Further Development of Semiparametric Volatility Models and their Applications to Value at Risk and Expected Shortfall

Der Fakultät für Wirtschaftswissenschaften der

Universität Paderborn

zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften

-Doctor rerum politicarum-

vorgelegte Dissertation

von

Xuehai Zhang, M.A.

August, 2019

Abstract

Following the implementation of Basel III and its forthcoming finalization, risk management has been already paid high attention by firms, financial institutions and even the Basel Committee on Banking Supervision. The risk measurement and financial risk models become the crucial prerequisites in risk management.

Firstly, some recently developed financial volatility models are provided in the dissertation. In order to examine the long-term and short-term market risk, a time-varying scale function is introduced based on parametric models, such as the GARCH, the ACD and the MEM models. Then, the market risk is decomposed into a short-term risk by a parametric part and a long-term risk by a semiparametric trend function. Due to no parametric model assumptions, the set-up semiparametric models are model free and general in the parametric part. Besides, to reduce the moments requirement of the considered data, the power transformation is employed in the general semiparametric models, i.e. some general power transformed semiparametric models are built up, such as the general Box-Cox SemiGARCH model, the general Box-Cox SemiACD model and the general Box-Cox SemiMEM model. The estimation of the scale function is not related to any parametric specification. Due to the shortcomings of the kernel and the local linear approaches, a non-negative constrained local linear estimator of the trend, which is next proposed to descale a suitable parametric model to the standardized residuals, is under consideration. Iterative plug-in algorithms are developed to estimate the bandwidth and the power parameter. For the power parameter estimation, various criteria, such as the Jarque-Bera test, the maximum likelihood estimation, the Shapiro-Wilk test and quantile-quantile regression, are employed. Further, the block bootstrap simulation is carried out to estimate the confidence interval of the power parameter. The empirical findings are illustrated by applying the algorithms to real financial market data, e.g. the returns, the trading duration,

the trading volume and the trading numbers, indicating the good performance of the general power transformed semiparametric models.

For another, the value at risk and expected shortfall are also predicted by the general semiparametric models. In the stationary process, we use the conditional t-distribution as the assumption in both risk measures. Well known models, such as the GARCH class models, including GARCH, APARCH, EGARCH, etc., based on the conditional t-distribution, are as parametric extensions. In addition, the backtesting with the semiparametric approach for both value at risk and expected shortfall are also discussed. Although the Kupiec POF test and independence test are carried out, the robustness of the results is challenged. Following Basel III, the traffic light tests, considering the cumulative probability, are applied. For expected shortfall, a breach indicator is introduced to obtain a similar traffic light test of value at risk. Loss functions from different viewers, such as the regulator and the firms, are also discussed. It is shown that different market participants prefer using different loss functions to maximize own profits. From the empirical cases, the semiparametric models are necessary tools in risk management.

Keywords: semiparametric, volatility, value at risk, expected shortfall, risk management

Acknowledgements

First of all, I would like to deeply appreciate my supervisor Prof. Dr. Yuanhua Feng for the provided kind guidance on my research and in this dissertation. The discussion with him helped produce the achieved results in the past few years. Also, his knowledge and appreciation of the current challenges in risk management extremely helped shape the direction of my research.

I am also indebted to my second supervisor Prof. Dr. Bernard Michael Gilroy and my defense committee members Prof. Dr. André Uhde and Prof. Dr. Stefan Jungblut. Their comments are of great benefit to the dissertation, contributed to the modification and improved the final thesis.

Besides, special thanks should also go to all of my colleagues, Dr. Sarah Forstinger, Dr. Christian Peitz and Sebastian Letmathe. Thanks for the encouragement during the thesis writing.

Also, I would like to express my gratitude to the China Scholar Council for the partial financial support of my Ph.D. research.

Finally, my thanks would send to my beloved Dad and all the family for their forever love and support throughout my abroad life. Especially, Mum, you are always with me, in my heart, thank you!

Contents

List of Tables

List of Figures

List of Abbreviations

-AP	Loss function with APARCH model
-EG	Loss function with EGARCH model
-OG	Loss function with GARCH model
a.s.	Almost surely
ABC	Approximate bootstrap confidence
ACD	Autoregressive conditional duration
ACV	Autoregressive conditional volume
AG	Aktiengesellschaft
AIC	Akaike information criterion
ALV	Allianz SE
APARCH	Asymmetric power ARCH
AR	Absolute returns
ARB	Autoregressive bootstrap
ARCH	Autoregressive conditional heteroskedasticity
ARMA	Autoregressive moving average
BACD	Burr ACD
BC	Bias-corrected
BCa	Bias-corrected and accelerated
BCBS	Basel Committee on Banking Supervision
BIC	Bayesian information criterion
BMW	Bayerische Motoren Werke AG
Breach	Breach indicator

BRO	Brent Crude Oil Futures
BSN	SP BSE SENSEX index
BTS	Business time sampling
CGARCH	Component GARCH
CI	Confidence interval
coef.	Coefficient
CS-	Parametric model
CTS	Calendar time sampling
CV	Cross-validation
CVaR	Conditional VaR
DAX	Deutscher Aktienindex
DBK	Deutsche Bank AG
DPI	Direct plug-in
e.g.	Exempli gratia (for example)
EACD	Exponential ACD
EARCH	Exponential GARCH
EIM	Exponential inflation method
EPOT	POT of expected shortfall
Eq.	Equation
ES	Expected shortfall
EST	Euro STOXX 50 index
FIACD	Fractionally integrated ACD
Fig.	Figure
FIGARCH	Fractionally integrated GARCH

