the department of econometrics, the department of energy economics, and the dean's office as well as to Prof. Antonio Roldan-Ponce, who supported me a lot in the beginning. Additionally, I want to thank the colleagues from other universities I met along the way: Phillipp Lauenstein, Paul Bui Quang, Duc Khuong Nguyen and Krzysztof Piontek. I am very thankful to the Deutsche Bundesbank, who partly financed my research stay in Vietnam. I thank my colleagues at the School of Business, International University–National University Ho Chi Minh City for their hospitality. I gratefully acknowledge the financial support of the Graduate Academy, Technische Universität Dresden, financed by *The Excellence Initiative* of the German Federal Ministry of Education and Research (BMBF) and the German Research Foundation (DFG). Moreover, I am thankful for the financial support provided by the Faculty of Business and Economics of the Technische Universität Dresden.

I would not enjoyed my journey as much if it was not for friends and family. I appreciate the help of my sister in-law Hiền Phạm Thu, who experienced the same struggles. I cannot thank my parents, Siegfried and Ursula, as well as my sister Anja enough for always supporting me.

Most of all, I thank my wonderful wife for her unconditional love, her understanding, and support.

To
my wife *Đức Anh*
and
my son *Leonard Minh*

1 Introduction

Financial econometrics is concerned with the statistical analysis of financial time series and it is relatively popular within the field of finance. In 2003, CLIVE W.J. GRANGER and ROBERT F. ENGLE received the "Nobel Prize in Economic Sciences" for their development of techniques for time series analysis. Especially the work of the latter influences how risk can be described from a financial perspective. By introducing the Autoregressive Conditional Heteroscedasticity (ARCH) model, ENGLE started a stream of literature, which is still ongoing. In his seminal paper, ENGLE (1982) develops a model that describes volatility as a process of past serially uncorrelated innovations. Hitherto, the volatility was modelled to be constant over time, i.e. homoscedastic. With the Generalised ARCH (GARCH), BOLLERSLEV (1986) extends ENGLE's framework and provides one of the widest used models in financial risk management.

In addition, at least three different streams of volatility modelling exist: Firstly, the *realised volatility* aggregates higher-frequency data to estimates of the volatility (e.g. PARK and LINTON, 2012). Secondly, the *stochastic volatility* is a modelling concept comparable to ARCH models. However, the volatility is driven by its own stochastic process (TAYLOR, 1995, pp. 70-75). Finally, the *implied volatility* is derived by using market-data with inverted option price formulas and may be seen as the market's future expectations (e.g. FRANKE, HÄRDLE, and HAFNER, 2015, pp. 112f.).

In finance, where risk is the volatility of returns fluctuating around their mean, ARCH models have a great impact on various risk related areas. For example, the framework allows the quantification of risk, which is an essential part of risk management. With its various augmentations, ARCH models account for many so-called stylised facts. These facts are properties, which are usually observed in financial time series. Among others, CONT (2001, p. 224) mentions:

- *heavy tails*: the occurrence of extreme events,
- *volatility clustering*: the fact that volatility groups in clusters of high and low volatility over time,
- *long memory*: slowly decaying autocorrelation in absolute returns, and
- *leverage effect*: the different impact of positive and negative returns on volatility.

Moreover, *structural breaks*—the change of the unconditional volatility over time—could possibly be explained by business cycles. Hence, incorporating these effects into ARCH models produces a more realistic depiction of risk, which is essential for applications in risk management.

This thesis provides an overview of the most prominent ARCH specifications. Moreover, the application to market risk quantification is highlighted with special focus on the Value-at-Risk (VaR) and Expected Shortfall (ES). These two parts build the methodological framework for six essays, which demonstrate the usage of ARCH models in the financial markets of equity, foreign exchange (FX), and commodities. The first two papers are concerned with commodity markets, namely crude oil and tanker freight rates. The third and fourth paper analyse the FX rates volatility of countries in transition (e.g. Poland). The fifth essay concentrates on the Vietnamese stock market. Lastly, the sixth paper presents a methodology for rapid computation of long memory ARCH models. In the following, a brief overview of each of the six papers is provided.[1]

1. *Oil Price Volatility Forecast with Mixture Memory GARCH*
   The first paper investigates the applicability of the Mixture Memory GARCH model (MMGARCH) on oil price volatility, which is of interest for numerous industries, e.g. the leisure and transportation industry or utilities. Previous studies investigated either long memory behaviour of oil price volatility or identified different regimes in the time series. The MMGARCH combines GARCH processes with short and long memory. The study reveals different memory structures in the main crude oil blends, the U.S. West Texas Intermediate (WTI) and the European Brent. The in- and out-of-sample performance of MMGARCH is compared to other standard GARCH models incorporating stylised facts such as asymmetry and long memory. It is found that both effects are present in crude oil volatility. The results show that MMGARCH outperforms all other models regarding the in-sample as well as the out-of-sample (variance and VaR forecast) analysis (KLEIN and WALTHER, 2016).

2. *Forecasting Volatility of Tanker Freight Rates Based on Asymmetric Regime-Switching GARCH Models*
   While the the first essay is focused on the product crude oil, the second paper analyses the volatility of tanker freight rates. As an essential part of oil transportation, the freight rates are of special interest due to the different origins of supply and demand. The demand side is mainly driven by the demand for oil, but the supply side is somewhat inelastic if one considers the size of the available fleet and the costs and time to increase it. Recent research reveals regimes of different structure of the volatility in the tanker freight market, while empirical evidence indicates the leverage effect. In addition to symmetric and asymmetric GARCH models, the performance of Markov-Regime-Switching GARCH variants is investigated in order to bring the two aforementioned aspects together. The underlying data includes the freight rates of Very Large Crude Carriers on the major global routes in the period 2000-2015. After seasonally adjusting the freight rates, regime-switching GARCH

[1] See Appendix A for the corresponding literature references.

models are found to outperform their single-regime complements in terms of in-sample fit and out-of-sample forecasting accuracy. The applicability of the models in freight risk management is compared by means of VaR and ES back testing procedures. The results show that accounting for volatility regimes and asymmetry does not enhance the performance of one-day-ahead forecasts (LAUENSTEIN and WALTHER, 2016).

3. *Empirical Evidence of Long Memory and Asymmetry in EUR/PLN Exchange Rate Volatility*
   This and the following study focus on the volatility of FX rates. Since most exchange rates follow a free floating regime, the volatility is an important indicator for the stability of a currency and vital to investors with trades affected by foreign currencies. The latter is especially true in central and eastern European countries, where most of the trades are related to countries within the European Economic and Monetary Union (EMU). In this work, the volatility of the exchange rate between the Polish Złoty and the Euro is modelled by implementing a variety of GARCH models under different return distributions. It is shown that the volatility exhibits an asymmetric and a long memory effect, separately and jointly. Hence, a GARCH model incorporating both effects is found to be superior over other models when forecasting the VaR (KLEIN, PHAM THU, and WALTHER, 2016).

4. *True or Spurious Long Memory in European Non-EMU Currencies*
   In addition to the Polish Złoty, this study analyses the Croatian Kuna, the Czech Koruna, the Hungarian Forint, the Romanian Leu, and the Swedish Krona. It is examined whether their Euro exchange rates volatility exhibits true or spurious long memory. It is well known that structural breaks might lead to spurious long memory behaviour. In a refined test strategy, true long memory is discriminated from spurious long memory for the six exchange rates. The findings suggest that Czech Koruna and Hungarian Forint only feature spurious long memory, while the rest of the series have both structural breaks and true long memory. Moreover, it is demonstrated how to extend existing models to depict both properties jointly yielding superior fit and better VaR forecasts (WALTHER et al., 2017).

5. *Expected Shortfall in the Presence of Asymmetry and Long Memory: An Application to Vietnamese Stock Markets*
   As a member of large upcoming multinational free trade agreements, Vietnam is in the focus of foreign investors. However, literature on market properties is rather scarce. This study analyses the conditional volatility of the two major Vietnamese stock indices with a specific focus on the application to risk management. After testing for long memory in returns and squared returns, GARCH models are used to account for asymmetry and long memory effects. These models are then used

to estimate the Value-at-Risk and the Expected Shortfall. The main results are that both indices have long memory in their squared returns, but differ in the asymmetric impact of negative and positive news on volatility as well as for the persistence of shocks. Long memory GARCH models perform best when estimating risk measures for both series (WALTHER, 2017).

6. *Fast Fractional Differencing in Modeling Long Memory of Conditional Variance for High-Frequency Data*
   In contrast to the aforementioned empirical studies, the last essay proposes a new method to compute the conditional volatility of long memory GARCH models by using Fast Fourier transforms. It is demonstrated how calculation times of parameter estimations benefit from this new approach without changing the estimation procedure. A more precise depiction of long memory behaviour becomes feasible. The new approach offers a computational advantage to most long memory GARCH models. Risk management applications like rolling-window Value-at-Risk predictions are substantially sped up. This new approach allows to calculate the conditional volatility of high-frequency data in a practicable amount of time (KLEIN and WALTHER, 2017).

By applying GARCH models and incorporating different stylised facts, the aforementioned essays provide deeper insight into the structure of variance in commodity, equity, and foreign exchange markets. Special focus is set to models with asymmetric effect, long memory behaviour, and structural breaks. The analysed stylised facts and the content of the essays are summarised in Tab. 1.

No.	Data	Asymmetry	Long Memory	Structural Breaks	Risk Measures
1	Commodities	✓	✓	✓	VaR
2	Commodities	✓		✓	VaR, ES
3	FX	✓	✓		VaR
4	FX		✓	✓	VaR
5	Equity indices	✓	✓		VaR, ES
6	Simulation		✓		

Table 1: Summary of the essays with overview of analysed stylised facts.

The remainder is structured as follows: Chapter 2 reviews several ARCH specifications and the corresponding stylised facts. Chapter 3 provides an overview of the estimation of risk measures in combination with GARCH models. Finally, Chapter 4 concludes this work and offers possible further research opportunities.

2 Models of Conditional Variance

The following equations formulate the basis of the econometric framework used in this work (BAUWENS, HAFNER, and LAURENT, 2012, pp. 3-5):

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t, \\ \varepsilon_t &= \sqrt{h_t} z_t, \qquad \text{with } z_t \text{ i.i.d. } \forall t \in \mathbb{Z},\ \mathbb{E}[z_t] = 0,\ \text{and } \mathbb{V}[z_t] = 1, && (1) \\ \mu_t &= \mathbb{E}[r_t | \Omega_{t-1}], \\ h_t &= \mathbb{V}[r_t | \Omega_{t-1}], && (2) \end{aligned}$$

where $(r_t)_{t \in \mathbb{Z}}$ is a return series and z_t is a realisation of an independent and identically distributed (i.i.d.) random variable. The conditional mean μ_t and the conditional variance h_t are measurable functions with respect to the sigma-algebra Ω_{t-1}, which is generated by all returns and possibly other variables up to time $t - 1$. The random variable z_t is drawn from a continuous distribution[2] and is independent from Ω_{t-1}. For μ_t the class of Autoregressive Moving Average (ARMA) models and its (fractionally) integrated variations are considerable (GRANGER, 1980 and BOX, JENKINS, and REINSEL, 2008). In what follows, various possible representations of h_t, representing different stylised facts, are considered. Furthermore, it is shown how to estimate the parameters, derive standard errors, and forecast with the different variance models.

Note that the focus is set on univariate models. However, multivariate ARCH models, especially in combination with conditional correlation exist, but are not covered in this work.[3]

2.1 ARCH Model and its Extensions

2.1.1 Autoregressive Conditional Heteroscedasticity Models

In his empirical analysis of speculative prices MANDELBROT (1963, p. 418) finds that "large changes tend to be followed by large changes—of either sign—and small changes tend to be followed by small changes". What the author describes is commonly know as *volatility clustering*. Figure 1 shows the weekly returns of the DAX30 index. Especially in the years 2001-2003, 2009, and 2011-2012, it appears that the amplitude of the returns is higher than in the rest of the sample, non-regarding whether the returns are positive or negative. To quote FAMA (1965, pp. 56-58):

[2] The standard Normal distribution is often used, but the choice set is not limited to this particular distribution.
[3] For an introduction to multivariate ARCH models see e.g. LÜTKEPOHL (2006, pp. 557-584) and FRANCQ and ZAKOIAN (2010, pp. 273-310).

“It may be that the distribution of price changes at any point in time is normal, but across time the parameters of the distribution change. A company may become more or less risky, and this may bring about a shift in the variance of the first differences.”

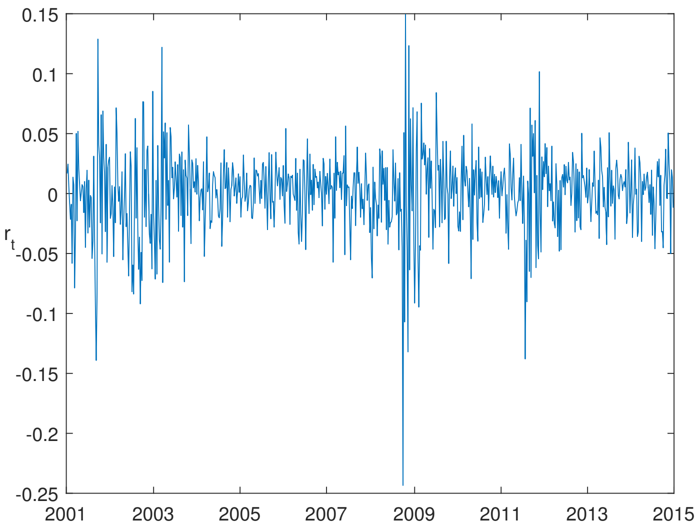

Figure 1: Weekly DAX30 returns January 2, 2001-December 29, 2014.

Hence, using unconditional second-order moments to measure risk over the whole sample, neglects the time-varying property of the variance. Combining the two ideas of dependent and varying variance, ENGLE (1982) introduces the Autoregressive Conditional Heteroscedasticity model, which is given by:

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2. \tag{3}$$

As mentioned before, h_t is the conditional variance (Eq. 2). In the ARCH(q) regression model, it is characterised by constant variance level ω and the lags on the squared residuals with order q, where $\varepsilon_t^2 = (r_t - \mu_t)^2$. In order to maintain stationarity and non-negativity, it has to hold that $\omega, \alpha_i \geq 0$ for all $i = 1, \ldots, q$ and $\sum_{i=1}^{q} \alpha_i < 1$ (ENGLE, 1982, p. 993, Theorem 2).

Empirical studies using ARCH often need a high lag-order and hence have the necessity to estimate many parameters. To reduce the amount of model parameters, ENGLE (1983) implements a linear declining weight function for an ARCH(8) model. ENGLE, LILIEN, and ROBINS (1987) even use twelfth-order ARCH models. Interestingly, the authors incorporate the ARCH model in the mean equation and formulate the so-called ARCH-in-mean (ARCH-M) model. This concept allows for time-varying variance and can be interpreted as a risk premium on financial returns.

2.1.2 Generalised ARCH Models

BOLLERSLEV (1986) presents a generalisation of ENGLE's model. The Generalised ARCH is augmented with an autoregressive term on the conditional variance of order p. This yields a smooth and exponentially declining autocorrelation function. Furthermore, it allows for a more parsimonious structure and hence, fewer parameters. The GARCH(p,q) process can be described as follows:

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}. \tag{4}$$

Here, additional parameter restrictions are the non-negativity of β_j for all $j = 1, \dots, p$ and the relation $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ for stationarity. BOLLERSLEV (1986, p. 310, Theorem 1) shows that the GARCH process (Eq. 4) is wide-sense stationary, i.e. covariance or weakly stationary, with $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{V}[\varepsilon_t] = \frac{\omega}{1-\sum_{i=1}^{q}\alpha_i - \sum_{j=1}^{p}\beta_j}$, and $\mathbb{C}\mathrm{ov}[\varepsilon_t, \varepsilon_s] = 0$ for $t \neq s$, if and only if $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$. Moreover, NELSON (1990) argues that $\mathbb{E}[\log(\beta_1 + \alpha_1 z_t^2)] < 0$ is a sufficient condition for GARCH(1,1) to be strictly stationary. BOUGEROL and PICARD (1992, pp. 116-118) formulate the condition for GARCH(p,q).

In some cases it is desirable to apply a non-stationary, i.e. non-mean-reverting, variant of GARCH. The Integrated GARCH (IGARCH), introduced by ENGLE and BOLLERSLEV (1986a), is similar to an integrated ARMA model on the conditional mean process. It examines the case where the polynomial $1 - \sum_{i=1}^{q} \alpha_i z^i - \sum_{j=1}^{p} \beta_j z^j$ has at least one unit root. The authors consider two types of IGARCH:

(1) without trend ($\omega = 0$), and

(2) with trend ($\omega > 0$).

Given the restriction $\alpha_1 + \beta_1 = 1$, the IGARCH(1,1) can be formulated as:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + (1 - \alpha_1) h_{t-1}.$$

It is important to mention that IGARCH does not have finite variance and thus, is not weakly stationary. However, it is still strictly stationary as NELSON (1990, p. 321) points out. Moreover, the IGARCH is said to be persistent in variance (ENGLE and BOLLERSLEV, 1986a, p. 27), i.e. all past shocks influence future predictions of the process.[4]

The IGARCH(1,1) without trend is also known as RiskMetrics (J. P. MORGAN, 1996, pp. 77-102). RiskMetrics has pre-set parameters $\alpha_1 = 0.06$ and $\beta_1 = 0.94$ for daily data and $\alpha_1 = 0.03$ and $\beta_1 = 0.97$ for monthly data. These "optimal" parameters are

[4] NELSON (1990, pp. 322-325) discusses the definition of "persistence" more deeply. However, for the purpose of this work, only the definition in ENGLE and BOLLERSLEV (1986a) is considered. See also BOLLERSLEV and ENGLE (1993) for the multivariate case of co-persistence.

derived by using the Root Mean Squared Error (RMSE) as a criterion. The authors estimate the parameters with the smallest RMSE for a large set of countries and financial time series and conclude that the proposed parameter set is the weighted average over all observed markets. The perception for this simplification is mixed (e.g. McMillan and Kambouroudis, 2009). However, its advantage is that it can be incorporated into a spread sheet without having to estimate the parameters.

2.1.3 Asymmetric GARCH Models

One drawback of the standard GARCH model lies in its nature to depend on the squared residual ε_t^2. Consequently, there is no discrimination between positive and negative shocks in the standard GARCH model. However, empirical studies show that "good news" and "bad news" impact volatility differently. Various explanation for the asymmetric effect are given in literature. Some works also name it *leverage effect*. Christie (1982, pp. 423-425) argues that financial leverage is positively correlated with equity volatility. Hence, it is said that negative returns reduce the equity and given a fixed debt, an increased debt-to-equity ratio, i.e. financial leverage (Franke, Härdle, and Hafner, 2015, p. 285). French, Schwert, and Stambaugh (1987), Campbell and Hentschel (1992), and Bekaert and Wu (2000) advocate the idea of volatility feedback, i.e. time-varying risk premiums. These authors show that the leverage ratio is not the only source of the effect and asymmetry still exists after filtering for financial leverage. While these explanations might fit to equity volatility, they do not account for other asset classes, where this effect is also present.[5] Alternatively, Avramov, Chordia, and Goyal (2006) present selling or trading activity in general as a different reason and show that stocks without leverage appear to have the same effect. Lastly, Smith (2016) presents results that the differences between negative and positive innovations are varying for different weekdays, which cannot be explained by any of the aforementioned theories.

Nevertheless, the asymmetric effect on volatility is incorporated in many GARCH augmentations and subsequently empirically proven, albeit no final solution to the "leverage puzzle" has been found yet. In the following, the most prominent asymmetric GARCH models are presented.

Nelson (1991) proposes the exponential GARCH (EGARCH) model. Following Engle and Ng (1993), a possible EGARCH(1,1) representation is given by:

$$\log(h_t) = \omega + \gamma_1 z_{t-1} + \alpha_1 (|z_{t-1}| - \mathbb{E}[|z_{t-1}|]) + \beta_1 \log(h_{t-1}). \tag{5}$$

The additional coefficient γ_1 measures whether "good" or "bad" news impact the conditional variance more ($\gamma_1 < 0$ or $\gamma_1 > 0$, respectively). While γ_1 measures the sign of

[5] Among others, Klein (2017) finds an inverted leverage effect for precious metals. Klein and Walther (2016) report the leverage effect for major crude oil volatility.

the standardised residual $z_t = \frac{\varepsilon_t}{\sqrt{h_t}}$ (sign effect), the coefficient α_1 accounts for the size or magnitude of z_t (size effect). If $\alpha_1 > 0$ ($\alpha_1 < 0$) then shocks above the expected size of the innovations z_t increase (decrease) the $\log(h_{t+1})$. Since the logarithm of h_t is modelled, the process does not need any restrictions to maintain non-negativity for the conditional variance. HE, TERÄSVIRTA, and MALMSTEN (2002, pp. 870f.) show that EGARCH is strictly stationary if and only if $|\beta_1| < 1$. Furthermore, the process has finite moments, if the underlying distribution of z_t has finite unconditional moments. Additionally, $\mathbb{E}[|z_t|]$ is also dependent on the distribution of z_t. If z_t is drawn from a Normal distribution, it can be shown that $\mathbb{E}[|z_t|] = \sqrt{2/\pi}$.[6]

The model proposed by GLOSTEN, JAGANNATHAN, and RUNKLE (1993, p. 1787) is often referred to as GJR. The authors distinguish positive and negative shocks by means of an indicator function:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 I_{\{\varepsilon_{t-1}<0\}} \varepsilon_{t-1}^2 + \beta_1 h_{t-1}.$$

The indicator function $I_{\{\varepsilon_{t-1}<0\}}$ is one if the last shock is negative, otherwise it is zero. Similar to GJR, ZAKOIAN (1994) introduces the Threshold GARCH (TGARCH). In its simplest form it can be written as:

$$\sqrt{h_t} = \omega + \alpha_1 I_{\{\varepsilon_{t-1}>0\}} \varepsilon_{t-1} + \gamma_1 I_{\{\varepsilon_{t-1}<0\}} \varepsilon_{t-1} + \beta_1 \sqrt{h_{t-1}}.$$

Other models incorporating asymmetric shocks in some way are the Asymmetric GARCH (AGARCH, ENGLE, 1990)[7], the Nonlinear GARCH (NGARCH, HIGGINS and BERA, 1992), or the VGARCH (ENGLE and NG, 1993). However, more prominently used than the aforementioned models is the Asymmetric Power ARCH (APARCH) by DING, GRANGER, and ENGLE (1993). The APARCH(p,q) can be formulated as follows:

$$h_t^{\frac{\delta}{2}} = \omega + \sum_{i=1}^{q} \alpha_i \left(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}\right)^{\delta} + \sum_{j=1}^{p} \beta_j h_{t-j}^{\frac{\delta}{2}}. \tag{6}$$

The standard GARCH restrictions are augmented with $\delta \geq 0$ and $\gamma_i \in (-1, 1)$ for all $i = 1, \ldots, p$. Here, $\gamma_i > 0$ indicates that negative shocks have more impact on the conditional variance than positive shocks. The APARCH combines the asymmetric effect and the flexibility to model a different power of the conditional standard deviation. In many empirical studies, the Box-Cox power transformation parameter δ tends to be less than 2 (e.g. KLEIN and WALTHER, 2016, p. 52). Interestingly, the model includes seven other models: ARCH, GARCH, GJR, TGARCH, and NGARCH to mention the ones described

[6] See e.g. LAURENT and PETERS (2002, pp. 453f.) for $\mathbb{E}[|z_t|]$ if the underlying distribution of z_t is a (skewed) Student-t or General Error distribution.

[7] Sometimes the AGARCH is also mentioned as Quadratic GARCH (QGARCH). See e.g. FRANSES and VAN DIJK (1996, p. 230). A more general form is discussed by SENTANA (1995).

above. The augmented GARCH by DUAN (1997) additionally includes EGARCH.

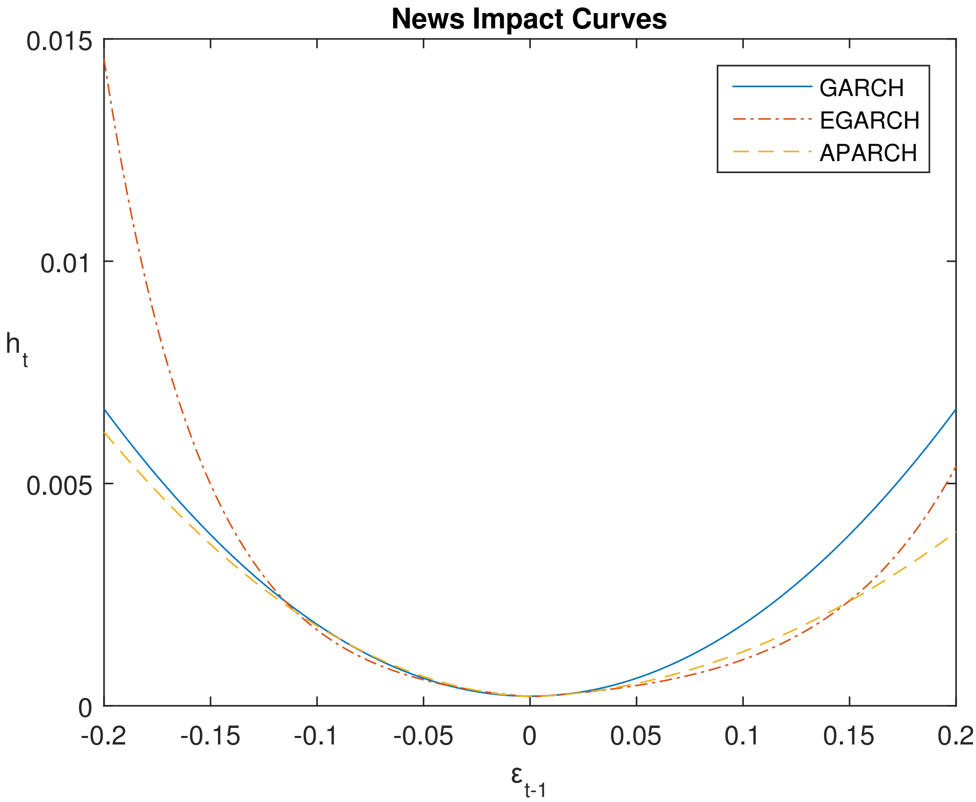

Figure 2: News impact curve for GARCH, EGARCH, and APARCH with Student-t distribution based on the results of WALTHER (2017) for the Vietnamese stock index VNI in the period July, 15 2005-December, 31 2015. The lagged conditional variance is set to the unconditional variance $h_{t-1} = \bar{\sigma}_t^2 = 2.6127 \cdot 10^{-4}$.

Once the parameters of the asymmetric GARCH models have been estimated, the leverage effect can be analysed. ENGLE and NG (1993) introduce the news impact curve—a graphical approach to visualise the influence of shocks on volatility. Figure 2 shows the news impact curve for GARCH, EGARCH, and APARCH for the data of WALTHER (2017). While the symmetric GARCH model responses with the same impact on the conditional variance h_t for positive and negative shocks ε_{t-1}, EGARCH and APARCH behave differently. In the case of EGARCH, the conditional volatility h_t is more influenced by negative shocks, given the steeper slope for $\varepsilon_{t-1} < 0$ in comparison to GARCH. On the contrary, the APARCH model has the same impact for negative shocks as GARCH, but places less weight on positive innovations. Additionally to the news impact curve, ENGLE and NG (1993, pp. 1757-1763) propose diagnostics to test the sign bias, the negative size bias, and the positive size bias, individually and jointly.

2.1.4 Long Memory GARCH Models

Another important stylised fact is the so-called *long memory* or long range dependence. It states that past distant observations still impact recent ones. One possible definition of the effect is that the auto-correlation function ρ of a stationary process z_t is not summable (FRANKE, HÄRDLE, and HAFNER, 2015, p. 318):

$$\lim_{T\to\infty} \sum_{k=-T}^{T} |\rho(k)| = \infty.$$

In finance, the autocorrelation function of empirically observed absolute or squared returns declines very slowly (e.g. hyperbolically). Usually, squared returns are used as a proxy for variance. Thus, the slowly declining autocorrelation in squared returns indicate long memory behaviour. Figure 3 shows the returns and the squared returns of the Brent oil price used in the study of KLEIN and WALTHER (2016). In the upper plot, the autocorrelation declines immediately. Contrary in the lower plot, the autocorrelation is slowly declining up to 100 lags.

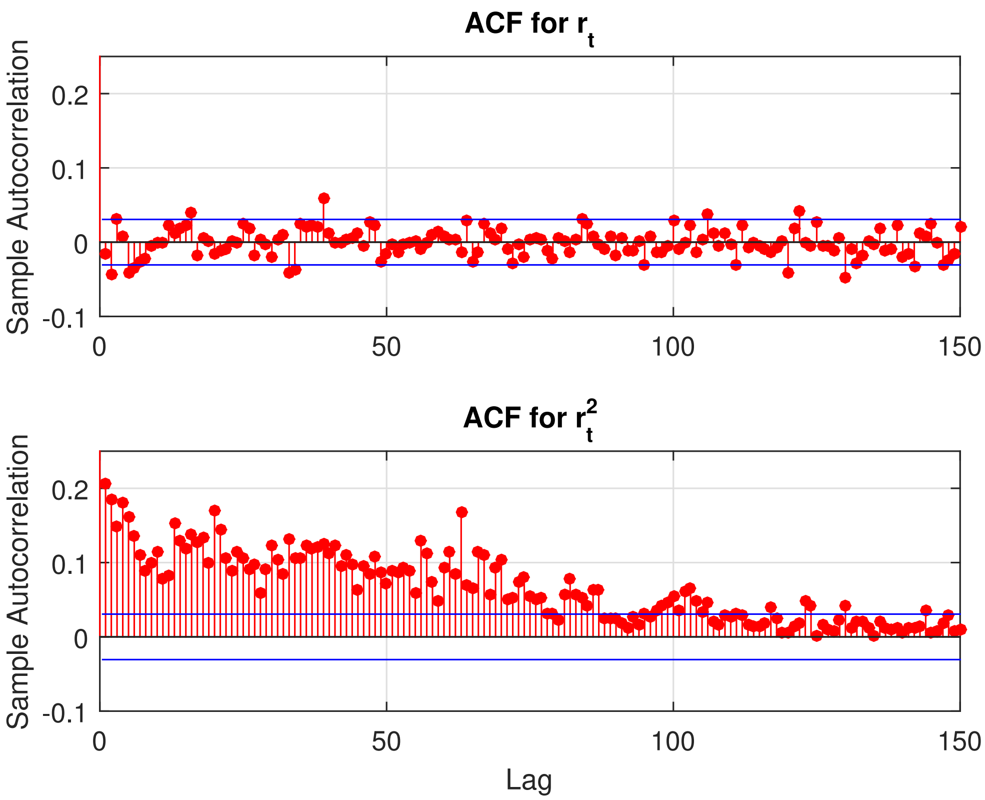

Figure 3: Sample autocorrelation function (ACF) for Brent oil price returns r_t and squared returns r_t^2, January 2, 1998-December 31, 2014. The blue bounds indicate the 95% confidence interval for the estimated autocorrelation.

The fractional integration is a way to incorporate this effect into the modelling of financial returns. GRANGER and JOYEUX (1980) and GRANGER (1980) introduce the fractional integration into ARMA models. For volatility models, BAILLIE, BOLLERSLEV, and MIKKELSEN (1996) formulate the Fractionally Integrated GARCH (FIGARCH). In contrast to the original GARCH, the FIGARCH is able to depict (1) long memory with only one additional parameter (d) and (2) a slowly, hyperbolically decaying auto-correlation instead of an exponential decay. The FIGARCH(1,d,1) can be described as:

$$\begin{aligned} h_t &= \frac{\omega}{1-\beta_1} + \left(1 - \frac{(1-\phi_1 L)(1-L)^d}{1-\beta_1 L}\right)\varepsilon_t^2 \\ &= \frac{\omega}{1-\beta_1} + \sum_{i=1}^{\infty} \lambda_i^{FI} \varepsilon_{t-i}^2, \end{aligned} \tag{7}$$

where

$$\begin{aligned} \lambda_1^{FI} &= \phi_1 - \beta_1 - d, \\ \lambda_i^{FI} &= \beta_1 \lambda_{i-1}^{FI} + \left(\frac{i-1-d}{i} - \phi_1\right)\left(\frac{(i-2-d)!}{i!(1-d)!}\right), \end{aligned} \tag{8}$$

L is the lag operator with $Lr_t = r_{t-1}$. The long memory parameter d is the real valued order of fractional integration. The last line in Eq. (7) corresponds to the ARCH(∞) representation of FIGARCH with weights λ_i^{FI} for all $i \in \mathbb{N}$ as defined in Eq. (8). The sufficient non-negativity constraints $\omega > 0$, $0 \leq \beta_1 \leq \phi_1 + d$, and $0 \leq d \leq 1 - 2\phi_1$ have to hold in order to refer to admissible parameters. A wider range of necessary and sufficient conditions can be found in CONRAD and HAAG (2006). However, the discussion on conditions for weak and strict stationarity of FIGARCH is still ongoing.[8] KAZAKEVIČIUS and LEIPUS (2003) question the existence of a stationary solution. DAVIDSON (2004, p. 20) points out that FIGARCH does not have a finite unconditional variance for any d.