FLF	Firm loss function
FTSE	FTSE 100 index
GACD	Gamma ACD
GARCH	Generalized ARCH
GCV	Generalized cross-validation
GDP	Gross domestic product
GJR-GARCH	Glosten-Jagannathan-Runkle GARCH
IF	Influence function
Ind	Christoffersen independence test
IPI	Iterative plug-in
JB	Jarque-Bera test
LC-	Log-transformed model
LL1-	Power transformation with loss function by Lopez
LL2-	Power transformation with loss function by Sarma
LL3-	Power transformation with loss function by Feng
LMGARCH	Long-memory GARCH
Log-ACD	Logarithmic ACD
log-vMEM	Logarithmic vector multiplicative error model
Loss1	Regulator loss function by Lopez
Loss2	Firm loss function by Sarma
Loss3	Firm loss function by Feng
MBB	Moving block bootstrap
MD	Mean duration

MEM	Multiplicative error model
MEM-J	Multiplicative error model with volatility jumps
MIDAS	Mixed data sampling
MIM	Multiplicative inflation method
MISE	Mean integrated squared error
Mix	Mixed Kupiec test
MLE	Maximum likelihood estimation
MSE	Mean squared error
Nemax	Neuer Markt Index
NGARCH	Nonlinear GARCH
Norm	Normal distribution
POF	Proportion of failures
POT	Peaks over threshold
QMLE	Quasi maximum likelihood estimate
QQr	Quantile-quantile regression
RLF	Regulator loss function
RUT	Russell 2000 index
RV	Realized volatility
s.e.	Standard error
SE	Societas Europaea
Semi-FI-Log-ACD	Semiparametric fractionally logarithmic ACD
Semi-FI-Log-MEM	Semiparametric fractionally logarithmic MEM
Semi-FI-MEM	Semiparametric fractionally MEM

Semi-Log-ACD	Semiparametric logarithmic ACD
Semi-Log-MEM	Semiparametric logarithmic MEM
SemiACD	Semiparametric ACD
SemiAPARCH	Semiparametric APARCH
SemiCGARCH	Semiparametric CGARCH
SemiEGARCH	Semiparametric EGARCH
SemiGARCH	Semiparametric GARCH
SemiMEM	Semiparametric MEM
SIE	Siemens AG
SP	Standard and Poor's 500 Index
Std	Student t-distribution
SW	Shapiro-Wilk test
TGARCH	Threshold GARCH
TickTS	Tick-time sampling
TR	Trading number
TS-GARCH	Taylor-Schwert GARCH
TTS	Transaction time sampling
TUFF	Time until first failure
UHF	Ultra high-frequency (data)
VaR	Value at risk
VO	Trading volume
VPOT	POT of value at risk
VSACD	Varying scale ACD
VSGARCH	Varying scale GARCH
VSMEM	Varying scale MEM

WACD	Weibull ACD

Chapter 1

Introduction

Since the global financial crisis, it has been realized that more close attention should be paid to the significance of quantitative risk management of the financial market operation. The financial econometrics and quantitative risk management are definitely related to not only the market economy operation but also the micro- and macroeconomics performance of the countries all over the world in every field. Financial risk engineer N. Taleb, the author of the book *Black Swan*, strongly warned about the banks' approaches applied to risk management and their irresponsibility against potential risk in the modern financial system from the coming crisis and its negative consequences.

In the financial world, volatility is an important concept in financial econometrics and is widely used in investment portfolios, asset pricing, product pricing and risk management. Currently, the empirical research and analysis on the volatility of financial prices have become one of the essential problems in modern financial risk management research. The financial modeling in the research process is the crucial technique to analyze the volatility and risk decomposition in markets.

Price volatility in financial markets is often considered and measured by classical variance models. In these models, the assumption has always recognized the variance as a constant at different time points. However, with the development of financial science research, it is found that this assumption can not reveal the real volatility movement in the financial market and a large number of financial time series such as stock prices, GDP indexes, interest rates and currency exchange rates show that the variance is not fixed, but time-varying. The ARCH (autoregressive conditional heteroskedasticity, Engle, 1982) and GARCH (general-

ized ARCH, Bollerslev, 1986) are tremendous successes for modeling volatility on financial markets. In literature, there is a huge number of extensions of the original GARCH model. For instance, the asymmetric power ARCH (APARCH, also called APGARCH, Ding et al., 1993) model, the EGARCH (exponential GARCH, Nelson, 1991) to discuss the short memory property in the volatility. There are also volatility models to reveal the long-memory property of squared returns, such as FIGARCH (fractionally integrated GARCH, Baillie et al., 1996), the LMGARCH (long-memory GARCH, Karanasos et al., 2004, Conrad, 2006 and Conrad and Karanasos, 2006) and so on. Those extensions are all defined as stationary time series. In practice, it is found that the unconditional variance of asset returns in a long period usually change over time. The SemiGARCH (semiparametric GARCH, Feng, 2004) and the Spline-GARCH (Engle and Rangel, 2008) model are hence proposed to analyze this nonstationary property of volatility, where a nonparametric scale function is introduced to parametric volatility models.

The high-frequency financial data is being focused in financial modeling. The research on ultra high-frequency data, such as the duration, has attracted more and more scholars. To analyze the ultra high-frequency data, Engle and Russell (1998) proposed an autoregressive conditional duration (ACD) model, which is an important tool in the high-frequency financial data analysis. Afterwards a lot of extensions of the ACD model were proposed. To analysis the long-term dependencies in the duration series, Jasiak (1998) extended the ACD model to the fractionally integrated ACD (FIACD) model. A logarithmic version of the ACD (Log-ACD) model was introduced by Bauwens and Giot (2000). The ACD models were extended to semiparametric methods. Feng (2013) proposed a SemiACD model and applied a local linear method to estimate the diurnal pattern. Feng and Zhou (2015) discussed a Semi-Log-ACD process by introducing the scale function into the logarithm ACD models. Indeed, it is a special case of the Semi-FI-Log-ACD model with fractional differencing parameter $d = 0$.