Alternatively, DAVIDSON (2004) presents a generalised model: the hyperbolic GARCH (HYGARCH). Following CONRAD (2010, pp. 443-446), the HYGARCH(1,d,1) can be formulated:

$$\begin{aligned} h_t &= \omega + \left(1 - \frac{1-\phi_1 L}{1-\beta_1 L}\left(1 + b\left[(1-L)^d - 1\right]\right)\right)\varepsilon_t^2 \\ &= \frac{\omega}{1-\beta_1} + \sum_{i=1}^{\infty} \lambda_i^{HY} \varepsilon_{t-i}^2, \end{aligned} \tag{9}$$

where

$$\begin{aligned} \lambda_1^{HY} &= bd + \phi_1 - \beta_1, \\ \lambda_i^{HY} &= \beta_1 \lambda_{i-1}^{HY} + b\left(\frac{i-1-d}{i} - \phi_1\right)\left(\frac{(i-2-d)!}{i!(1-d)!}\right). \end{aligned} \tag{10}$$

The extra coefficient $b \in [0, 1]$ allows the special cases GARCH ($b = 0$) and FIGARCH ($b = 1$). Thus, the HYGARCH can be interpreted as a mixture of both models and remain

[8] DOUC, ROUEFF, and SOULIER (2008) show the existence of some FIGARCH processes. A recent review on the matter is provided by DAVIDSON and LI (2014).

non-negative, if the respective conditions for FIGARCH and GARCH are met. CONRAD (2010) provides necessary and sufficient non-negativity conditions for HYGARCH which are less restrictive.

The last two models, which are presented in this subsection, combine the stylised facts of long memory and the above mentioned leverage effect. Corresponding to the EGARCH model (Eq. 5), BOLLERSLEV and MIKKELSEN (1996) postulate the Fractionally Integrated EGARCH (FIEGARCH) by alternating Eq. (7):

$$\begin{aligned}\log(h_t) &= \frac{\omega}{1-\beta_1} + \left(1 - \frac{(1-\phi_1 L)(1-L)^d}{1-\beta_1 L}\right)(\gamma_1 z_t + \alpha_1(|z_t| - \mathbb{E}[|z_t|])),\\ &= \frac{\omega}{1-\beta_1} + \sum_{i=1}^{\infty} \lambda_i^{FI}(\gamma_1 z_{t-i} + \alpha_1(|z_{t-i}| - \mathbb{E}[|z_{t-i}|])).\end{aligned}$$

Furthermore, TSE (1998) combines the APARCH model (Eq. 6) with hyperbolically decay of shocks and formulates the Fractionally Integrated APARCH (FIAPARCH):

$$\begin{aligned}h_t^{\frac{\delta}{2}} &= \frac{\omega}{1-\beta_1} + \left(1 - \frac{(1-\phi_1 L)(1-L)^d}{1-\beta_1 L}\right)(|\varepsilon_t| - \gamma_1\varepsilon_t)^{\delta},\\ &= \frac{\omega}{1-\beta_1} + \sum_{i=1}^{\infty} \lambda_i^{FI}(|\varepsilon_{t-i}| - \gamma_1\varepsilon_{t-i})^{\delta}.\end{aligned}$$

Both models can also be transferred to their HYGARCH representations by changing the ARCH(∞) weights.

From the ARCH(∞) representations of the aforementioned long memory GARCH models, it can be seen that the infinite sum must be truncated to suit practical purposes. BAILLIE, BOLLERSLEV, and MIKKELSEN (1996, pp. 12f.) suggest to use at least 1,000 lags. Nonetheless, a data set of T observations and a truncation lag of B, translates to $T \cdot B$ calculations to obtain the full path of conditional variance. In view of parameter estimation and forecasting exercises, where the whole path has to be evaluated several times, the process is relatively time consuming. To ease this problem, KLEIN and WALTHER (2017) adopt the idea from JENSEN and NIELSEN (2014) to use Fast Fractional Fourier transforms (COOLEY and TUKEY, 1965) to compute the conditional variance. The computations reduce to $T \cdot \log(B)$ and offer an enormous potential for time savings.[9]

2.1.5 Regime Switching GARCH Models

Heretofore, all presented models keep the same structure when applied to actual data. By doing so, one neglects the possibility of different e.g. economic environments in the sam-

[9] In a Monte Carlo simulation, KLEIN and WALTHER (2017) show e.g. for T =5,000 and B =1,000 the computation time of FIGARCH(1,d,1) parameter estimation reduces from 10.92 seconds to 0.54 seconds.

ple period. Hence, in less (high) volatile times, the estimated parameters from a GARCH model yield a conditional variance, which is to high (low).[10] CAI (1994, p. 310) argues that the strong persistence in variance is due to structural changes. To overcome this possible bias, the Markov-Regime-Switching (MRS) framework introduced by HAMILTON (1989) can be used. Based on a Markov-Chain, each regime possesses its own set of parameters. HAMILTON and SUSMEL (1994) and CAI (1994) are the first to formulate Markov-Regime-Switching ARCH models. For R regimes with unobservable states $S_t \in \{1, \ldots, R\}$ at time t, the MRS-ARCH(q) process reads as follows:

$$\begin{aligned} r_t &= \mu_{t,S_t} + \sqrt{h_{t,S_t}} z_t \\ h_{t,S_t} &= \omega_{S_t} + \sum_{i=1}^{q} \alpha_{i,S_t} \varepsilon_{t-i}^2. \end{aligned}$$

The underlying first order Markov-Chain determines the current state S_t. The transition probabilities $P_{i,j} = \mathbb{P}[S_t = j | S_{t-1} = i]$ of moving from Regime i to j are collected in the transition matrix

$$\mathbf{P} = \begin{bmatrix} P_{1,1} & P_{2,1} & \cdots & P_{R,1} \\ P_{1,2} & P_{2,2} & \cdots & P_{R,2} \\ \vdots & \vdots & \ddots & \vdots \\ P_{1,R} & P_{2,R} & \cdots & P_{R,R} \end{bmatrix},$$

where each column in **P** sums up to unity, i.e. for the i-th column $\sum_{j=1}^{R} P_{i,j} = 1$. Note that $\mathbb{P}[S_t = i] > 0$ for all $i \in \{1, \ldots, R\}$. The transition probabilities are estimated along with the other model parameters (HAMILTON and SUSMEL, 1994, p. 316).

However, the formulation of a MRS-GARCH is much more cumbersome. Given the GARCH structure, the whole set of states $\{S_t, S_{t-1}, S_{t-2}, \ldots\}$ of the Markov-Chain has to be known in order to recursively calculate the current conditional variance h_{t,S_t}. For R regimes, R^T states have to be considered, which is practically impossible for larger sample sizes (CAI, 1994, p. 310).

GRAY (1996) circumvents the problem. For a MRS-GARCH(1,1) with R regimes, the author proposes to calculate the conditional expected value of h_t given the information at $t-1$, i.e.

$$h_{t,S_t} = \omega_{S_t} + \alpha_{S_t} \varepsilon_{t-1}^2 + \beta_{S_t} h_{t-1},$$

[10] The same motivation is used in the German article LOCAREK-JUNGE and WALTHER (2017).

with

$$
\begin{aligned}
h_t &= \mathbb{E}[h_{t,S_t}|\Omega_{t-1}] \\
&= \sum_{j=1}^{R} P_{t,S_t=j}\left(\mu_{t,S_t=j}^2 + h_{t,S_t=j}\right) - \left(\sum_{j=1}^{R} P_{t,S_t=j}\mu_{t,S_t=j}\right)^2 \\
\varepsilon_t &= r_t - \sum_{j=1}^{R} P_{t,S_t=j}\mu_{t,S_t=j},
\end{aligned}
$$

where $P_{t,S_t=j} = \mathbb{P}\left[S_t = j|\Omega_{t-1}\right]$ for $j = 1, \ldots, R$ is the probability of being in state j at time t. Hence, the variance h_{t,S_t}, conditional of time t and state S_t, is calculated given the information set Ω_{t-2}. KLAASSEN (2002) alternates the process and uses the information set Ω_{t-1}. Lastly, HAAS, MITTNIK, and PAOLELLA (2004b) use a different approach. Instead of conditioning the regime variance h_{t,S_t} on one mutual variance path, it is proposed that each regime has its own variance path. Thus, a MRS-GARCH(1,1) could read as follows:

$$h_{t,S_t} = \omega_{S_t} + \alpha_{S_t}\varepsilon_{t-1}^2 + \beta_{S_t}h_{t-1,S_t}. \tag{11}$$

Stationarity conditions for the MRS-GARCH models are discussed in HAAS, MITTNIK, and PAOLELLA (2004b); LIU (2006), and ABRAMSON and COHEN (2007). Once the parameters of the MRS-GARCH model are estimated, one can derive smoothed state probabilities P_{t,S_t} to improve inference with the algorithm presented in KIM (1994).

The GARCH variants presented in Sec. 2.1.1-2.1.4 can be used to substitute the underlying GARCH process in each regime (PEREZ-QUIROS and TIMMERMANN, 2001; ALOUI and JAMMAZI, 2009; HENRY, 2009). Figure 4 shows the two regimes from the Very Large Crude Carrier Route TD4 for monthly returns derived from a MRS-APARCH model (LAUENSTEIN and WALTHER, 2016). Here, the blue block indicates a regime of high volatility.

Another generalisation of the MRS models is to relax the assumption of constant transition probabilities. DIEBOLD, LEE, and WEINBACH (1994) introduce time-varying transition probabilities for the general class of MRS models. Among others KRÄMER (2008) and HENRY (2009) use this specification in a MRS-GARCH framework.

2.1.6 Mixture GARCH Models

Closely related to the discussed MRS-GARCH models above, is the class of Mixture GARCH models. Instead of having different regimes, this model class mixes distributions to obtain a better fit on the empirical distribution.[11] As mentioned earlier, the Normal dis-

[11] NOMIKOS and POULIASIS (2011, p. 322) mention that for the Mixture GARCH models, “what is important is the overall regime probability;” while for MRS-GARCH models “the probability of each observation belonging to any given regime is more important.”

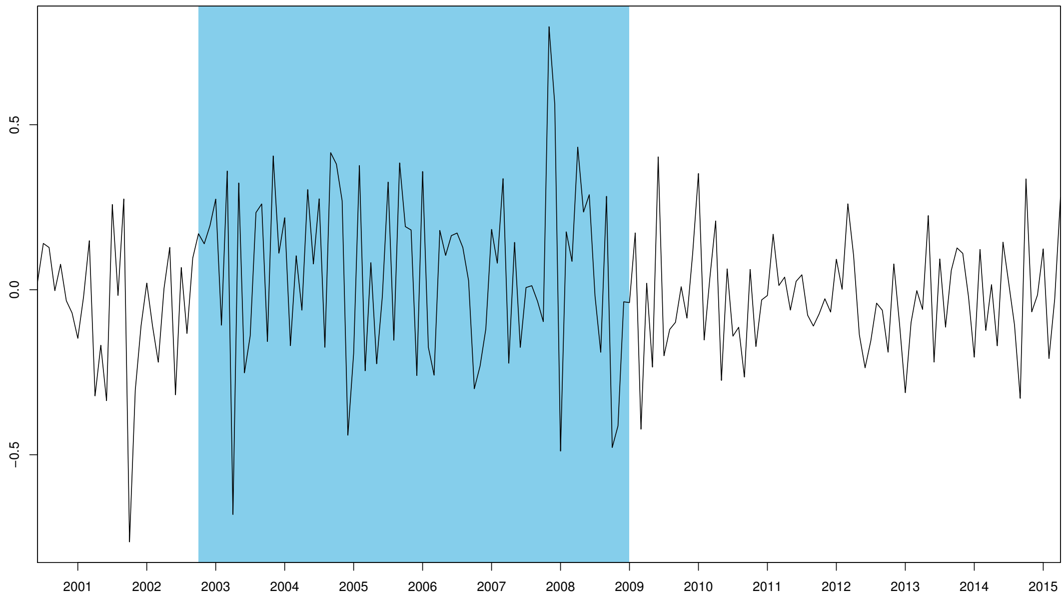

Figure 4: Regimes in tanker freight rates (Very Large Crude Carrier Route TD4, June 1, 2000-May 29,2015). The blue block indicates the "high volatility" regime and is derived from the smoothed probabilities from a MRS-APARCH with monthly returns. The data is based on the work of LAUENSTEIN and WALTHER (2016).

tribution is not capable to depict certain stylised facts, such as the fat tails. However, the mix of e.g. two Normal distributions is able to do so. HAAS, MITTNIK, and PAOLELLA (2004a) present the Mixture Normal GARCH model. The formulation of the GARCH process does not differ from the one presented in Eq. (11), except that one does not consider time-dependent regimes S_t, but constant mixture components $S \in \{1, \ldots, K\}$. Thus, the mixed conditional variance is given by:

$$h_t = \sum_{i=1}^{K} P_{t,S=i} h_{t,S=i},$$

where the probability of i-th component $P_{t,S=i} = \mathbb{P}[S = i]$ is constant over time and can be interpreted as a weight. A more flexible approach is advocated by CHENG, YU, and LI (2009). The authors' Dynamic Mixture GARCH model allows for time-varying mixtures. In a two component setting, the state probability is given by e.g. a logistic link function

$$P_{t,S=1} = \frac{1}{1 + \exp\left(\kappa_0 + \kappa_1 r_{t-1}\right)},$$

with $P_{t,S=2} = (1 - P_{t,S=1})$ and κ_0 and κ_1 as autoregressive parameters on r_t. LI, LI, and LI (2013) extend the idea and mix a standard GARCH with a FIGARCH component. The resulting Mixture Memory (MM-)GARCH can depict a component with short memory and one component with long memory. The model is applied by KLEIN and WALTHER (2016) on oil prices. The component-wise and full conditional density for the time series

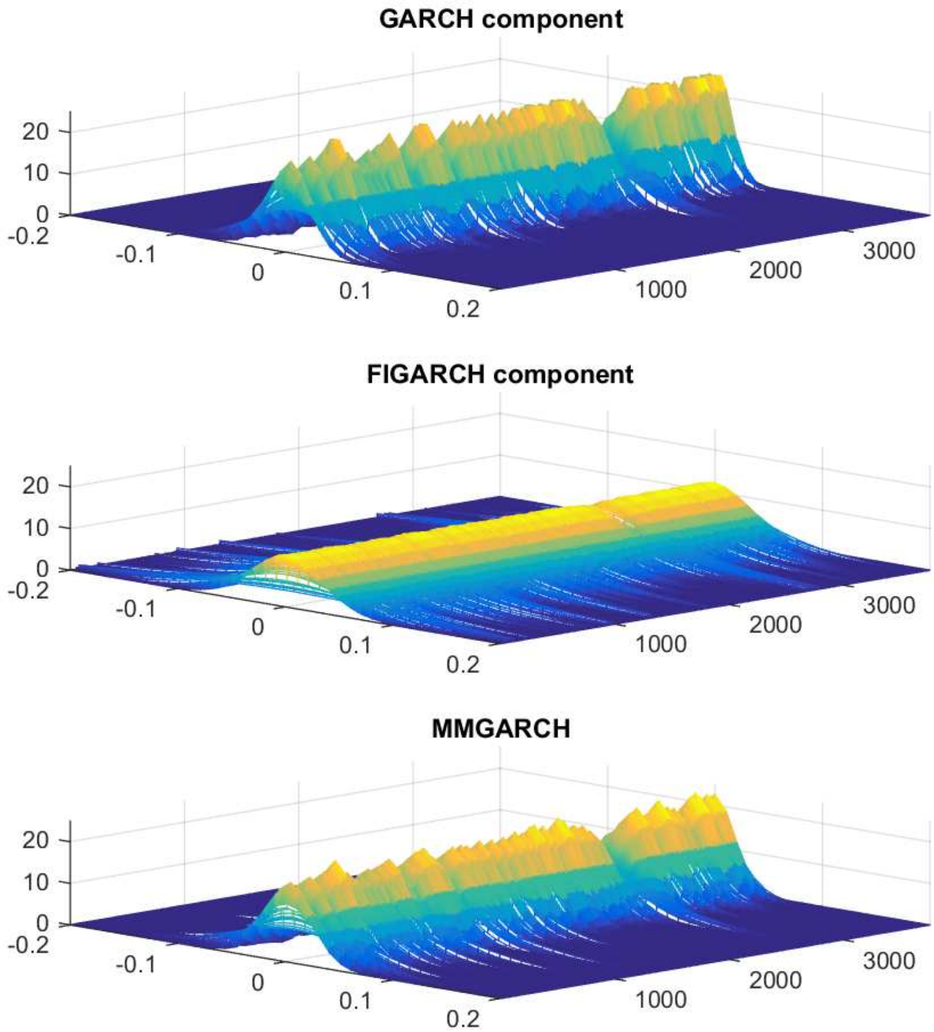

Figure 5: Component-wise probability density function of the Mixture Memory GARCH. The data for the West Texas Intermediate crude oil returns (1995-2014) is based on the work of KLEIN and WALTHER (2016).

of the oil blend WTI is presented in Figure 5. It can be seen that the two components have different volatilities and that the MMGARCH is mainly driven by the GARCH.

Other mixture GARCH variations are presented in VLAAR and PALM (1993); PALM and VLAAR (1997); and LIN and YEH (2000).

2.1.7 Component GARCH Models

The last set of models presented in this work, are the component GARCH models. The first variant is the component GARCH of DING and GRANGER (1996). By weighting single GARCH processes, the authors propose a model to better depict long memory behaviour (as in Sec. 2.1.4). The component GARCH specification of ENGLE and LEE (1999) goes a different direction. The model disentangles the variance into a long- (τ_t) and short-run (g_t) part. The model of additive nature reads as follows:

$$\begin{aligned}
h_t &= \tau_t + g_t, \\
g_t &= (\alpha + \beta)\, g_{t-1} + \alpha \left(\varepsilon_{t-1}^2 - h_{t-1}\right), \\
\tau_t &= \omega + \psi \tau_{t-1} + \zeta \left(\varepsilon_{t-1}^2 - h_{t-1}\right).
\end{aligned}$$

The parameter restrictions $1 > \psi > \alpha + \beta > 0$, $\beta > \zeta > 0$, and $\alpha, \beta, \zeta, \omega > 0$ are sufficient to guarantee stationarity and non-negativity. Moreover, the condition provides that the persistence in the long-run process τ_t dies out at a slower rate than in the short-run process g_t.

ENGLE and RANGEL (2008) suggest another approach. The Spline-GARCH decomposes the variance into low- and high-frequency factors. The low-frequency part τ_t is described by an exponential quadratic spline. The Spline(k)-GARCH with k splines is described as:

$$\begin{aligned}
h_t &= \tau_t g_t, \\
g_t &= (1 - \alpha - \beta) + \alpha \left(\frac{\varepsilon_{t-1}^2}{\tau_{t-1}}\right) + \beta g_{t-1}, \\
\tau_t &= c \exp\left(\omega_0 \frac{t}{T} + \sum_{i=1}^{k} \omega_i \max\left(\frac{t - t_{i-1}}{T}; 0\right)^2\right),
\end{aligned}$$

where $\{t_0 = 0, t_1, t_2, \ldots, t_k = T\}$ are the equidistant knots of the splines in τ_t. Interestingly, the expected value of the mean-reverting high-frequency part g_t is 1 by construction:

$$\begin{aligned}
\mathbb{E}\left[g_t\right] &= \mathbb{E}\left[(1 - \alpha - \beta) + \alpha z_{t-1}^2 + \beta g_{t-1}\right] \\
&= (1 - \alpha - \beta) + \alpha \mathbb{E}\left[z_{t-1}^2\right] + \beta \mathbb{E}\left[g_{t-1}\right], \\
\Leftrightarrow (1 - \beta)\, \mathbb{E}\left[g_t\right] &= (1 - \alpha - \beta) + \alpha, \\
\Leftrightarrow \mathbb{E}\left[g_t\right] &= 1,
\end{aligned}$$

provided that $\mathbb{E}[z_t^2] = \mathbb{V}[z_t] = 1$ (Eq. 1) and $\mathbb{E}[g_t] = \mathbb{E}[g_{t-1}]$. Thus, the unconditional variance is determined by the low-frequency part, i.e.

$$\mathbb{E}[h_t] = \mathbb{E}[\tau_t g_t] = \tau_t \mathbb{E}[g_t] = \tau_t. \tag{12}$$

Building on that idea, other model variations have emerged. AMADO and TERÄSVIRTA (2013) propose an additive and multiplicative Time-Varying GARCH and GJR with a smooth transition part described by a logistic transition function.[12] PASCALAU, THOMANN,

[12] GONZÁLEZ-RIVERA (1998) and BELKHOUJA and BOUTAHARY (2011) follow a similar idea.

and GREGORIOU (2010) and BAILLIE and MORANA (2009) use flexible Fourier forms (GALLANT, 1984) instead of a spline to describe the low-frequency part, while the high-frequency is driven by a GARCH and FIGARCH process, respectively. Finally, ENGLE, GHYSELS, and SOHN (2013) replace the non-parametric spline with the Mixed Data Sampling approach by GHYSELS, SANTA-CLARA, and VALKANOV (2004).

In the Spline-GARCH, the number of knots k has to be set in advance or selected up on an information criterion (see Sec. 2.2). WALTHER et al. (2017) suggest to use a structural break point test instead. The break points of the Iterated Cumulative Sum of Squares algorithm (INCLAN and TIAO, 1994; SANSÓ, ARAGÓ, and CARRION, 2004) are then used to set the knots in the Spline-GARCH.[13] By this means, the knots are not necessarily equidistant. Figure 6 shows how in the multiplicative component Spline-GARCH model, the high-frequency part fluctuates around a common trend represented by the low-frequency component.

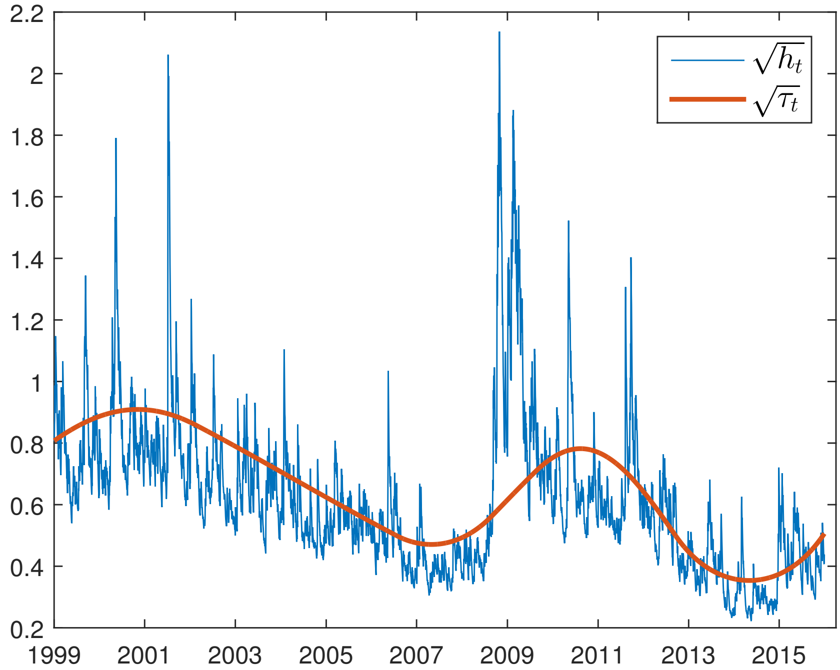

Figure 6: Spline-GARCH on Polish Zloty to Euro exchange rate returns in the period 1999-2015 based on daily closing prices. The knots of the splines are selected with Iterated Cumulative Sum of Squares approach (SANSÓ, ARAGÓ, and CARRION, 2004). The data is based on the work of WALTHER et al. (2017).

2.2 Estimation and Model Selection

The parameters in GARCH models can be estimated by various means: Ordinary Least Squares (ENGLE, 1982); Bayesian or Monte Carlo Estimation (GEWEKE, 1989); Whittle Estimation (GIRAITIS and ROBINSON, 2001); Least Absolute Deviation (PENG, 2003); and even a closed-form estimator (KRISTENSEN and LINTON, 2006). However, the most

[13] Actually, WALTHER et al. (2017) use a Spline-FIGARCH.

prominent method to estimate the parameters of GARCH models is the Maximum Likelihood Estimation (MLE). In what follows, the MLE estimation for GARCH models is presented.[14]

Given the information set Ω_{t-1} and the assumption that $(\varepsilon_t)_{t\in\mathbb{Z}}$ are i.i.d., the conditional variance $(h_t(\theta))_{t\in\mathbb{Z}}$ with the parameter vector θ, e.g. $\theta = (\omega, \alpha, \beta)'$ in the case of GARCH(1,1)[15], the conditional likelihood function can be written:

$$\mathcal{L}(\theta) = \prod_{t=1}^{T} \ell_t(\theta|\Omega_{t-1}),$$

where $\ell_t(\theta|\Omega_{t-1})$ is the conditional likelihood (equal to the conditional density function $f_t(\theta|\Omega_{t-1})$). In case of a Normal distribution of ε_t, the conditional likelihood is

$$\ell_t(\theta|\Omega_{t-1}) = \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{\varepsilon_t^2}{2h_t}\right). \tag{13}$$

In practice, however, the conditional log-likelihood function is used:

$$\begin{aligned}\log \mathcal{L}(\theta) &= \sum_{t=1}^{T} \log \ell_t(\theta|\Omega_{t-1}) \\ &= \sum_{t=1}^{T} \left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log h_t - \frac{\varepsilon_t^2}{2h_t}\right).\end{aligned} \tag{14}$$

The parameter estimate $\hat{\theta}$ is obtained by maximisation of the log-likelihood function:

$$\hat{\theta} = \underset{\theta\in\Theta}{\arg\max} \log \mathcal{L}(\theta),$$

where Θ is the admissible parameter space with regards to non-negativity and stationarity conditions. Since the calculation of $(h_t(\theta))_{t\in\mathbb{Z}}$ includes the values h_t and ε_0^2 for $t \leq 0$, presample values are needed. ENGLE and BOLLERSLEV (1986b, p. 24) and BOLLERSLEV (1986, p. 316) suggests to use the sample mean $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2$.

A prerequisite to use the MLE is that the underlying model is the "true" model. As stated above, especially financial data is not Normally distributed. Hence, the model is misspecified when using Eq. (13) and (14), which leads to inconsistent estimators and biased standard errors (WHITE, 1982). Therefore, the use of the Quasi Maximum-Likelihood Estimation (QMLE) is suggested, which applies under certain conditions even if the model is misspecified. The MLE and QMLE only differ in a robust covariance

[14] The description of the MLE is similar to the one presented in LOCAREK-JUNGE, KLEIN, and WALTHER (2014, pp. 1350f.).

[15] Note that the parameter indices for first-order GARCH specification, e.g. GARCH(1,1), are left out for the sake of simplicity. Thus, α_1 is denoted as α etc. Moreover, prime denotes transposition. Hence, θ is a column vector.

matrix for the parameter estimates (BOLLERSLEV, 2010, p. 158).

In case of the correct model, the covariance matrix for the estimator can be obtained either from the Outer-Product (first-order derivative) D_T^{-1}/T with

$$D_T = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial \log \ell_t(\theta)}{\partial \theta}\frac{\partial \log \ell_t(\theta)}{\partial \theta'}\right),$$

or the Hessian (second-order derivative) form H_T^{-1}/T with

$$H_T = -\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2 \log \ell_t(\theta)}{\partial \theta \partial \theta'}\right).$$

For the correct model, both should be the same. When the model is assumed to be misspecified, one can obtain robust standard errors by using the so-called sandwich estimator for the covariance matrix from BOLLERSLEV and WOOLDRIDGE (1992, pp. 148f.) i.e.

$$D_T^{-1} H_T D_T^{-1}/T.$$

The standard errors for $\hat{\theta}$ are the square root of the diagonal elements of the covariance estimator (MCNEIL, FREY, and EMBRECHTS, 2015, pp. 124-127 and RUPPERT and MATTESON, 2015, pp. 104-107).

The likelihood $\ell_t(\theta|\Omega_{t-1})$ can be chosen to better fit the empirical data, e.g. to account for fat tails. One possibility is to use the density function of the standardised Student-t distribution (BOLLERSLEV, 1987, p. 543 and TSAY, 2013, pp. 189f.):

$$\ell_t(\theta|\Omega_{t-1}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\left(\nu-2\right)h_t}}\left(1+\frac{\varepsilon_t^2}{\left(\nu-2\right)h_t}\right)^{-(\nu+1)/2},$$

where $\Gamma(\cdot)$ is the Gamma function

$$\Gamma\left(x\right) = \int_0^\infty y^{x-1}\exp\left(-y\right)\mathrm{d}y\,,$$

and ν is the degree of freedom, which can be estimated along with the rest of the parameters, i.e. for GARCH(1,1): $\theta = (\omega, \alpha, \beta, \nu)'$. Instead of Eq. (14) it follows:

$$\begin{aligned}\log \mathcal{L}\left(\theta\right) = T&\left(\log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log\left(\pi\left(\nu-2\right)\right)\right)\\ &-\frac{1}{2}\sum_{t=1}^{T}\left(\log h_t + \left(\nu+1\right)\log\left(1+\frac{\varepsilon_t^2}{\left(\nu-2\right)h_t}\right)\right).\end{aligned}$$

The QMLE works for the models presented in Sec. 2.1.1-2.1.4 and 2.1.7. In case of MRS- and Mixture GARCH models, the series $(S_t)_{t\in\mathbb{Z}}$ is not observable. Hence, one needs to calculate the $R \times 1$ conditional probability vector

$$\hat{\xi}_{t|s} = \begin{bmatrix} \mathbb{P}[S_t = 1|\theta; \Omega_s] \\ \mathbb{P}[S_t = 2|\theta; \Omega_s] \\ \vdots \\ \mathbb{P}[S_t = R|\theta; \Omega_s] \end{bmatrix}.$$

HAMILTON (1994, pp. 690-696) suggests to derive the state probabilities iteratively by

$$\begin{aligned} \hat{\xi}_{t|t} &= \frac{\left(\hat{\xi}_{t|t-1} \odot \eta_t\right)}{\mathbf{1}'\left(\hat{\xi}_{t|t-1} \odot \eta_t\right)}, \\ \hat{\xi}_{t+1|t} &= \mathbf{P}\hat{\xi}_{t|t}, \end{aligned}$$

where η_t is the R-dimensional vector of the conditional density functions

$$\eta_t = \begin{bmatrix} f_t\left(\theta|S_t = 1; \Omega_{t-1}\right) \\ f_t\left(\theta|S_t = 2; \Omega_{t-1}\right) \\ \vdots \\ f_t\left(\theta|S_t = R; \Omega_{t-1}\right) \end{bmatrix},$$

and $\odot$ is the element-wise multiplication operator. The log-likelihood is obtained as a by-product of this algorithm with

$$\begin{aligned} \log \ell_t\left(\theta|\Omega_{t-1}\right) &= \log\left(\mathbf{1}'\left(\hat{\xi}_{t|t-1} \odot \eta_t\right)\right), \\ &= \log \sum_{i=1}^{R} P_{t,S_t=i} f_t\left(\theta|S_t = i; \Omega_{t-1}\right). \end{aligned}$$

Another possibility is the so-called Expectation-Maximisation (EM) algorithm (DEMPSTER, LAIRD, and RUBIN, 1977). Based on starting parameters $\hat{\theta}^{(0)}$ a first expectation for $\hat{\xi}^{(1)}_{t|t-1}$ is calculated. This expectation is used to estimate the parameters $\hat{\theta}^{(1)}$ by maximising the log-likelihood function. However, since the data is incomplete (the regimes are not observable), the log-likelihood is replaced by an expected log-likelihood:

$$\hat{\theta}^{(1)} = \underset{\theta \in \Theta}{\arg\max} \log \mathcal{L}^*, \tag{15}$$

$$\log \mathcal{L}^* = \sum_{t=1}^{T} \sum_{i=1}^{R} \hat{\xi}^{(1)}_{t|t-1} \log\left(\mathbb{P}[S_t = i|\theta; \Omega_{t-1}] f_t\left(\theta|S_t = i; \Omega_{t-1}\right)\right). \tag{16}$$

The second expectation step uses $\hat{\theta}^{(1)}$ and so on. The algorithm stops, when $\hat{\theta}^{(k)} \approx \hat{\theta}^{(k-1)}$ (HAMILTON, 1990, pp. 46-51 and KLEIN and WALTHER, 2016, pp. 48f.).

Once, the model parameters are estimated, one can compare the goodness-of-fit. Popular measures are the Akaike Information Criterion (AIC, AKAIKE, 1974, p. 719) and the Bayesian Information Criterion (BIC, SCHWARZ, 1978, p. 461):

$$\begin{aligned}\text{AIC} &= -2\log\mathcal{L} + 2n,\\ \text{BIC} &= -2\log\mathcal{L} + n\log T,\end{aligned}$$

where n is the number of parameters of a specific model. When comparing two models, the model with the lower AIC or BIC has the better goodness-of-fit. This procedure can also be exercised to identify e.g. the lag-order p and q of GARCH(p,q) models or the number of splines k in the Spline(k)-GARCH as suggested for model selection by BOX, JENKINS, and REINSEL (2008, pp. 211f.) for ARMA models.

2.3 Forecasting

In this section, the forecasting or prediction with GARCH models is reviewed. Generally, there are two cases that are considered: (1) one-period ahead and (2) multi-periods ahead. The latter can be additionally subdivided into point or accumulated volatility forecast.