In the dissertation, the definition, estimation and properties of the semiparametric models and the methods of bandwidth selection are discussed with various data types. In the financial return series, a SemiGARCH model is considered to

model the volatility in the long run by introducing a time-varying scale function and after removing the trend, the stationary return series can be analyzed with any classical GARCH type model. The semiparametric analysis can also be carried out in the research on the non-negative financial variables, such as trading volume, trading number, average transaction duration and volatility indexes and the SemiACD or Semi-Log-ACD model is applied as an extension of parametric duration model to describe the mean movement in decades. Furthermore, the semiparametric algorithm is available to not only the financial data but also the macroeconomic data. In practice, the semiparametric modeling with macroeconomic data, such as GDP, inflation rate, interest rate and so on, works also very well.

The power transformation is a key idea proposed in the dissertation. The Box-Cox transformation (Box and Cox, 1964) is developed as an efficient power transform technique, applying the non-linear transformation, especially in the analysis of the variance. To apply the Box-Cox transformation, the expectation of the variable possesses a simple structure and the error of the variable is also consistent. Furthermore, the most important advantage of the Box-Cox is, after Box-Cox transformation, the distribution of the variable ís closer to the normal distribution. Manly (1976) proposed an exponential power transformation and in this exponential transformation, the negative values of the variable are also considered. Modulus Transformation (John and Draper, 1980), a modified power transformation, introduced a sign function in the data transform and it is effective with the symmetric distribution data set, however, the transformation seems to be invalid when the transformation parameter is zero. Yeo and Johnson (2000) developed another modification by minimizing the Kullback-Leibler distance between the normal distribution and the transformed distribution and considered not only the transformation of the negative values but also the logarithm form with power zero.

In the analysis of Box-Cox transformation, the focus is the determination of the power transformation parameter. Box and Cox discussed two approaches, the Maximum Likelihood method (MLE) and the Bayesian method. MLE is the

most common method applied in the searching of the power transformation parameter due to the feasibility in computation and confidence interval calculation. Besides, the normality (goodness of fit) test can also be applied in the power transformation parameter selection. Rahman (1999) and Rahman and Pearson (2008) applied the Shapiro-Wilk test and the Anderson-Darling test to obtain the transformation, respectively. Asar et al. (2017) summarized the common normality test, such as Shapiro-Wilk test, Anderson-Darling test, Cramer-von Mises test, Pearson chi-square test, Kolmogorov-Smirnov test, Jarque-Bera test and a method of artificial covariate and developed these normality tests in the searching algorithms to find maximum or minimum statistic values rather than numerical calculation. Furthermore, graphical methods is also an alternative method in the power transformation searching, by comparing the histogram of the sample data and a normal distribution curve, for example, the quantile-quantile regression (QQr) method.

The dissertation is organized as follows. In Chapter 2, the definition of parametric GARCH and ACD model is first introduced. Then, a variety of the common volatility models and duration models are detailed discussed. In the volatility models, ARCH, GARCH, APARCH, EGARCH and CGARCH models are selected as the representative and in the duration models, we discussed ACD and log ACD models. In the subsection of each model, we described the definitions, statistic properties and the estimation method. Besides, as extensions of the parametric models, we investigated the models with time-varying components. The SemiGARCH, SemiACD and Semi-Log-ACD models are the parametric modifications by introducing scale function.

In Chapter 3, a framework for general SemiGARCH models is built by introducing time-varying trends to present short- and long-term market performance through daily trading data. The scale function reveals long-term risk components and the classical parameter GARCH model exhibits short-term market risk. If the scale function is not considered, the restriction on the parameter GARCH model no longer exists and generally, the SemiGARCH framework does not require the assumption of the parameter part, which means that the parameter model is free.

Also, a power transformation is proposed to reduce the moment requirement of the GARCH model. The IPI algorithm is executed to estimate the power parameters to reach a convergence value.

In Chapter 4, as Basel III and its imminent completion, we will use the parametric and semiparametric models to examine VaR and ES predictions and backtesting. An important innovation is the ES traffic light test. In the empirical study, the ES backtesting test works well and does provide a simple and straightforward method for ES backtesting. In addition, some practical cases have been found to support semiparametric models, which can reasonably reveal market risks and meet regulatory agencies and enterprises, proving that semiparametric models are necessary risk management tools and complement the parametric models.

Then, in Chapter 5, the duration models are considered. The aim of the chapter is to describe the semiparametric models to analyze the non-negative data, such as mean transaction duration, trading number, trading volume, realized volatility and volatility indexes. A SemiACD model is proposed to discuss the scale function in the mean duration. The mean duration considered in this chapter still follows a multiplicative process, modeling with a time-varying MEM model. In the scale function estimation, the Box-Cox-transformation is still considered, however, in the estimation algorithm, not only the selection of bandwidth is considered but also the constant factor in the asymptotic variance.

A wide class of semiparametric GARCH models is interpreted in Chapter 6 by introducing a scale function into a GARCH-type model for featuring long-run volatility dynamics caused by changing the macroeconomic environment. The dynamics volatility can be modeled as an MEM with a varying scale function. Furthermore, the scale function is estimated under weak moment condition by means of the Box-Cox-transformation of the constrained returns and the scale function is estimated independent of any GARCH specification. Asymptotic properties of the proposed nonparametric and parametric estimators are studied in detail and an iterative plug-in algorithm is developed for selecting the bandwidth and the parametric estimation of the stationary part is also independent.

A further topic on the non-negative intraday high-frequency data with the

SemiMEM model is discussed in Chapter 7. At the end of the dissertation, the summary of the main contributions is concluded. In view of the current work, the shortcoming of the research is pointed out and design a reasonable research prospect.

Chapter 2

Parametric and Semiparametric Models

The GARCH model and its extensions are the most popular approaches to model the conditional heteroskedasticity in financial returns. However, financial returns possess not only the conditional heteroskedasticity but also time heteroskedasticity, implying the nonstationarity of financial returns in a long period. The feature requires a new GARCH-type model, modeling both the conditional volatility dynamics and the long-run risk. In this chapter, the selected parametric volatility and duration models are discussed. The properties of these models are also summarized. Besides, semiparametric models with a time-varying trend function are introduced, decomposing the long-run volatility.

2.1 Introduction

Financial time series models develop rapidly in recent decades, especially the GARCH-type models. From the ARMA model to the recent GARCH-type model, the model has experienced a process from linear to nonlinear, from parametric to nonparametric or semiparametric approach. It is known that volatility clustering exists in financial time series, and the distribution of random variables appears the fat tails. Different from the classical models, the Autoregressive Conditional Heteroskedasticity (ARCH) model suggests that the conditional variance could change over time as a function of past errors. In practical applications of the ARCH model, a relatively long lag in the conditional variance equation is often required, which might lead to an increase in the complexity of estimating parameters and decrease the freedom degree. However, the restrict condition is exactly needed in this model to ensure conditional variance to be non-negative. Therefore, there are many economists tried to improve ARCH models. Among these researches, the

generalized ARCH (GARCH) model, which is introduced by Bollerslev (1986), is the most widely well-known one with a better framework to study time-varying volatility in financial markets.

2.2 Overview of the volatility models

It is well known, financial markets are often volatile. The volatility is an important variable to indicate the risk of assets and reflect the uncertainty of asset returns. To measure the volatility, parametric models, such as (G)ARCH models and their extensions, are regarded as the most commonly used approaches in investment analysis and futures pricing. In this section, some of the parametric GARCH models will be introduced.

2.2.1 The ARCH model

In the study of ARCH models, the conditional mean and conditional variance of financial returns are introduced. Suppose $r_1, r_2, \ldots, r_t$ are the time series random variables and their conditional variances depend on a great number of information according to the past periods. The conditional mean does not depend on the past information,

$$E(r_t|r_{t-1}, \ldots, r_1) \doteq 0, \tag{2.1}$$

whereas the conditional variance

$$h_t = \text{var}\,(r_t|r_{t-1}, \ldots, r_1), \tag{2.2}$$

obviously depends on $r_{t-1}, \ldots, r_1$.

According to Engle (1982), the uncorrelated but dependent process X_t, can be defined as ARCH(p),

$$r_t = \eta_t h_t^{1/2}, \quad r_t|\mathcal{F}_{t-1} \sim N(0, h_t), \tag{2.3}$$

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2, \tag{2.4}$$

where η_t is a sequence of i.i.d. random variables with mean 0 and variance 1, $\sqrt{h_t}$ is the conditional standard deviation, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, p$ and $\mathcal{F}_{t-1}$ denotes the past information. Here, h_t is the conditional variance, which depends on p periods information in the past. The conditions $\alpha_0 > 0$ and $\alpha_i \geq 0$ guarantee a positive conditional variance. Obviously, conditional variance depends on the squared past observations. ARCH model indicates that r_t and r_{t+k} are not independent. Equation (2.4) describes the independence of r_t using a simple quadratic function of its lagged value.

An ARCH(p) model can also be written as AR (p) model. For an ARCH (p) with var $(X_t) < \infty$, define a martingale sequence

$$\varepsilon_t = r_t^2 - h_t, \tag{2.5}$$

ε_t are uncorrelated i.i.d. random variables with $E(\varepsilon_t) = 0$ and var $(\varepsilon_t) = 1$. We have

$$r_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \varepsilon_t. \tag{2.6}$$

The Equation (2.6) is an AR(p) model for r_t^2 with innovations ε_t. Obviously, it indicates that the ARCH model is related to the AR model.

2.2.2 The GARCH model

Bollerslev (1986) proposed the Generalized Autoregressive conditional Heteroskedasticity (GARCH) model based on the ARCH model. The GARCH model is regressive, as a result of the tool entirely explains the stylized facts observed in financial markets returns. Different from the ARCH model, the GARCH model can model the variance of the errors in addition. That is to say, the conditional variances in the GARCH process depend not only on the squared past observations but also on conditional variances in the past. Therefore, it is very suitable to analyze and forecast volatility.

A GARCH(p,q) model is defined by

$$r_t = \eta_t h_t^{1/2}, \quad r_t|\mathcal{F}_{t-1} \sim N(0, h_t), \tag{2.7}$$

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{2.8}$$

where η_t and h_t are as defined before, $p > 0, q > 0, \alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1, \ldots, p$, $\beta_j \geq 0$ for $j = 1, \ldots, q$. The higher p and q are, the smaller α_i are. The same as the ARCH(p) model, $\mathcal{F}_t$ denotes the information set of all information through time t. The conditions $\alpha_0 > 0, \alpha_i \geq 0$ and $\beta_j \geq 0$ guarantee the positivity of the conditional variance.

From the above formula of the GARCH model, we can see the difference between ARCH models and GARCH models. That is, following the GARCH models, the conditional variance depends on not only squared past observation in the previous p periods but also conditional variances in the previous in the past q periods. If $q = 0$, the GARCH(p, q) process decreases to the ARCH(p) process.

Let $\mathcal{M}(B)$ and $\mathcal{N}(B)$ be $\sum_{i=1}^{p} \alpha_i B^p$ and $1 - \sum_{j=1}^{q} \beta_j B^q$, respectively. As an infinite lag polynomial can be expressed as the quotient of two finite lag polynomial

$$\Phi(B) = \sum_{i=1}^{\infty} \phi_i B^i = \frac{\mathcal{M}(B)}{\mathcal{N}(B)}, \tag{2.9}$$

the GARCH(p, q) model can also be written as an ARCH(∞),

$$\begin{aligned} h_t &= \left(1 - \sum_{j=1}^{q} \beta_j B^j\right)^{-1} \left(\omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2\right) \\ &= \omega^* + \sum_{i=1}^{\infty} \phi_i r_{t-i}^2, \end{aligned} \tag{2.10}$$

where

$$\phi_i = \alpha_i + \sum_{j=1}^{J} \beta_j \phi_{i-j}, \quad i = 1, \ldots, p$$

and $J = \min\{q, i-1\}$, B is the backshift operator, $\omega^* = \omega/\mathcal{N}(1)$. If $\mathcal{N}(1) > 0$, ϕ_i will decrease for i is greater that $m = \max\{p, q\}$ (Bollerslev, 1986). Obviously, if $\omega^* \geq 0$ and $\phi_i \geq 0$, then $h_t \geq 0$.