For GARCH(1,1), the one-period ahead variance forecast $\mathbb{E}[h_{T+1}|\Omega_T] = \hat{h}_{T+1}$ is trivial. Given all information Ω_T and the estimated parameters $\hat{\theta}$, the Eq. (4) can be used, i.e.

$$\hat{h}_{T+1} = \hat{\omega} + \hat{\alpha}\varepsilon_T^2 + \hat{\beta}h_T. \tag{17}$$

For the 2-periods ahead, the equation can be formulated as

$$\hat{h}_{T+2} = \hat{\omega} + \hat{\alpha}\varepsilon_{T+1}^2 + \hat{\beta}h_{T+1}.$$

Since ε_{T+1}^2 and h_{T+1} are unknown, they can be substituted by their conditional expectation, i.e.

$$\hat{h}_{T+2} = \hat{\omega} + \hat{\alpha}\mathbb{E}[\varepsilon_{T+1}^2|\Omega_T] + \hat{\beta}\hat{h}_{T+1}.$$

Given that $\mathbb{E}[\varepsilon_{T+1}^2|\Omega_T] = \hat{h}_{T+1}$, it follows

$$\hat{h}_{T+2} = \hat{\omega} + \left(\hat{\alpha} + \hat{\beta}\right)\hat{h}_{T+1},$$

where $\hat{h}_{T+1}$ can be substituted by Eq. (17):

$$\begin{aligned}\hat{h}_{T+2} &= \hat{\omega} + \left(\hat{\alpha} + \hat{\beta}\right)\left(\hat{\omega} + \hat{\alpha}\varepsilon_T^2 + \hat{\beta}h_T\right)\\ &= \hat{\omega} + \hat{\omega}\left(\hat{\alpha} + \hat{\beta}\right) + \left(\hat{\alpha} + \hat{\beta}\right)\left(\hat{\alpha}\varepsilon_T^2 + \hat{\beta}h_T\right).\end{aligned}$$

The s-period ahead prediction, for $s \geq 3$, is obtained by further recursive substitution:

$$\begin{aligned}
\hat{h}_{T+s} &= \hat{\omega} + \left(\hat{\alpha} + \hat{\beta}\right) \hat{h}_{T+s-1} \\
&= \hat{\omega} + \left(\hat{\alpha} + \hat{\beta}\right) \left(\hat{\omega} + \left(\hat{\alpha} + \hat{\beta}\right) \hat{h}_{T+s-2}\right) \\
&= \hat{\omega} + \hat{\omega} \left(\hat{\alpha} + \hat{\beta}\right) + \left(\hat{\alpha} + \hat{\beta}\right)^2 \hat{h}_{T+s-2} \\
&\dots \\
&= \hat{\omega} \sum_{i=0}^{s-2} \left(\hat{\alpha} + \hat{\beta}\right)^i + \left(\hat{\alpha} + \hat{\beta}\right)^{s-1} \hat{h}_{T+1} \\
&= \hat{\omega} \sum_{i=0}^{s-1} \left(\hat{\alpha} + \hat{\beta}\right)^i + \left(\hat{\alpha} + \hat{\beta}\right)^{s-1} \left(\hat{\alpha}\varepsilon_T^2 + \hat{\beta} h_T\right).
\end{aligned} \tag{18}$$

From Eq. (18), it is obvious that for $s \to \infty$, $\hat{h}_{T+s} \to \frac{\hat{\omega}}{1-\hat{\alpha}-\hat{\beta}}$, provided that $\hat{\alpha} + \hat{\beta} < 1$, which coincides with the unconditional variance and demonstrates the mean-reverting property of the model (TSAY, 2013, pp. 200f. and MCNEIL, FREY, and EMBRECHTS, 2015, pp. 130f.).

Forecasting with asymmetric GARCH models (Sec. 2.1.3) is a bit more complex and often depends on the underlying distributional assumption due to the conditional expectations. TSAY (2013, pp. 220f.) provides the s-period ahead forecast for EGARCH(1,1) with Normal distribution. In order to do so, the Eq. (5) needs to be transformed to

$$\begin{aligned}
h_t &= \exp\left(\omega + g\left(z_{t-1}\right) + \beta \log h_{t-1}\right) \\
&= \exp\left(\omega\right) \exp\left(g\left(z_{t-1}\right)\right) h_{t-1}^{\beta},
\end{aligned}$$

with $g\left(z_t\right) = \gamma z_t + \alpha \left(|z_t| - \sqrt{2/\pi}\right)$, since h_{T+s} and not $\log h_{T+s}$ is to be forecasted.[16] Thus, for the one-period ahead prediction the equation is

$$\hat{h}_{T+1} = \exp\left(\hat{\omega}\right) \exp\left(\hat{g}\left(z_T\right)\right) h_T^{\hat{\beta}},$$

where all data is known after estimation at time T. Any further forecast needs the expec-

[16] With Jensen's inequality it follows that $\exp\left(\mathbb{E}[\log h_{T+s}]\right) \leq \mathbb{E}[\exp\left(\log h_{T+s}\right)]$.

tation of $\exp\left(g\left(z_t\right)\right)$, i.e.

$$\begin{aligned}
\mathbb{E}\left[\exp\left(g\left(z_t\right)\right)\right] =& \mathbb{E}\left[\exp\left(\gamma z_t + \alpha\left(\left|z_t\right| - \sqrt{2/\pi}\right)\right)\right] \\
=& \int_{-\infty}^{\infty} \exp\left(\gamma z_t + \alpha\left(\left|z_t\right| - \sqrt{2/\pi}\right)\right) \varphi\left(z_t\right) \mathrm{d}z_t \\
=& \exp\left(-\alpha\sqrt{2/\pi} + \frac{\left(\gamma+\alpha\right)^2}{2}\right) \Phi\left(\gamma+\alpha\right) \\
&+ \exp\left(-\alpha\sqrt{2/\pi} + \frac{\left(\gamma-\alpha\right)^2}{2}\right) \Phi\left(\gamma-\alpha\right),
\end{aligned}$$

with $\varphi(\cdot)$ and $\Phi(\cdot)$ as the probability density function and cumulative distribution function of the standard Normal distribution, respectively. Hence, the two-period and s-period, for $s \geq 3$, ahead forecasts are:

$$\begin{aligned}
\hat{h}_{T+2} &= \exp\left(\hat{\omega}\left(1+\hat{\beta}\right) + \hat{\beta}\hat{g}\left(z_T\right)\right) h_T^{\hat{\beta}^2} \mathbb{E}\left[\exp\left(\hat{g}\left(z_t\right)\right)\right], \\
\hat{h}_{T+s} &= \exp\left(\hat{\omega}\sum_{i=0}^{s} \beta^i + \hat{\beta}^{s-1}\hat{g}\left(z_T\right)\right) h_T^{\hat{\beta}^s} \mathbb{E}\left[\exp\left(\hat{g}\left(z_t\right)\right)\right]^{\sum_{i=0}^{s-2}\beta^i}.
\end{aligned}$$

The prediction for GJR-GARCH follows the one for the normal GARCH in Eq. (18). For a symmetrical distribution, $\mathbb{E}[\varepsilon_t^2|\varepsilon_t < 0; \Omega_{t-1}] = \frac{1}{2}h_t$. Consequently the s-period ahead forecast is

$$\hat{h}_{T+s} = \hat{\omega}\sum_{i=0}^{s-1}\left(\hat{\alpha} + \hat{\gamma}/2 + \hat{\beta}\right) + \left(\hat{\alpha} + \hat{\gamma}/2 + \hat{\beta}\right)^{s-1}\left(\hat{\alpha}\varepsilon_T^2 + \hat{\gamma}I_{\{\varepsilon_{t-1}<0\}}\varepsilon_T^2 + \hat{\beta}h_T\right).$$

For the APARCH(1,1) forecast with Normal innovations, it is referred to KLEIN and WALTHER (2016, p. 49).

To forecast long memory GARCH models (Sec. 2.1.4), the ARCH(∞) representation is used:

$$\hat{h}_{T+s} = \frac{\hat{\omega}}{1-\hat{\beta}} + \sum_{i=1}^{\infty}\hat{\lambda}_i\varepsilon_{T+s-i}^2.$$

For $i = 1, \ldots, s-1$ the squared residuals are unknown and must be replaced by their conditional expectation:

$$\hat{h}_{T+s} = \frac{\hat{\omega}}{1-\hat{\beta}} + \sum_{i=1}^{s-1}\hat{\lambda}_i\hat{h}_{T+s-i} + \sum_{i=s}^{\infty}\hat{\lambda}_i\varepsilon_{T+s-i}^2.$$

By iteratively calculating $\hat{h}_{T+1}, \hat{h}_{T+2}, \ldots, \hat{h}_{T+s-1}$, the prediction for $\hat{h}_{T+s}$ is estimated. In practice, the infinite sum needs to be truncated. In most applications, a truncation lag of $1,000$ is common (KLEIN and WALTHER, 2017).

MRS- and Mixture GARCH models follow their single regime counterparts. The only difference is that a forecast for the regime/component probabilities has to be drawn. In case of the MRS models, HAMILTON (1994, p. 694) shows that

$$\hat{\xi}_{t+s|T} = \mathbf{P}^s \hat{\xi}_{t|T}.$$

The best guess for Mixture models, however, is, to simply use the probabilities at time T for forecasts to $T+s$.

Lastly, in multiplicative Component GARCH models, the expectation for the long-term component is given in Eq. (12). Thus, only the short-term component needs to be forecasted and is equivalent to the various GARCH models described above with the simple exception that the unconditional variance for the short-term component is 1. Hence, for a GARCH(1,1) the forecast is

$$\hat{h}_{T+s} = \tau_T \left((1 - \hat{\alpha} - \hat{\beta}) \sum_{i=0}^{s} \left(\hat{\alpha} + \hat{\beta} \right)^i + \left(\hat{\alpha} + \hat{\beta} \right)^s g_T \right).$$

The above mentioned procedures yield point forecasts, i.e. the variance at time $T+s$. However, in some cases, the econometrician wants to have an aggregated forecast, i.e. the variance for the period $T+1$ to $T+s$. For homoscedastic frameworks with symmetric error distribution, the rule-of-square-root usually applies. The weekly volatility is simply $\sigma^{(w)} = \sqrt{5}\sigma^{(d)}$ for five trading days and $\sigma^{(d)}$ as the daily volatility. In a GARCH framework, one has to sum up the daily variance point forecasts (POON, 2005, p. 16):

$$\hat{h}^{(w)}_{T+1:T+5} = \sum_{i=1}^{5} \hat{h}^{(d)}_{T+i},$$

with the weekly volatility $\sqrt{\hat{h}^{(w)}_{T+1:T+5}}$.

To evaluate the forecast accuracy, loss functions in an out-of-sample exercise are used. The sample is divided into a training data set of length N with $t = 1, \ldots, N$ and a test data set with length M where $t = N+1, \ldots, N+M$. One estimates the model's parameter from the training data set and makes predictions for the realisation in the test data set. Afterwards, the predictions and the observations are compared using loss functions. A variety of loss function is presented in HANSEN and LUNDE (2005, pp. 877), POON (2005, pp. 23f.), and PATTON (2011, p. 248). However, the most common ones are the above mentioned RMSE

$$\text{RMSE} = \sqrt{\sum_{i=1}^{M} \left(\hat{h}_{N+i} - h_{N+i} \right)^2},$$

and the mean absolute error (MAE)

$$\text{MAE} = \sum_{i=1}^{M} \left| \hat{h}_{N+i} - h_{N+i} \right| ,$$

where h_{N+i} is the realised variance to compare the forecast $\hat{h}_{N+i}$ with. Since just the observation r_{N+i} and not the realised variance is observable, proxies have to be used. A frequently utilised proxy is the squared observation r_t^2 (e.g. daily squared return), even though it is widely known that it inherits a lot of noise. Therefore, observations at a higher frequency can be combined to build a proxy for the wanted frequency (e.g. accumulated intra-day squared returns for daily variance) (ANDERSEN and BOLLERSLEV, 1998).

After calculating the loss function for several models, the model with the lowest loss function yields the best performance. Nonetheless, another problem arises. Using the same data for different models makes it more likely that the results are driven by chance rather then the superiority in forecast of one model (WHITE, 2000, p. 1098). In order to identify the models with the best forecasting performance, multiple tests exist to circumvent the so-called data-snooping problem. DIEBOLD and MARIANO (1995) propose a test for equal predictive ability. The tests of WHITE (2000) and HANSEN (2005), however, test for superior predictive ability, i.e. the null hypothesis is that the model of interest is not inferior to its peers. The aforementioned tests are all constructed in a way that benchmark models are needed to compare the other models with. HANSEN, LUNDE, and NASON (2011) suggests the Model Confidence Set to extract models of equal superiority out of a choice of forecasting models.

3 Risk Measures with GARCH Models

This chapter reviews methodologies to estimate the shortfall risk measures VaR and ES. Special focus is set on possibilities to forecast the VaR and ES using GARCH models. Moreover, popular back test methodologies are presented.

3.1 Estimating Value-at-Risk & Expected Shortfall

VaR and ES are so-called shortfall risk measures, as they intend to describe risk as a negative deviation from a base scenario. In contrast, e.g. the standard deviation is a symmetric risk measure. Financial institutions and regulators use VaR and ES for various purposes. According to JORION (2007, p. 380), VaR has the following three main applications: to report risk, to control risk, and to allocate risk. Within the regulatory framework of the BASEL COMMITTEE ON BANKING SUPERVISION (BCBS, 2016), VaR and ES are utilised to set minimum capital requirements for financial institutions.

The VaR is the minimum loss that occurs at a given confidence level $(1-a)$ over a given period of time. Formally, the VaR can be defined as:

$$\text{VaR}_{1-a} = \inf\{x|F(x) \geq 1-a\}, \tag{19}$$

where $F(\cdot)$ is the cumulative distribution function of the returns.[17] The right hand side of Eq. (19) can be expressed as the $(1-a)$-quantile F^{-1} of the distribution F, i.e.

$$\text{VaR}_{1-a} = F^{-1}(1-a).$$

For the Normal distribution with mean μ and standard deviation σ, the VaR is

$$\text{VaR}_{1-a} = \mu + \sigma\Phi^{-1}(1-a).$$

Alternatively, for the Student-t distribution with $\nu > 2$, the VaR is given as

$$\text{VaR}_{1-a} = \mu + \sigma F_{\text{t}}^{-1}(1-a, \nu),$$

where F_{t}^{-1} is the quantile function of the Student-t distribution with ν degrees of freedom. In practice, 95%, 97.5%, or 99% are used for $1-a$. In case of the Normal distribution, the 99% quantile is approximately 2.3263. For the Student-t distribution with $\nu = 3$ the 99% quantile is 2.6065 and with $\nu = 4$ it is 2.6495. Thus, the Student-t distribution provides

[17] Note that most literature defines VaR based on a general loss variable. If the VaR is defined for the return of an asset x is either $-r_t$ or r_t, depending on the trader's position (either long or short).

Distribution	**95%**	**97.5%**	**99%**
Standard Normal	1.6449	1.9600	2.3263
Student-t ($\nu = 3$)	1.3587	1.8374	2.6216
Student-t ($\nu = 4$)	1.5074	1.9632	2.6495

Table 2: Overview of quantiles for different distributions.

heavier tails. An overview for other important quantiles is provided in Tab. 2. To estimate the VaR by means of GARCH models, the unconditional mean μ and variance σ^2 are replaced by their conditional complements μ_t and h_t (TSAY, 2013, pp. 329-334).

Figure 7 depicts a comparison of the 99% VaR with Normal and Student-t distribution with $\mu = 0$ and $\sigma = 1$. It can be seen that the Student-t distribution with $\nu = 4$ has a higher kurtosis and "fatter tails", i.e. observations are more concentrated to the centre and more probability is shifted to the extremes.

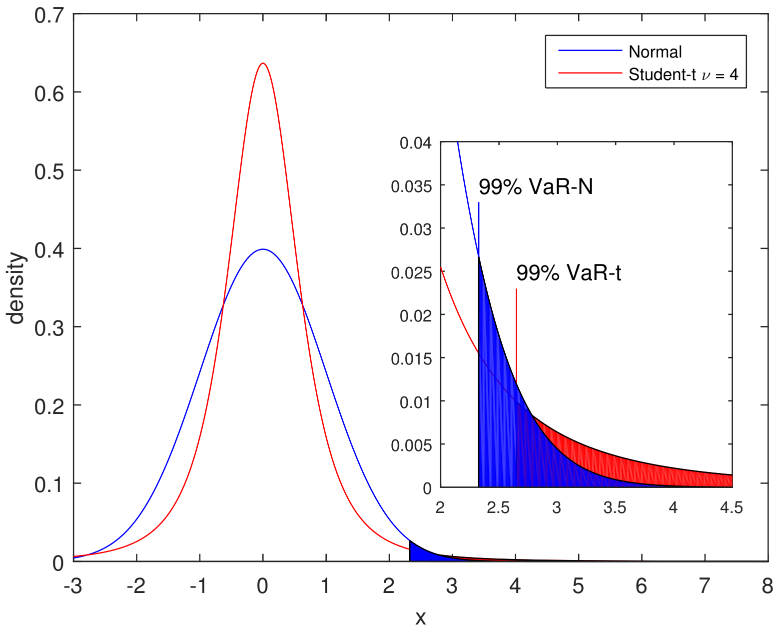

Figure 7: Comparison of Value-at-Risk with Normal and Student-t distribution.