We have the conditional mean $E(r_t|\mathcal{F}_{t-1}) = 0$ and hence $E(r_t) = 0$. Furthermore, it is almost the same as a ARCH(p) model, we have $\text{cov}\,(r_t|\mathcal{F}_{t-1}, r_{t+k}|\mathcal{F}_{t+k-1}) = 0$, for $k > 0$, hence $\gamma(k) = \text{cov}\,(r_t, r_{t+k}) = 0$. That is, r_t by a GARCH process are an uncorrelated series but not independent. However, the observations are not independent, because of the dependence between their variances or squared values, $\text{cov}\,(r_t^2, r_{t+k}^2) \neq 0$.

The weak stationary of a GARCH(p, q) process requires a necessary and sufficient condition. That is, if and only if the sum of all the coefficients is smaller than 1, i.e. the unconditional variance exists ($\text{var}\,(X_t) < \infty$). Under this condition, the unconditional variance of a GARCH(p, q) model can be also calculated as the constant parameter divided by the difference between the sum of all the coefficient and 1. It is clear to see that whether $\text{var}\,(r_t) < \infty$ does not depend on α_0 but only $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j$. However, if a unit GARCH is considered, α_0 is not a free parameter any longer and its value is exactly equal to $1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j$.

Consider the original GARCH model with conditional normal distribution, the method of conditional maximum likelihood estimation (MLE, Bollerslev, 1986) is always applied to estimate a GARCH model. It is required the existence of the fourth moment because h_t is the sum of squared returns.

Assume $E(r_t^4) < \infty$. Let $\theta = (\alpha_0, \alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q)'$ be the unknown parameters vector. Using the conditional normality, the conditional Gaussian log-likelihood

$$L(\theta) = \frac{1}{n} \sum_{t=1}^{n} l_t. \tag{2.11}$$

Taking the logarithm and neglecting the constant term, we obtain the following log-likelihood function

$$l_t = -\frac{1}{2} ln[h_t(\mathcal{F}_{t-1}; \theta)] - \frac{r_t^2}{2h_t(\mathcal{F}_{t-1}; \theta)}, \tag{2.12}$$

and the maximum value of $L(\theta)$, denoted by $\hat{\theta}$, is the MLE of θ.

2.2.3 The APARCH model

The asymmetry of the effect of positive and negative returns which we have mentioned in the last subsection is regarded as the leverage effect. According to Black (1976), the leverage effect of the stock market is well-known in finance literature and higher volatility responses to negative past returns (bad news), while lower volatility responses to positive past returns (good news) (Nelson, 1991).

The asymmetric power ARCH (APARCH, also called APGARCH) model introduced by Ding et al. (1993) is the formulae of conditional variances different for positive or negative returns. r_t in the APARCH model is similar to the Equation (2.7) in a GARCH (p, q) process. Then, a general APARCH(p, q) is defined as

$$h_t^{d/2} = \alpha_0 + \sum_{i=1}^{p} \alpha_i (|r_{t-i}| - \gamma_i r_{t-i})^d + \sum_{j=1}^{q} \beta_j h_{t-j}^{d/2}, \tag{2.13}$$

where $h_t^{d/2}$ is the conditional standard deviation, $0 < d \leq 2$ is a power index of this model, $\alpha_0 > 0$, $\alpha_i, \beta_j \geq 0$ are similar to those in a GARCH model and $|\gamma_i| < 1$ are introduced to model possible asymmetric information effect. APARCH models include several models as special cases, particularly for cases with $d = 1$ or $d = 2$. This point will be detailed introduced in the following chapter.

The APARCH model includes several ARCH models as special cases. If the values of δ and γ_i are changed, APARCH model derives into the following models, the standard GARCH model, the GJR-GARCH model, the TS-GARCH model(Taylor and Schwert model), the NGARCH model (Nonlinear GARCH model) and the TGARCH model (threshold GARCH model).

When $\delta = 2, \gamma_i = 0$, the APARCH model turns into a GARCH model with the covariance stationary condition for ε_t as $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ (Bollerslev, 1986).

When $\delta = 2, \gamma_i \neq 0$, the APARCH model can be named as the GJR (Glosten, Jagannathan and Runkle, 1993) model

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \gamma_i r_{t-i}^2 I(r_{t-i} < 0) + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{2.14}$$

where $I(\cdot)$ is the indicator function to simulate the asymmetric influence of the positive and negative shocks on the conditional variance.

When $\delta = 1, \gamma_i = 0$, the APARCH model transforms into the TS-GARCH (Schwert, 1989 and Taylor, 1986) model

$$h_t^{1/2} = \omega + \sum_{i=1}^{p} \alpha_i |r_{t-i}| + \sum_{j=1}^{q} \beta_j h_{t-j}^{1/2}. \tag{2.15}$$

When $\delta = 1, \gamma_i \neq 0$, an asymmetric Taylor-Schwert model is obtained, which is named as the TGARCH (threshold GARCH, Zakoïan, 1994) model

$$h_t^{1/2} = \omega + \sum_{i=1}^{p} \alpha_i |r_{t-i}| + \gamma_i |r_{t-i}| I(r_{t-i} < 0) + \sum_{j=1}^{q} \beta_j h_{t-j}^{1/2}. \tag{2.16}$$

2.2.4 The EGARCH model

The exponential GARCH (EGARCH) model introduced by Nelson (1991) is a popular extension of the GARCH model. The standard GARCH model has some limitations compared with EGARCH. To guarantee the conditional variance at each time point to be positive, many restrictions must be added to the parameters. After that, an asymmetric response to shocks can't be treated with standard GARCH models. To overcome the drawback, Nelson (1991) claimed a logarithmic transformation of the volatility and obviously, the adoption of the 'nature device' guaranteed the positivity of the variance without any restrictions on the parameters.

r_t is said to be a family of EGARCH(p, q) models if it satisfies Equation (2.7) and an equation described as the following expression (Nelson, 1991),

$$\ln h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i g(r_{t-i}) + \sum_{j=1}^{q} \beta_j \ln h_{t-j}, \tag{2.17}$$

where

$$g(r_{t-i}) = \vartheta r_{t-j} + \kappa(|r_{t-i}| - E|r_{t-i}|), i = 1, \ldots, p,$$

$\alpha_0, \alpha_i, \beta_i, \vartheta$ and κ are real number coefficients. The model effects, sign effect and

size effect, can be reflected from the parameters α_i and β_i, respectively. A very useful point in an asset pricing is that the sign and the magnitude of r_{t-i} can be allowed to have separate effects on the volatility by the formulation of $g(r_{t-j})$.

According to the Theorem 2.1 from Nelson and the definition of the stationary, when γ and θ do not equal to zero at the same time, the EGARCH process of order (1,1) is strictly stationary and ergodic if and only if $\alpha_1^2 < \infty$ and $|\beta_1| < 1$.

2.2.5 The component GARCH model

The GARCH model has not distinguished the long-term and short-term components. However, it is known that stock prices always fluctuate around an average value. This phenomenon is called mean-revert. It is also found that the mean-revert of short-term volatility is more rapid than for the long-term one and the market volatility must have enough persistence to influence the stock returns in the long-run (Xu and Taylor, 1994). To explain this different response between short-term and long-term, Engle and Lee (1999) introduced the component GARCH model. In this model, the conditional variance is decomposed into a permanent and transitory component. So that the model can be used to investigate the long-run and short-run movements of volatility affecting securities.

The component GARCH (CGARCH) model is defined by

$$h_t = q_t + \sum_{i=1}^{p} \alpha_i (r_{t-i}^2 - q_{t-i}) + \sum_{j=1}^{q} \beta_j (h_{t-j} - q_{t-j}), \tag{2.18}$$

and

$$q_t = \omega + \rho q_{t-1} + \varphi (r_{t-1}^2 - h_{t-1}), \tag{2.19}$$

where q_t the permanent component of the conditional variance and $(h_{t-j} - q_{t-j})$ is the transitory component of the conditional variance (Engle and Lee, 1999).

The parameters ρ is used to examine the persistence of shock impacts on the long-run component. If $0 < \rho < 1$, the long-run volatility component follows an AR process and will converge to a constant level defined by $\omega/(1-\rho)$. When

ρ is extremely close to 1, usually between 0.99 and 1, the long-run volatility component converges to ω very slowly. If $0 < (\alpha_1 + \beta_1) < \rho < 1$, the impact of volatility shocks on the long-run volatility component diminishes as well but be more persistent than that of the short-run component. φ shows the sensitivity of the long-run component to volatility shocks. α expresses the sensitivity of the short-run component to volatility shocks. When $\alpha \geq \varphi$, the immediate impact of volatility shocks on the long-run component would be smaller than that on the short-run component.

2.3 The ACD model

Engle and Russel (1998) proposed an autoregressive conditional duration (ACD) model to analyze the transaction duration. Let $t_0, t_1, \ldots, t_N$ with $t_0 < t_1 < \ldots < t_N$ be a sequence of time, where $N = N(d)$ is a random number and t_i indicates the time of the i-th transaction. The transaction durations are defined as $x_i = t_i - t_{i-1}$, for $i = 1, 2, \ldots, N$. Furthermore let ψ_i be the expectation of the i-th duration

$$E(x_i|x_{i-1}, \ldots, x_1) = \psi_i(x_{i-1}, \ldots, x_i; \theta) \equiv \psi_i. \tag{2.20}$$

The conditional mean interacts multiplicatively with the error term, so that the class of ACD models consists of various parameterizations of (2.20),

$$x_i = \psi_i \varepsilon_i, \tag{2.21}$$

where $\varepsilon_i > 0$ are i.i.d. random variable and $E(\varepsilon_i) = 1$.

Engle and Russell (1998) proposed the ACD(p, q) model and defined the conditional duration by a linear parameter process of ψ_i as:

$$\psi_i = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{k=1}^{q} \beta_k \psi_{i-k}, \tag{2.22}$$

where $\alpha_0 > 0$, $\alpha_j \geq 0$, $\beta_k \geq 0$.

The weak stationarity of an ACD(p, q) process requires to satisfy the following

necessary and sufficient condition,

$$\sum_{j=1}^{p} \alpha_j + \sum_{k=1}^{q} \beta_k < 1, \tag{2.23}$$

then, the unconditional mean of the ACD model is summable ($E(x_i) < \infty$) and under the condition of Eq. (2.23), the unconditional mean of an ACD(p, q) model is similar as the unconditional variance of the GARCH model.

2.4 The semiparametric GARCH model

The GARCH model has many advantages. The function form is accessible, and parameters could be easily estimated. If the model assumptions were correct, the estimation is consistent with reality. However, the drawbacks of parametric volatility models are more upsetting. Firstly, a preselected parametric model may not fit unexpected features, due to too restricted or too low dimensional. Secondly, sometimes the regression function seems to be too complicated or difficult to be defined. Thirdly, because different sequences will be witnessed when different conditional distributions are selected in the process of prediction by using parametric models, there will most possibly exist the problem of misclassification, which may result in a excessively high model bias and loss of efficiency, unless the assumed function perfectly matches the true error distribution (Di and Gangopadhyay, 2011). An important drawback of the parametric volatility models is that the unconditional variance is assumed constant. However, it is found that sometimes the scale change of the time series is not constant significantly (Beran and Ocker, 2001).