Another important risk measure is the ES. ARTZNER et al. (1999, pp. 208-210) define four criteria for risk measures in order to be *coherent*, i.e. monotonicity, positive homogeneity, translation invariance, and sub-additivity. While VaR fulfils the first three axioms, it violates the sub-additivity in some cases. Furthermore, VaR only represents a certain threshold which is not exceeded at a given confidence level, while the ES provides a measure of the expected loss, once this threshold is violated. MCNEIL, FREY, and EMBRECHTS (2015, pp. 69f.) define the ES for continuous distributions by

$$\begin{aligned} \mathrm{ES}_{1-a} &= \frac{1}{a} \int_{1-a}^{1} F^{-1}(u)\,\mathrm{d}u \\ &= \frac{1}{a} \int_{1-a}^{1} \mathrm{VaR}_u \mathrm{d}u. \end{aligned}$$

In Figure 7, the ES is the expected value of the filled areas for the corresponding distributions. Some literature refer to ES also as Conditional VaR (e.g. ROCKAFELLAR and URYASEV, 2002).[18]

In order to retrieve closed-form expressions for the ES, the distribution function of x must be known. For the Normal distribution, the ES is

$$\text{ES}_{1-a} = \mu + \frac{\varphi\left(\Phi^{-1}\left(1-a\right)\right)}{a}\sigma,$$

and for the Student-t distribution

$$\text{ES}_{1-a} = \mu + \frac{f_{\text{t}}\left(F_{\text{t}}^{-1}\left(1-a,\nu\right),\nu\right)}{a}\left(\frac{\nu + \left(F_{\text{t}}^{-1}\left(1-a,\nu\right)\right)^2}{\nu - 1}\right)\sigma,$$

where f_{t} is the probability density function of the Student-t distribution (TSAY, 2013, pp. 334-336 and MCNEIL, FREY, and EMBRECHTS, 2015, pp. 70f.).

The presented forms to estimate the VaR and ES are not limited to these cases. To get an estimate for both risk measures, the distribution of the loss variable has to be obtained by some means, to derive the quantile. A very popular way to do so is the historical simulation, where the quantile is taken from the empirical distribution of former realisations of the loss variable. However, the historical simulation has two main drawbacks: (1) the results are very sensitive to the chosen timespan of data; (2) the scenarios are limited to cases which happen in the past (BEST, 1998, pp. 34-38). The Monte-Carlo simulation overcomes these shortcomings by drawing random scenarios from a pre-specified distribution. Obviously, choosing the "right" distribution is not an easy task, given the range of stylised facts (JORION, 2007, pp. 265-268, 307-329).

Besides the aforementioned approaches, literature offers several other possibilities to estimate the VaR and the ES: e.g. Mixture Densities using Neural Networks (LOCAREK-JUNGE and PRINZLER, 1998), filtered historical simulation (HULL and WHITE, 1998 and BARONE-ADESI, GIANNOPOULOS, and VOSPER, 1999), extreme value theory (MCNEIL and FREY, 2000 and HERRERA and SCHIPP, 2013), and conditional auto-regressive VaR (ENGLE and MANGANELLI, 2004). Recent literature proposes expectile regression to determine VaR (KUAN, YEH, and HSU, 2009) and ES (TAYLOR, 2007).

To illustrate the VaR and ES forecast, Figure 8 shows estimated values for the WTI between 2010 and 2015 for long and short trading positions. The data is taken from KLEIN and WALTHER (2016). The estimates are obtained from forecasting GARCH with Normal distribution one day ahead. A 99% VaR forecast for the given period of 1,261 days should have about 13 violations, i.e. returns that exceed the VaR. Here, the short trading position counts four hits and the long trading position 18 hits. Thus, the short trading position is

[18] Additionally, ES is also called Average VaR, Tail VaR, and Conditional Tail Expectation. Confusingly, these names also refer to slightly different definitions, e.g. $\mathbb{E}\left[x|x \geq \text{VaR}_{1-a}\right]$ (HUSCHENS, 2017, pp. 83-86).

modelled too conservatively and the long trading position could be improved. Moreover, many violations even exceed the estimated ES. Clearly, modelling the tails of the distribution must be improved, e.g. by using fat tailed distributions. The next section reviews methods to evaluate estimated VaR and ES.

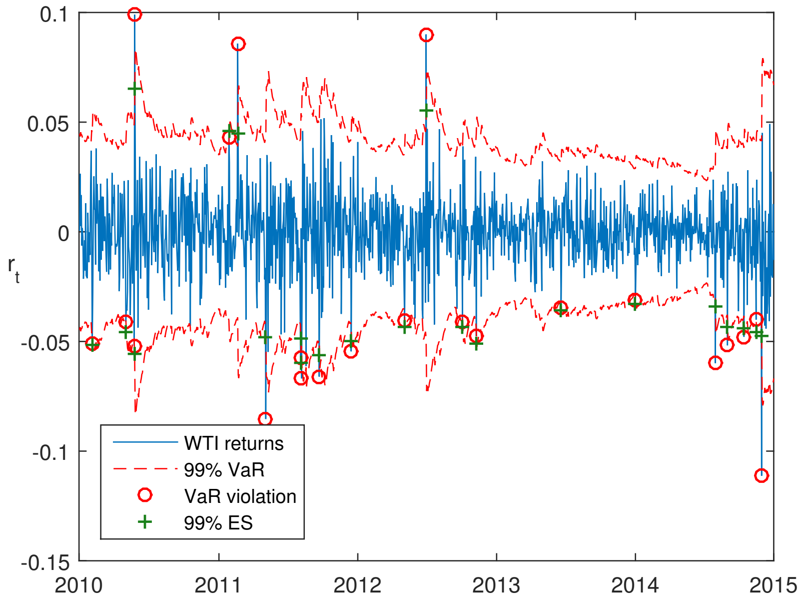

Figure 8: Daily Value-at-Risk and Expected Shortfall estimations for WTI in period 2010-2015 using GARCH with Normal distribution. Data is retrieved from KLEIN and WALTHER (2016).

3.2 Back Testing

To measure the performance of the various means to estimate the VaR and the ES, *back tests* have to be conducted. Therefore, the same framework as for the evaluation of variance estimates (see Sec. 2.3) can be used, i.e. using the training data to forecast the risk measures for the out-of-sample period. The actual realisations in the out-of-sample period are used to obtain test statistics.

The BASLE COMMITTEE ON BANKING SUPERVISION (1996) advocates a very simple approach based on the binomial probability of the 99% VaR. To distinguish between erroneously rejected and accepted models, the BCBS set three traffic light colour zones, i.e. green, yellow, and red. Banks have to back test their internal models on a daily basis for the last 250 trading days ($M = 250$). If the bank's losses exceed the 99% VaR not more than four times during that period, the model is considered to be in the green zone and presumed accurate. For four to nine exceptions, a model is placed in the yellow zone. Depending on the number of exceptions the supervisor of the bank can increase the bank's scaling factor for capital requirements. Lastly, the red zone indicates models that have at least ten exceptions in the out-of-sample period. Since the probability of erroneously rejected models (e.g. due to bad luck) is very low, the supervisor will increase the bank's scaling factor by one point and may forbid the usage of the model.

More sophisticated back testing approaches for VaR are reviewed by PIONTEK (2010, p. 482). The author classifies existing approaches into three groups of VaR back tests: (1) tests based on the frequency of failures, (2) tests based on the distribution, and (3) tests based on loss functions. Here, only examples for the first class of back tests are presented. See also CHRISTOFFERSEN (2010) for an overview.

One of the first tests based on the frequency of failure is proposed by KUPIEC (1995). A series of VaR_{1-a} violations $\left(I_t\left(a\right)\right)_{t\in\mathbb{Z}}$ is defined by

$$I_t(a) = \begin{cases} 1 & \text{if } \mathrm{VaR}_{1-a,t} \leq x_t \\ 0 & \text{if } \mathrm{VaR}_{1-a,t} > x_t. \end{cases} \tag{20}$$

The unconditional coverage tests of KUPIEC (1995, p. 79) is a Log-Likelihood ratio and compares the two binomial likelihoods of the level a with the actually level

$$a^* = \frac{M^*}{M}$$

with

$$M^* = \sum_{i=1}^{M} I_{N+i}.$$

The test statistic is

$$\mathrm{Kup}_a = 2\log\left(\frac{\left(1-a^*\right)^{M-M^*}\left(a^*\right)^{M^*}}{\left(1-a\right)^{M-M^*}a^{M^*}}\right),$$

and is asymptotically χ^2 distributed with one degree of freedom. Hence, critical values for the null hypothesis H_0: $a = a^*$ are 2.7055, 3.8415, and 6.6349 for 10%, 5%, and 1% level of significance, respectively.

KUPIEC tests whether a model to estimate the VaR has the wanted coverage over a specified time period (out-of-sample). The test suggested by CHRISTOFFERSEN (1998) is concerned with the fact that VaR violations might cluster as the volatility does. Thus, a good VaR model yields a wanted coverage ratio as well as independent violations. The independence part of the null hypothesis stands against a first order Markov chain as the alternative. The test statistic for the conditional coverage reads as follows

$$\mathrm{Chr}_a = 2\log\left(\frac{\left(\frac{n_{00}}{n_{00}+n_{01}}\right)^{n_{00}}\left(\frac{n_{01}}{n_{00}+n_{01}}\right)^{n_{01}}\left(\frac{n_{10}}{n_{10}+n_{11}}\right)^{n_{10}}\left(\frac{n_{11}}{n_{10}+n_{11}}\right)^{n_{11}}}{\left(1-a\right)^{M-M^*}a^{M^*}}\right),$$

where n_{ij} corresponds to the number of observations in I_t where the value i is followed by j. In particular, $M = n_{00} + n_{01} + n_{10} + n_{11}$ and $M^* = n_{10} + n_{11}$. The null hypothesis is rejected if the test statistics is greater than the critical values from the χ^2 distribution with two degrees of freedom (CHRISTOFFERSEN, 1998, pp. 845-847).

Alternatives to these two VaR tests are manifold. ZIGGEL et al. (2014) suggest to compare the unconditional and the conditional coverage tests with distributions drawn from a Monte-Carlo simulation. LOPEZ (1998) and SARMA, THOMAS, and SHAH (2003) propose loss function based tests, which also includes excess of the VaR violation. CRNKOVIC and DRACHMAN (1996), DIEBOLD, GUNTHER, and TAY (1998), and BERKOWITZ (2001) provide tests based on the whole density instead of certain quantiles. The duration based approaches are related to the time between VaR violations (CHRISTOFFERSEN and PELLETIER, 2004 and CANDELON et al., 2011). Multi-level VaR back tests are provided by CAMPBELL (2006) and PÉRIGNON and SMITH (2008). Lastly, ENGLE and MANGANELLI (2004) suggest a dynamic quantile test.

Back testing frameworks for the ES are rather scarce compared to the variety of VaR tests. GNEITING (2011, p. 756) sees the explanation for that in the lack of the so-called elicitability. In brief, the property of elicitability states that a statistic minimises the expected value of a score function, e.g. the mean minimises the quadratic score (BELLINI and BIGNOZZI, 2015). This seems necessary in order to compare the forecasts of different models. While VaR possesses this property, ES does not (ZIEGEL, 2014). However, EMMER, KRATZ, and TASCHE (2015) provide that ES is conditional elicitable and that it is back testable in a two-step procedure. ACERBI and SZEKELY (2014) point out that elicitability is only important to compare models, but not to back test. Thus, the authors present three non-parametric back tests for the ES. The "direct ES" test is based on the joint evaluation of VaR and ES by combining the number and the size of VaR violations. The test statistic reads as follows

$$\mathrm{AS}_a = \frac{\sum_{t=1}^{M} \frac{x_t I_t(a)}{\mathrm{ES}_{1-a,t}}}{Ma} + 1,$$

where $I_t(a)$ refers to Eq. (20). The p-values can be drawn from a Monte-Carlo simulation (ACERBI and SZEKELY, 2014, pp. 3-6, 10). An appropriate model yields test statistics around 0. If $\mathrm{AS}_a < 0$ then the model has either to many or to high VaR violations. On the contrary, $\mathrm{AS}_a > 0$ indicates that the underlying model is too conservative. Other ES back tests are proposed by e.g. WONG (2008), MCNEIL, FREY, and EMBRECHTS (2015, pp. 354f.), and EMMER, KRATZ, and TASCHE (2015).

4 Conclusion

The aim of the essays of this thesis is to give further insight to stylised facts of financial time series, especially in the commodity, foreign exchange, and equity markets. A variety of GARCH models is presented incorporating the empirically observed properties aiming for more precise risk measures. To this end, 15 GARCH specifications with three different distributions yielding a total of 27 model-distribution combinations are employed to account for heavy tails, volatility clustering, the leverage effect, long memory, and structural breaks.[19]

In many cases, the most sophisticated model yields the best in-sample fit. However, in terms of goodness-of-fit, i.e. considering the trade-off between the number of parameters and fit, the most sophisticated model is not necessarily the best choice. Interestingly, most of the examined financial time series exhibit the stylised facts mentioned at the beginning, but to a different extent. In some cases these effects are even overlapping, e.g. the situation when structural breaks imitate the behaviour of long memory as shown for FX rates in WALTHER et al. (2017). These spurious effects need to be identified in order to avoid misspecification, which leads to biased forecasts.

In addition, forecasting risk measures for returns of financial assets must take into account the trading position. A trader, who is short (long) on an asset, suffers losses when the return is positive (negative). These positions refer to different tails of the distribution. It is often shown that using symmetrical distributions such as the Normal or the Student-t distribution does not account for the asymmetry of these positions. Thus, VaR and ES forecasts work really well for only one tail. On the opposite tail, the models fail to provide accurate measures (e.g. see Figure 8), however. Hence, using more flexible distributions, which allow for skewness, could help to overcome this shortcoming (KLEIN, PHAM THU, and WALTHER, 2016, pp. 136,138). Promising approaches are presented by HARVEY and SIDDIQUE (1999) and BALI, MO, and TANG (2008), who use time-varying conditional skewness in addition to GARCH models.

Moreover, GARCH models are not limited to the application of risk measurement. Other applications, especially in a multivariate context are asset pricing, portfolio selection & optimisation (BOUBAKER and SGHAIER, 2013), option pricing (DUAN, 1995), or hedging (MANSUR, COCHRAN, and SHAFFER, 2007). The most appealing property of multivariate GARCH models is that they allow to take correlations between time series in consideration. Time-varying modelling of correlations (ENGLE, 2002) or copula-based approaches (LEE and LONG, 2009) allow to cover volatility spillover effects from one asset into another and take non-linearities of co-movements into account, e.g. that seemingly

[19] See Appendix A for a complete list of models and distributions applied in each essay.

uncorrelated assets are highly correlated in stressed market situations.

An important question, which is not covered within this thesis, is: Why do financial returns fluctuate? The presented models can incorporate certain patterns to reflect empirical properties in the models, but it is only possible to answer the question whether or not a certain stylised fact is present. The causes for the stylised facts or the variance fluctuations cannot be observed. SCHWERT (1989) tries to answer this questions by using macroeconomic data to describe the variance of financial time series. However, due to the fact that macroeconomic measures are mostly published on a monthly or quarterly basis, it is difficult to explain daily volatility. Based on the above presented Spline(k)-GARCH, ENGLE, GHYSELS, and SOHN (2013) provide a model that allows to combine observations of different frequencies to describe a daily GARCH process. This mixed data sampling model class has the potential to offer more inside into causes of market fluctuations and should be set into focus of further research.

A Essay Overview

The following tables contain an overview of the six essays associated with this dissertation. It provides author and publication[20] details as well as a list of presentations at seminars and international conferences[21]. Moreover, the list contains the applied GARCH models and the underlying distributions in each essay.

No. 1	**Oil Price Volatility Forecast with Mixture Memory GARCH**
Authors	KLEIN, TONY; WALTHER, THOMAS
Year	2016
Publication	Energy Economics, Vol. 58, pp. 45-58 (VHB: *B*, SJR: *3.02*)
Presentations	• Energy Finance Conference, London, United Kingdom, 2015* • International Ruhr Energy Conference, Essen, Germany, 2015*
Models	GARCH, RiskMetrics, EGARCH, APARCH, FIGARCH, HYGARCH, FIAPARCH, MMGARCH
Distributions	Normal

No. 2	**Forecasting Volatility of Tanker Freight Rates Based on Asymmetric Regime-Switching GARCH Models**
Authors	LAUENSTEIN, PHILIPP; WALTHER, THOMAS
Year	2016
Publication	International Journal of Financial Engineering and Risk Management, Vol. 2, No. 3, pp. 172-199
Presentations	• HypoVereinsbank PhD Seminar, Halle, Germany, 2016 • Energy & Commodity Finance Conference, Paris, France, 2016
Models	GARCH, EGARCH, APARCH, MRS-GARCH, MRS-EGARCH, MRS-APARCH
Distributions	Normal

[20] Where applicable, the VHB-JOURQAL3 (http://vhbonline.org/VHB4you/jourqual/vhb-jourqual-3) and/or the SJR 2015 (http://www.scimagojr.com) rankings are provided.

[21] Presentations of co-authors are marked with *, † denotes presentations which were awarded best paper award, ‡ denotes presentations which were awarded certificate of appreciation (five best papers).