Feng (2004) proposed a semiparametric GARCH (SemiGARCH) model by introducing a smooth scale function into the GARCH model. The squared residuals in SemiGARCH models can be estimated by an approximate kernel smoother or the local polynomial smoother. Further after removing the scale function, the parametric parameters are estimated by the classical parametric models estimation approach.

The SemiGARCH model combines a smooth scale function with the standard GARCH model:

$$r_t = \mu + \sigma(x_t)\varepsilon_t, \tag{2.24}$$

where μ is an unknown constant, $x_t = t/n$, $\sigma(x) > 0$ is the nonparametric component, a smooth, bounded scale function and $\{\varepsilon_t\}$ is the parametric component.

The conditional variance of $\{\varepsilon_t\}$ is assumed to follow a GARCH(p, q) process:

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{2.25}$$

where $h_t^{1/2}$ is the conditional standard deviations of the standardized process ε_t, $\omega > 0$; $\alpha_1, \ldots, \alpha_p \geq 0$ and $\beta_1, \ldots, \beta_q \geq 0$. To estimate the scale function, $E(r_t^8) < \infty$ is assumed to ensure Eq. (5.2) strictly stationary, which implies in particular that $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ (Feng, 2004).

The SemiGARCH model provides us a tool to decompose financial risk into an unconditional component $\sigma(x_t)$, a conditional component $h_t^{1/2}$ and the i.i.d. innovations η_t.

The estimation of the SemiGARCH model can combine the nonparametric estimation of the local variance $v(x) = \sigma^2(x)$, with parametric estimation of the unknown parameter vectors $\theta = (\alpha_0; \alpha_1, \ldots, \alpha_p; \beta_1, \ldots, \beta_q)$.

At first, the scale function can be estimated by some nonparametric regression approaches without any parametric assumptions. In this model the kernel estimation will be used. If the constant mean μ is replaced by a smooth function g, we can get a nonparametric regression with scale change and dependence

$$r_t = g(x_t) + \sigma(x_t)\varepsilon_t, \tag{2.26}$$

where ε_t is a zero mean stationary process.

Therefore, Eq. (2.24) can be transformed into a general nonparametric regression problem. Letting $r_t^* = r_t - \mu$, $Z_t = (r_t^*)^2$ and $\xi_t = \varepsilon_t^2 - 1 \geq -1$, which are zero mean stationary time series errors. Then, the model can be rewritten as

$$Z_t = v(x_t) + v(x_t)\xi_t. \tag{2.27}$$

Letting $\hat{\mu} = \bar{r}$ and $\hat{z}_t = (\hat{r}_t^*)^2$, in which $\hat{r}_t^*$ is then defined by $\hat{r}_t^* = r_t - \bar{r}$. The Nadaraya-Watson kernel regression is defined by

$$\hat{v}(x) = \frac{\sum_{t=1}^{n} K(\frac{x_t - x}{b})\hat{z}_t}{\sum_{t=1}^{n} K(\frac{x_t - x}{b})} =: \sum_{t=1}^{n} w_t \hat{z}_t, \tag{2.28}$$

where w_t is the weighting function $w_t = \frac{K(\frac{x_t - x}{b})}{\sum_{t=1}^{n} K(\frac{x_t - x}{b})}$, $K(u)$ is a second order kernel function with compact support [-1,1] and b is the bandwidth, the size of the weights (Fan, 1993).

Besides, $\nu(x_t)$ can also be estimated by the local linear regression. The local linear estimator of $\nu(x_t)$ at $0 \leq x_t \leq 1$ is obtained by minimizing

$$Q(x) = \sum_{t=1}^{n} \{r_t - \alpha_0 - \alpha_1(x_t - x)\}^2 K\left(\frac{x_t - x}{b}\right). \tag{2.29}$$

Obviously, we obtain $\tilde{\nu}(x_t) = \hat{\alpha}_0$. The bias of the local linear estimator is always of the order $O(b^2)$, which is important for application, because the forecasting of the trend is mainly carried out based on the estimation at the right boundary. However, $\tilde{\nu}(x_t)$ obtained above is sometimes negative, especially with a small sample size or bandwidth. To ensure the non-negativity, the constrained local linear regression is considered. We propose to use the constrained local estimator $\hat{\nu}(x_t) = |\tilde{\nu}(x_t)|$, which is a.s. positive. In this dissertation, the scale function is estimated via codes by Feng (2004). After the time-varying trend is removed, the descaled data is able to be applied in a parametric process.

According to the above assumptions, the estimator ε_t is now replaced by the standardized residuals

$$\hat{\varepsilon}_t = \hat{r}_t / \hat{\sigma}(x_t) = (r_t - \bar{r})/\hat{\sigma}(x_t). \tag{2.30}$$

Then the estimator of parametric vector θ can be obtained by the standard maxi-

mum likelihood method, which has been introduced in subsection 2.1.2. A suitable model can also be selected by using other methods, e.g. the Akaike information criterion (AIC), the Bayesian information criterion (BIC), etc. In the thesis, the parametric models are estimated by the *fGarch*[1] and *rugarch*[2] packages in R.

2.5 The semiparametric ACD model

Consider a stochastic process $\{t_0, t_1, \ldots, t_n, \ldots\}$ with $t_0 < t_1 < \ldots < t_n < \ldots$ and it represents a sequence of time series. $x_i = t_i - t_{i-1}$ defines the interval between the durations and ψ_i is the expectation of the ith duration. According to the arrival times, $N(t)$ refers to the counting function.