No. 3	**Evidence of Long Memory and Asymmetry in EUR/PLN Exchange Rate Volatility**
Authors	KLEIN, TONY; PHAM THU, HIEN; WALTHER, THOMAS
Year	2016
Publication	Research Papers of Wrocław University of Economics, No. 428, pp. 128-140
Presentations	• Science meets Social Science (S3), Wrocław University of Technology, Poland, 2015 • Wrocław Conference in Finance, Wrocław, Poland, 2015
Models	GARCH, APARCH, FIGARCH, FIAPARCH
Distributions	Normal, Student-t, Skewed Student-t

No. 4	**True or Spurious Long Memory in European Non-EMU Currencies**
Authors	WALTHER, THOMAS; KLEIN, TONY; PHAM THU, HIEN; PIONTEK, KRZYSZTOF
Year	2017
Publication	Research in International Business and Finance, Vol. 40C, pp. 217-230 (SJR: *0.43*)
Presentations	• HypoVereinsbank PhD Seminar, Leipzig, Germany, 2016 • Wrocław Conference in Finance, Wrocław, Poland, 2016 † • Macromodels International Conference, Łodz, Poland, 2016* ‡
Models	GARCH, FIGARCH, ICSS-FIGARCH, Spline-FIGARCH, ICSS-Spline-FIGARCH, Adaptive-FIGARCH
Distributions	Student-t

No. 5	**Expected Shortfall in the Presence of Asymmetry and Long Memory: An Application to Vietnamese Stock Markets**
Author	WALTHER, THOMAS
Year	2017
Publication	Pacific Accounting Review, Vol. 29, No. 2, pp. 132-151
Presentations	• Vietnam International Conference in Finance, Da Nang, Vietnam, 2016 • Joint Seminar on Finance, Wrocław, 2016
Models	GARCH, RiskMetrics, EGARCH, APARCH, FIGARCH, FIAPARCH
Distributions	Student-t, Skewed Student-t

No. 6	**Fast Fractional Differencing in Modeling Long Memory of Conditional Variance for High-Frequency Data**
Authors	KLEIN, TONY; WALTHER, THOMAS
Year	2017
Publication	Finance Research Letters, forthcoming (VHB: *B*, SJR: *0.41*)
Presentations	• Vietnam International Conference in Finance, Da Nang, Vietnam, 2016 • Statistische Woche, Augsburg, Germany, 2016 • Macromodels International Conference, Lodz, Poland, 2016 • HSC Seminar on Stochastic and Numerical Methods, Wrocław University of Technology, Poland, 2016* • Workshop of the German Operations Research Society (GOR e.V.), WG FIFI, Augsburg, Germany, 2016*
Models	FIGARCH, FIAPARCH
Distributions	Normal

Bibliography

ABRAMSON, ARI and COHEN, ISRAEL (2007): On the Stationarity of Markov-Switching GARCH Processes, in: Econometric Theory, Vol. 23, No. 03, pp. 485–500.

ACERBI, CARLO and SZEKELY, BALAZS (2014): Backtesting Expected Shortfall, in: Risk Magazine, pp. 76–81.

AKAIKE, HIROTUGU (1974): A new look at the statistical model identification, in: IEEE Transactionson Automatic Control, Vol. 19, No. 6, pp. 716–723.

ALOUI, CHAKER and JAMMAZI, RANIA (2009): The effects of crude oil shocks on stock market shifts behaviour: A regime switching approach, in: Energy Economics, Vol. 31, No. 5, pp. 789–799.

AMADO, CRISTINA and TERÄSVIRTA, TIMO (2013): Modelling volatility by variance decomposition, in: Journal of Econometrics, Vol. 175, No. 2, pp. 142–153.

ANDERSEN, TORBEN G. and BOLLERSLEV, TIM (1998): Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts, in: International Economic Review, Vol. 39, No. 4, pp. 885–905.

ARTZNER, PHILIPPE; DELBAEN, FREDDY; EBER, JEAN-MARC and HEATH, DAVID (1999): Coherent Measures of Risk, in: Mathematical Finance, Vol. 9, No. 3, pp. 203–228.

AVRAMOV, DORON; CHORDIA, TARUN and GOYAL, AMIT (2006): The impact of trades on daily volatility, in: Review of Financial Studies, Vol. 19, No. 4, pp. 1241–1277.

BAILLIE, RICHARD T.; BOLLERSLEV, TIM and MIKKELSEN, HANS OLE (1996): Fractionally integrated generalized autoregressive conditional heteroskedasticity, in: Journal of Econometrics, Vol. 74, No. 1, pp. 3–30.

BAILLIE, RICHARD T. and MORANA, CLAUDIO (2009): Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach, in: Journal of Economic Dynamics and Control, Vol. 33, No. 8, pp. 1577–1592.

BALI, TURAN G.; MO, HENGYONG and TANG, YI (2008): The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR, in: Journal of Banking and Finance, Vol. 32, No. 2, pp. 269–282.

BARONE-ADESI, GIOVANNI; GIANNOPOULOS, KOSTAS and VOSPER, LES (1999): VaR without correlations for portfolios of derivative securities, in: Journal of Futures Markets, Vol. 19, No. 5, pp. 583–602.

BASEL COMMITTEE ON BANKING SUPERVISION (2016): Minimum capital requirements for market risk, Technical Report January 2016. URL: www.bis.org/bcbs/publ/d352.pdf.

BASLE COMMITTEE ON BANKING SUPERVISION (1996): Supervisory Framework for the use of "Backtesting" in Conjunction With the Internal Models Approach to Market Risk Capital Requirements, Technical Report January 1996, Basle Committee on Banking Supervision. URL: www.bis.org/publ/bcbs22.pdf.

BAUWENS, LUC; HAFNER, CHRISTIAN and LAURENT, SÉBASTIEN (2012): Volatility Models, in: BAUWENS, LUC; HAFNER, CHRISTIAN and LAURENT, SÉBASTIEN (eds.), *Handbook of Volatility Models and Their Applications*, chapter 1, Hoboken, New Jersey: Wiley, pp. 1–45.

BEKAERT, GEERT and WU, GUOJUN (2000): Asymmetric Volatility and Risk in Equity Markets, in: Review of Financial Studies, Vol. 13, No. 1, pp. 1–42.

BELKHOUJA, MUSTAPHA and BOUTAHARY, MOHAMED (2011): Modeling volatility with time-varying FIGARCH models, in: Economic Modelling, Vol. 28, No. 3, pp. 1106–1116.

BELLINI, FABIO and BIGNOZZI, VALERIA (2015): On elicitable risk measures, in: Quantitative Finance, Vol. 15, No. 5, pp. 725–733.

BERKOWITZ, JEREMY (2001): Testing Density Forecasts, With Applications to Risk Management, in: Journal of Business & Economic Statistics, Vol. 19, No. 4, pp. 465–474.

BEST, PHILIP (1998): Implementing Value at Risk, Chichester: Wiley.

BOLLERSLEV, TIM (1986): Generalized autoregressive conditional heteroskedasticity, in: Journal of Econometrics, Vol. 31, No. 3, pp. 307–327.

BOLLERSLEV, TIM (1987): A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, in: The Review of Economics and Statistics, Vol. 69, No. 3, pp. 542–547.

BOLLERSLEV, TIM (2010): Glossary to ARCH (GARCH), in: BOLLERSLEV, TIM; RUSSELL, JEFFREY and WATSON, MARK W. (eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*, Oxford: Oxford University Press.

BOLLERSLEV, TIM and ENGLE, ROBERT F. (1993): Common Persistence in Conditional Variances, in: Econometrica, Vol. 61, No. 1, p. 167.

BOLLERSLEV, TIM and MIKKELSEN, HANS OLE (1996): Modeling and pricing long memory in stock market volatility, in: Journal of Econometrics, Vol. 73, No. 1, pp. 151–184.

BOLLERSLEV, TIM and WOOLDRIDGE, JEFFREY M. (1992): Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances, in: Econometric Reviews, Vol. 11, No. 2, pp. 143–172.

BOUBAKER, HENI and SGHAIER, NADIA (2013): Portfolio optimization in the presence of dependent financial returns with long memory: A copula based approach, in: Journal of Banking and Finance, Vol. 37, No. 2, pp. 361–377.

BOUGEROL, PHILIPPE and PICARD, NICO (1992): Stationarity of Garch processes and of some nonnegative time series, in: Journal of Econometrics, Vol. 52, No. 1-2, pp. 115–127.

BOX, GEORGE E. P.; JENKINS, GWILYM M. and REINSEL, GREGORY C. (2008): Time Series Analysis, Hoboken: Wiley, 4 edition.

CAI, JUN (1994): A Markov Model of Switching-Regime ARCH, in: Journal of Business & Economic Statistics, Vol. 12, No. 3, pp. 309–316.

CAMPBELL, JOHN Y. and HENTSCHEL, LUDGER (1992): No news is good news. An asymmetric model of changing volatility in stock returns, in: Journal of Financial Economics, Vol. 31, No. 3, pp. 281–318.

CAMPBELL, SEAN D. (2006): A Review of Backtesting and Backtesting Procedures, in: Journal of Risk, Vol. 9, No. 2, pp. 1–17.

CANDELON, BERTRAND; COLLETAZ, GILBERT; HURLIN, CHRISTOPHE and TOKPAVI, SESSI (2011): Backtesting value-at-risk: A GMM duration-based test, in: Journal of Financial Econometrics, Vol. 9, No. 2, pp. 314–343.

CHENG, XIXIN; YU, PHILIP L. H. and LI, WAI KEUNG (2009): On a Dynamic Mixture GARCH Model, in: Journal of Forecasting, Vol. 28, No. 3, pp. 247–265.

CHRISTIE, ANDREW A. (1982): The stochastic behavior of common stock variances. Value, leverage and interest rate effects, in: Journal of Financial Economics, Vol. 10, No. 4, pp. 407–432.

CHRISTOFFERSEN, PETER (2010): Backtesting, in: *Encyclopedia of Quantitative Finance*, Chichester, UK: John Wiley & Sons, Ltd.

CHRISTOFFERSEN, PETER and PELLETIER, DENIS (2004): Backtesting Value-at-Risk: A Duration-Based Approach, in: Journal of Financial Econometrics, Vol. 2, No. 1, pp. 84–108.

CHRISTOFFERSEN, PETER F. (1998): Evaluating Interval Forecasts, in: International Economic Review, Vol. 39, No. 4, pp. 841–862.

CONRAD, CHRISTIAN (2010): Non-negativity conditions for the hyperbolic GARCH model, in: Journal of Econometrics, Vol. 157, No. 2, pp. 441–457.

CONRAD, CHRISTIAN and HAAG, BERTHOLD R. (2006): Inequality Constraints in the Fractionally Integrated GARCH Model, in: Journal of Financial Econometrics, Vol. 4, No. 3, pp. 413–449.

CONT, RAMA (2001): Empirical properties of asset returns: stylized facts and statistical issues, in: Quantitative Finance, Vol. 1, No. 2, pp. 223–236.

COOLEY, JAMES W. and TUKEY, JOHN W. (1965): An Algorithm for the Machine Calculation of Complex Fourier Series, in: Mathematics of Computation, Vol. 19, No. 90, pp. 297–301.

CRNKOVIC, CEDOMIR and DRACHMAN, JORDAN (1996): Quality Control, in: Risk, Vol. 9, No. 9, pp. 138–143.

DAVIDSON, JAMES (2004): Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model, in: Journal of Business & Economic Statistics, Vol. 22, No. 1, pp. 16–29.

DAVIDSON, JAMES and LI, XIAOYU (2014): Strict stationarity, persistence and volatility forecasting in ARCH(oo) processes, in: Journal of Empirical Finance, Vol. 38, pp. 534–547.

DEMPSTER, ARTHUR P.; LAIRD, NAN M. and RUBIN, DONALD B. (1977): Maximum likelihood from incomplete data via the EM algorithm, in: Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 1, pp. 1–38.

DIEBOLD, FRANCIS X.; GUNTHER, TODD A. and TAY, ANTHONY S. (1998): Evaluating Density Forecasts with Applications to Financial Risk Management, in: International Economic Review, Vol. 39, No. 4, pp. 863–883.

DIEBOLD, FRANCIS X.; LEE, JOON-HAENG and WEINBACH, GRETCHEN C. (1994): Regime Switching with Time-Varying Transition Probabilities, in: HARGREAVES, C. (ed.), *Nonstationary Time Series Analysis and Cointegration*, Oxford: Oxford University Press, pp. 283–302.

DIEBOLD, FRANCIS X. and MARIANO, ROBERT S. (1995): Comparing Predictive Accuracy, in: Journal of Business & Economic Statistics, Vol. 13, No. 3, pp. 134–144.

DING, ZHUANXIN and GRANGER, CLIVE W. J. (1996): Modeling volatility persistence of speculative returns: A new approach, in: Journal of Econometrics, Vol. 73, No. 1, pp. 185–215.

DING, ZHUANXIN; GRANGER, CLIVE W. J. and ENGLE, ROBERT F. (1993): A long memory property of stock market returns and a new model, in: Journal of Empirical Finance, Vol. 1, No. 1, pp. 83–106.

DOUC, RANDAL; ROUEFF, FRANÇOIS and SOULIER, PHILIPPE (2008): On the existence of some ARCH(∞) processes, in: Stochastic Processes and their Applications, Vol. 118, No. 5, pp. 755–761.

DUAN, JIN-CHUAN (1995): The GARCH Option Pricing Model, in: Mathematical Finance, Vol. 5, No. 1, pp. 13–32.

DUAN, JIN-CHUAN (1997): Augmented GARCH (p,q) process and its diffusion limit, in: Journal of Econometrics, Vol. 79, No. 1, pp. 97–127.

EMMER, SUSANNE; KRATZ, MARIE and TASCHE, DIRK (2015): What is the best risk measure in practice? A comparison of standard measures, in: The Journal of Risk, Vol. 18, No. 2, pp. 31–60.

ENGLE, ROBERT (2002): Dynamic Conditional Correlation, in: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models, Vol. 20, No. 3, pp. 339–350.

ENGLE, ROBERT F. (1982): Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, in: Econometrica, Vol. 50, No. 4, pp. 987–1007.

ENGLE, ROBERT F. (1983): Estimates of the Variance of U. S. Inflation Based upon the ARCH Model, in: Journal of Money, Credit and Banking, Vol. 15, No. 3, pp. 286–301.

ENGLE, ROBERT F. (1990): Stock Volatility and the Crash of ’87: Discussion, in: The Review of Financial Studies, Vol. 3, No. 1, pp. 102–106.

ENGLE, ROBERT F. and BOLLERSLEV, TIM (1986a): Modelling the persistence of conditional variances, in: Econometric Reviews, Vol. 5, No. 1, pp. 1–50.

ENGLE, ROBERT F. and BOLLERSLEV, TIM (1986b): Reply, in: Econometric Reviews, Vol. 5, No. 1, pp. 81–87.

ENGLE, ROBERT F.; GHYSELS, ERIC and SOHN, BUMJEAN (2013): Stock Market Volatility and Macroeconomic Fundamentals, in: Review of Economics and Statistics, Vol. 95, No. 3, pp. 776–797.

ENGLE, ROBERT F. and LEE, GARY (1999): A long-run and short-run component model of stock return volatility, in: ENGLE, ROBERT and WHITE, HALBERT (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Oxford: Oxford University Press, pp. 475–497.

ENGLE, ROBERT F.; LILIEN, DAVID M. and ROBINS, RUSSELL P. (1987): Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model, in: Econometrica, Vol. 55, No. 2, pp. 391–407.

ENGLE, ROBERT F. and MANGANELLI, SIMONE (2004): CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, in: Journal of Business & Economic Statistics, Vol. 22, No. 4, pp. 367–381.

ENGLE, ROBERT F. and NG, VICTOR K. (1993): Measuring and Testing the Impact of News on Volatility, in: The Journal of Finance, Vol. 48, No. 5, pp. 1749–1778.

ENGLE, ROBERT F. and RANGEL, JOSE GONZALO (2008): The spline-GARCH model for low-frequency volatility and its global macroeconomic causes, in: Review of Financial Studies, Vol. 21, No. 3, pp. 1187–1222.

FAMA, EUGENE F. (1965): The behavior of stock-market prices, in: Journal of Business, Vol. 38, No. 1, pp. 34–105.

FRANCQ, CHRISTIAN and ZAKOIAN, JEAN-MICHEL (2010): GARCH Models: Structure, Statistical Inference and Financial Applications, Chichester: Wiley.

FRANKE, JÜRGEN; HÄRDLE, WOLFGANG KARL and HAFNER, CHRISTIAN MATTHIAS (2015): Statistics of Financial Markets, Heidelberg: Springer, 4th edition.

FRANSES, PHILIP HANS and VAN DIJK, DICK (1996): Forecasting stock market volatility using (non-linear) Garch models, in: Journal of Forecasting, Vol. 15, No. 3, pp. 229–235.

FRENCH, KENNETH R.; SCHWERT, G. WILLIAM and STAMBAUGH, ROBERT F. (1987): Expected stock returns and volatility, in: Journal of Financial Economics, Vol. 19, No. 1, pp. 3–29.

GALLANT, A. RONALD (1984): The Fourier Flexible Form, in: American Journal of Agricultural Economics, Vol. 66, No. 2, pp. 204–208.

GEWEKE, JOHN (1989): Exact predictive densities for linear models with arch disturbances, in: Journal of Econometrics, Vol. 40, No. 1, pp. 63–86.

GHYSELS, ERIC; SANTA-CLARA, PEDRO and VALKANOV, ROSSEN (2004): The MIDAS Touch: Mixed Data Sampling Regression Models, in: CIRANO Working Papers, Vol. 20, No. 919, pp. 1–33.

GIRAITIS, LIUDAS and ROBINSON, PETER M. (2001): Whittle Estimation of Arch Models, in: Econometric Theory, Vol. 17, No. 3, pp. 608–631.

GLOSTEN, LAWRENCE R.; JAGANNATHAN, RAVI and RUNKLE, DAVID E. (1993): On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, in: Journal of Finance, Vol. 48, No. 5, pp. 1779–1801.

GNEITING, TILMANN (2011): Making and Evaluating Point Forecasts, in: Journal of the American Statistical Association, Vol. 106, No. 494, pp. 746–762.

GONZÁLEZ-RIVERA, GLORIA (1998): Smooth-Transition GARCH Models, in: Studies in Nonlinear Dynamics & Econometrics, Vol. 3, No. 2.

GRANGER, CLIVE W. J. (1980): Long memory relationships and the aggregation of dynamic models, in: Journal of Econometrics, Vol. 14, No. 2, pp. 227–238.

GRANGER, CLIVE W. J. and JOYEUX, ROSELYNE (1980): An Introduction to Long-Memory Time Series Models and Fractional Differencing, in: Journal of Time Series Analysis, Vol. 1, No. 1, pp. 15–29.

GRAY, STEPHEN F. (1996): Modeling the conditional distribution of interest rates as a regime-switching process, in: Journal of Financial Economics, Vol. 42, pp. 27–62.

HAAS, M.; MITTNIK, STEFAN and PAOLELLA, MARC S. (2004a): Mixed Normal Conditional Heteroskedasticity, in: Journal of Financial Econometrics, Vol. 2, No. 2, pp. 211–250.

HAAS, MARKUS; MITTNIK, STEFAN and PAOLELLA, MARC S. (2004b): A New Approach to Markov-Switching GARCH Models, in: Journal of Financial Econometrics, Vol. 2, No. 4, pp. 493–530.

HAMILTON, JAMES D. (1989): A new approach to the economic analysis of nonstationary time series, in: Econometrica, Vol. 57, No. 2, pp. 357–384.

HAMILTON, JAMES D. (1990): Analysis of Time Series Subject to Changes in Regime, in: Journal of Econometrics, Vol. 45, pp. 39–70.

HAMILTON, JAMES D. (1994): Time Series Analysis, Princeton: Princeton University Press.

HAMILTON, JAMES D. and SUSMEL, RAUL (1994): Autoregressive conditional heteroskedasticity and changes in regime, in: Journal of Econometrics, Vol. 64, pp. 307–333.

HANSEN, PETER R.; LUNDE, ASGER and NASON, JAMES M. (2011): The Model Confidence Set, in: Econometrica, Vol. 79, No. 2, pp. 453–497.

HANSEN, PETER REINHARD (2005): A Test for Superior Predictive Ability, in: Journal of Business & Economic Statistics, Vol. 23, No. 4, pp. 365–380.

HANSEN, PETER REINHARD and LUNDE, ASGER (2005): A forecast comparison of volatility models: does anything beat a GARCH(1,1)?, in: Journal of Applied Econometrics, Vol. 20, No. 7, pp. 873–889.

HARVEY, CAMPBELL R and SIDDIQUE, AKHTAR (1999): Autoregressive Conditional Skewness, in: The Journal of Financial and Quantitative Analysis, Vol. 34, No. 4, pp. 465–487.

HE, CHANGLI; TERÄSVIRTA, TIMO and MALMSTEN, HANS (2002): Moment Structure of a Family of First-Order Exponential GARCH Models, in: Econometric Theory, Vol. 18, No. 04, pp. 868–885.

HENRY, ÓLAN T. (2009): Regime switching in the relationship between equity returns and short-term interest rates in the UK, in: Journal of Banking and Finance, Vol. 33, No. 2, pp. 405–414.

HERRERA, RODRIGO and SCHIPP, BERNHARD (2013): Value at risk forecasts by extreme value models in a conditional duration framework, in: Journal of Empirical Finance, Vol. 23, pp. 33–47.

HIGGINS, MATTHEW L. and BERA, ANIL K. (1992): A Class of Nonlinear Arch Models, in: International Economic Review, Vol. 33, No. 1, pp. 137–158.

HULL, JOHN C. and WHITE, ALAN D. (1998): Incorporating volatility updating into the historical simulation method for value-at-risk, in: The Journal of Risk, Vol. 1, No. 1, pp. 5–19.

HUSCHENS, STEFAN (2017): Risikomaße, in: Dresdener Beiträge zu Quantitativen Verfahren, Vol. 68.

INCLAN, CARLA and TIAO, GEORGE C. (1994): Use of cumulative sums of squares for retrospective detection of changes of variance, in: Journal of the American Statistical Association, Vol. 89, No. 427, pp. 913–923.

J. P. MORGAN (1996): RiskMetrics - Technical Document, Technical report. URL: www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

JENSEN, ANDREAS NOACK and NIELSEN, MORTEN ØRREGAARD (2014): A Fast Fractional Difference Algorithm, in: Journal of Time Series Analysis, Vol. 35, No. 5, pp. 428–436.

JORION, PHILIPPE (2007): Value at Risk: The New Benchmark for Managing Financial Risk, New York: McGraw-Hill, 3rd edition.

KAZAKEVIČIUS, VYTAUTAS and LEIPUS, REMIGIJUS (2003): A new theorem on the existence of invariant distributions with applications to ARCH processes, in: Journal of Applied Probability, Vol. 40, No. 1, pp. 147–162.

KIM, CHANG-JIN (1994): Dynamic linear models with Markov-switching, in: Journal of Econometrics, Vol. 60, No. 1-2, pp. 1–22.

KLAASSEN, FRANC (2002): Improving GARCH volatility forecasts with regime-switching GARCH, in: Empirical Economics, Vol. 27, No. 2, pp. 363–394.

KLEIN, TONY (2017): Conditional Variance Dynamics of Gold and Silver: On Correlation and Forecast Comparison with High Frequency Data. Unpublished Manuscript.

KLEIN, TONY; PHAM THU, HIEN and WALTHER, THOMAS (2016): Evidence of long memory and asymmetry in the EUR/PLN exchange rate volatility, in: Research Papers of Wroclaw University of Economics, Vol. 428, No. 428, pp. 128–140.

KLEIN, TONY and WALTHER, THOMAS (2016): Oil price volatility forecast with mixture memory GARCH, in: Energy Economics, Vol. 58, pp. 46–58.

KLEIN, TONY and WALTHER, THOMAS (2017): Fast fractional differencing in modeling long memory of conditional variance for high-frequency data, in: Finance Research Letters, forthcoming.

KRÄMER, WALTER (2008): Long memory with Markov-Switching GARCH, in: Economics Letters, Vol. 99, No. 2, pp. 390–392.

KRISTENSEN, DENNIS and LINTON, OLIVER (2006): A Closed-Form Estimator for the GARCH(1,1) Model, in: Econometric Theory, Vol. 22, No. 02, pp. 323–337.

KUAN, CHUNG MING; YEH, JIN HUEI and HSU, YU CHIN (2009): Assessing value at risk with CARE, the Conditional Autoregressive Expectile models, in: Journal of Econometrics, Vol. 150, No. 2, pp. 261–270.

KUPIEC, PAUL H. (1995): Techniques for Verifying the Accuracy of Risk Measurement Models, in: The Journal of Derivatives, Vol. 3, No. 2, pp. 73–84.

LAUENSTEIN, PHILIPP and WALTHER, THOMAS (2016): Forecasting volatility of tanker freight rates based on asymmetric regime-switching GARCH models, in: International Journal of Financial Engineering and Risk Management, Vol. 2, No. 3, pp. 172–199.

LAURENT, SÉBASTIEN and PETERS, JEAN-PHILIPPE (2002): G@rch 2.2: An OX Package for Estimating and Forecasting Various ARCH Models, in: Journal of Economic Surveys, Vol. 16, No. 3, pp. 447–485.

LEE, TAE HWY and LONG, XIANGDONG (2009): Copula-based multivariate GARCH model with uncorrelated dependent errors, in: Journal of Econometrics, Vol. 150, No. 2, pp. 207–218.

LI, MUYI; LI, WAI KEUNG and LI, GUODONG (2013): On Mixture Memory GARCH Models, in: Journal of Time Series Analysis, Vol. 34, No. 6, pp. 606–624.

LIN, BING-HUEI and YEH, SHIH-KUO (2000): On the distribution and conditional heteroscedasticity in Taiwan stock prices, in: Journal of Multinational Financial Management, Vol. 10, No. 3-4, pp. 367–395.

LIU, JI CHUN (2006): Stationarity of a Markov-Switching GARCH model, in: Journal of Financial Econometrics, Vol. 4, No. 4, pp. 573–593.

LOCAREK-JUNGE, HERMANN; KLEIN, TONY and WALTHER, THOMAS (2014): GARCH-Modelle, in: WISU - Das Wirtschaftsstudium, Vol. 43, No. 11, pp. 1348–1354.

LOCAREK-JUNGE, HERMANN and PRINZLER, RALF (1998): Estimating Value-at-Risk Using Neural Networks, in: WEINHARDT, CHRISTOF; MEYER ZU SELHAUSEN, HERMANN and MORLOCK, MARTIN (eds.), *Informationssysteme in der Finanzwirtschaft*, Berlin: Springer, pp. 385–397.

LOCAREK-JUNGE, HERMANN and WALTHER, THOMAS (2017): Markov-Regime-Switching-Modelle in der Finanzwirtschaft, in: WiSt - Wirtschaftswissenschaftliches Studium, Vol. 46, No. 1, pp. 4–9.

LOPEZ, JOSE A. (1998): Methods for evaluating value-at-risk estimates, in: Economic Policy Review, , No. August 1996, pp. 119–124.

LÜTKEPOHL, HELMUT (2006): New Introduction to Multiple Time Series Analysis, Berlin: Springer.

MANDELBROT, BENOIT (1963): The variation of certain speculative prices, in: The Journal of Business, Vol. 36, No. 4, pp. 394–419.

MANSUR, IQBAL; COCHRAN, STEVEN J. and SHAFFER, DAVID (2007): Foreign Exchange Volatility Shifts and Futures Hedging: An ICSS-GARCH Approach, in: Review of Pacific Basin Financial Markets and Policies, Vol. 10, No. 03, pp. 349–388.

MCMILLAN, DAVID G. and KAMBOUROUDIS, DIMOS (2009): Are RiskMetrics forecasts good enough? Evidence from 31 stock markets, in: International Review of Financial Analysis, Vol. 18, No. 3, pp. 117–124.

MCNEIL, ALEXANDER J. and FREY, RÜDIGER (2000): Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, in: Journal of Empirical Finance, Vol. 7, No. 7, pp. 271–300.

MCNEIL, ALEXANDER J.; FREY, RÜDIGER and EMBRECHTS, PAUL (2015): Quantitative Risk Management: Concepts, Techniques and Tools, Princeton: Princeton University Press, revised edition.

NELSON, DANIEL B. (1990): Stationarity and Persistence in the GARCH(1,1) Model, in: Econometric Theory, Vol. 6, No. 03, pp. 318–334.

NELSON, DANIEL B. (1991): Conditional heteroskedasticity in asset returns: A new approach, in: Econometrica, Vol. 59, No. 2, pp. 347–370.

NOMIKOS, NIKOS K. and POULIASIS, PANOS K. (2011): Forecasting petroleum futures markets volatility: The role of regimes and market conditions, in: Energy Economics, Vol. 33, No. 2, pp. 321–337.

PALM, FRANZ C. and VLAAR, PETER J. G. (1997): Simple Diagnostic Procedures for Modeling Financial Time Series, in: Allg. Statistisches Archiv, Vol. 81, No. 1, pp. 85–101.

PARK, SUJIN and LINTON, OLIVER (2012): Realized Volatility: Theory and Applications, in: BAUWENS, LUC; HAFNER, CHRISTIAN and LAURENT, SÉBASTIEN (eds.), *Handbook of Volatility Models and Their Applications*, chapter 13, Hoboken, New Jersey: Wiley, pp. 317–345.

PASCALAU, RAZVAN; THOMANN, CHRISTIAN and GREGORIOU, GREG N. (2010): Unconditional mean, Volatility and the Fourier-GARCH representation. URL: https://mpra.ub.uni-muenchen.de/35932/.

PATTON, ANDREW J. (2011): Volatility forecast comparison using imperfect volatility proxies, in: Journal of Econometrics, Vol. 160, No. 1, pp. 246–256.

PENG, LIANG (2003): Least absolute deviations estimation for ARCH and GARCH models, in: Biometrika, Vol. 90, No. 4, pp. 967–975.

PEREZ-QUIROS, GABRIEL and TIMMERMANN, ALLAN (2001): Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities, in: Journal of Econometrics, Vol. 103, No. 1-2, pp. 259–306.

PÉRIGNON, CHRISTOPHE and SMITH, DANIEL R. (2008): A New Approach to Comparing VaR Estimation Methods, in: The Journal of Derivatives, Vol. 16, No. 2, pp. 54–66.

PIONTEK, KRZYSZTOF (2010): The analysis of power for some chosen VaR backtesting procedures: Simulation approach, in: FINK, ANDREAS; LAUSEN, BERTHOLD; SEIDEL, WILFRIED and ULTSCH, ALFRED (eds.), *Advances in Data Analysis, Data Handling and Business Intelligence*, Heidelberg: Springer-Verlag, pp. 481–490.

POON, SER-HUANG (2005): A Practical Guide to Forecasting Financial Market Volatility, John Wiley & Sons.

ROCKAFELLAR, R. TYRRELL and URYASEV, STANISLAV (2002): Conditional value-at-risk for general loss distributions, in: Journal of Banking and Finance, Vol. 26, No. 7, pp. 1443–1471.

RUPPERT, DAVID and MATTESON, DAVID S. (2015): Statistics and Data Analysis for Financial Engineering, New York: Springer, 2nd edition.

SANSÓ, A; ARAGÓ, V and CARRION, JL (2004): Testing for changes in the unconditional variance of financial time series, in: Revista de Economía financiera, Vol. 4, pp. 32–53.

SARMA, MANDIRA; THOMAS, SUSAN and SHAH, AJAY (2003): Selection of value-at-risk models, in: Journal of Forecasting, Vol. 22, No. 4, pp. 337–358.

SCHWARZ, GIDEON (1978): Estimating the Dimension of a Model, in: The Annals of Statistics, Vol. 6, No. 2, pp. 461–464.

SCHWERT, G. WILLIAM (1989): Why Does Stock Market Volatility Change Over Time?, in: The Journal of Finance, Vol. 44, No. 5, pp. 1115–1153.

SENTANA, ENRIQUE (1995): Quadratic ARCH Models, in: The Review of Economic Studies, Vol. 62, No. 4, pp. 639–661.

SMITH, GEOFFREY PETER (2016): Weekday variation in the leverage effect: A puzzle, in: Finance Research Letters, Vol. 17, pp. 193–196.

TAYLOR, JAMES W. (2007): Estimating Value at Risk and Expected Shortfall Using Expectiles, in: Journal of Financial Econometrics, Vol. 6, No. 2, pp. 231–252.

TAYLOR, STEPHEN J. (1995): Modelling Financial Time Series, Chichester: Wiley.

TSAY, RUEY S. (2013): An Introduction to Analysis of Financial Data with R, Hoboken, New Jersey: Wiley.

TSE, YIU KUEN (1998): The conditional heteroscedasticity of the yen-dollar exchange rate, in: Journal of Applied Econometrics, Vol. 13, No. 1, pp. 49–55.

VLAAR, PETER J G and PALM, FRANZ C (1993): The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps, in: Journal of Business & Economic Statistics, Vol. 11, No. 3, pp. 351–360.

WALTHER, THOMAS (2017): Expected Shortfall in the Presence of Asymmetry and Long Memory: An Application to Vietnamese Stock Markets, in: Pacific Accounting Review, Vol. 29, No. 2, pp. 132–151.

WALTHER, THOMAS; KLEIN, TONY; PHAM THU, HIEN and PIONTEK, KRZYSZTOF (2017): True or spurious long memory in European Non-EMU currencies, in: Research in International Business and Finance, Vol. 40C, pp. 217–230.

WHITE, HALBERT (1982): Maximum Likelihood Estimation of Misspecified Models, in: Econometrica, Vol. 50, No. 1, pp. 1–25.

WHITE, HALBERT (2000): A Reality Check for Data Snooping, in: Econometrica, Vol. 68, No. 5, pp. 1097–1126.

WONG, WOON K. (2008): Backtesting trading risk of commercial banks using expected shortfall, in: Journal of Banking & Finance, Vol. 32, No. 7, pp. 1404–1415.

ZAKOIAN, JEAN MICHEL (1994): Threshold heteroskedastic models, in: Journal of Economic Dynamics and Control, Vol. 18, No. 5, pp. 931–955.

ZIEGEL, JOHANNA F. (2014): Coherence and elicitability, in: Mathematical Finance, Vol. 26, No. 4, pp. 901–918.

ZIGGEL, DANIEL; BERENS, TOBIAS; WEISS, GREGOR N.F. and WIED, DOMINIK (2014): A new set of improved Value-at-Risk backtests, in: Journal of Banking & Finance, Vol. 48, pp. 29–41.