The duration expectation is defined as,

$$E[x_i|x_{i-1}, \ldots, x_{i-p}, \psi_{i-q}] = \vartheta[x_i|x_{i-1}, \ldots, x_{i-p}, \psi_{i-1}, \ldots, \psi_{i-q}, \Omega] \equiv \psi_i, \quad (2.31)$$

in the formula, x_i follows the definition in parametric ACD models,

$$x_i = \psi_i \varepsilon_i, \quad (2.32)$$

where ε_i are i.i.d. with unit mean.

Feng (2014) proposed a local linear method to estimate the diurnal pattern in the SemiACD model. Consider the diurnal pattern of the ACD model (Russell and Engle, 2002, 2010),

$$x_i = \phi(t_i)\psi_i \varepsilon_i, \quad (2.33)$$

where $\phi(t_i)$ is the deterministic diurnal pattern and the local mean of x_i, ψ_i is the conditional expectation of the diurnally durations above or below the average value of the day.

[1] *fGarch* is written by Wuertz et al. in *Rmetrics-Autoregressive Conditional Heteroskedastic Modeling.*

[2] *rugarch* is developed by Ghalanos and Kley in *Univariate GARCH Models.*

Let $y_i = \psi_i \varepsilon_i$ and $E(y_i) = 1$, model (2.33) can be rewritten as

$$x_i = \phi(t_i) + \phi(t_i)\xi_i, \tag{2.34}$$

where $\xi_i = y_i - 1$. It is obvious, model (2.34) discusses the estimation of the scale and mean function $\phi(t_i)$ to a general nonparametric regression process. If ψ_i follow a unit ACD process,

$$\psi_i = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{k=1}^{q} \beta_k \psi_{i-k}, \tag{2.35}$$

then, model (2.33) and (2.36) are a semiparametric ACD process and can be estimated by a semiparametric procedure combining nonparametric estimation of $\phi(t_i)$ and parametric estimation of unit the ACD model. In the SemiACD model, because of the unit mean of y_i, the constant parameter α_0 is

$$\alpha_0 = 1 - \sum_{j=1}^{p} \alpha_j - \sum_{k=1}^{q} \beta_k. \tag{2.36}$$

Here, the constant in the ACD process cannot be chosen freely.

Due to $\phi(t_i)$ depending on N, we define $\phi_N(t_i) = m(t_i)/N$, then Eq. (2.34) can be written as

$$x_i^* \approx m(t_i) + m(t_i)\xi_i, \tag{2.37}$$

The estimation of the scale function $\phi(t_i)$ in the SemiACD model can apply the local polynomial method by minimizing the least squares to estimate $m(t_i)$. Let $K(u)$ be a kernel function and $b > 0$ be the bandwidth. The local linear estimator of $m(t_i)$ is obtained by means of the following locally weighted least square approach

$$Q(b) = \sum_{i=1}^{N} \{x_i^* - \alpha_0 - \alpha_1(t_i - t)\}^2 K\left(\frac{t_i - t}{b}\right) \Rightarrow \min, \tag{2.38}$$

where $\tilde{m}(t_i) = \hat{\alpha}_0$ is the local linear estimate and the kernel function K may be different to that used above. The bias of $\tilde{m}(t_i)$ is always of the order $O(b)$, which is important to the forecasting based on the estimation at the right end point.

2.6 Final remarks

In this dissertation, the considered data is decomposed by the semiparametric models by removing a scale function. In Chapter 3 and Chapter 4, GARCH models are applied as the parametric part with normal and t-distribution of innovations, respectively. Some GARCH model extensions are used in the stationary process, such as APARCH, EGARCH and CGARCH models. In Chapter 6 and 7, ACD models are discussed to analyze the nonnegative financial data.

Chapter 3

A class of SemiGARCH models estimated based on the Box-Cox transformation[1]

The chapter proposes a wide class of semiparametric GARCH models by introducing a scale function into a GARCH class model for featuring long-run volatility dynamics, which can be thought of as an MEM (multiplicative error model) with a varying scale function. The focus is to estimate the scale function under suitable weak moment conditions through the Box-Cox transformation of the absolute returns. The estimation of the scale function is independent of any GARCH specification. To overcome the drawbacks of the kernel and the local linear approaches, a non-negatively constrained local linear estimator of the scale function, which is then proposed to fit a suitable parametric GARCH model to the standardized residuals, is considered. Asymptotic properties of the proposed nonparametric and parametric estimators are studied in detail and iterative plug-in algorithms are developed for selecting the bandwidth and transformation parameters, which are selected by MLE and JB statistic. The algorithms are also carried out independently without any parametric specification in the stationary part. Application shows that the proposals fit well to real data.

[1]Chapter 3 is based on the working paper: *A general class of SemiGARCH models based on the Box-Cox Transformation* (Zhang et al., 2017), CIE, 2017–06.

3.1 Introduction

Despite the success of the ARCH (autoregressive conditional heteroskedasticity, Engle, 1982) and GARCH (generalized ARCH, Bollerslev, 1986) models for modeling conditional (short-run) volatility dynamics in stock market returns, their implications for long-run volatility are restrictive, in the sense that these models imply a constant unconditional long-run volatility, i.e. it implies that the stock market returns are stationary. However, in recent years it was realized that this feature does not seem to be consistent with the time series behavior of volatilities of stock returns. Different extensions of the standard GARCH model are hence proposed for capturing the long-run volatility patterns observed in the data. For instance, Feng (2004) introduced a SemiGARCH (semiparametric GARCH) model by employing a smooth volatility trend (also called the scale function) into the standard GARCH model and proposed to estimate it using data-driven kernel regression. Van Bellegem and von Sachs (2004) discussed the forecasting of financial time series under the time-varying unconditional variance. A general time-varying ARCH process was introduced by Dahlhaus and Rao (2006). Engle and Rangel (2008) put forward a Spline-GARCH model with a nonparametric volatility trend, which is defined as a function of the observation time, i.e. the location, and exogenous macroeconomic variables and is estimated by an exponential quadratic spline. Engle et al. (2008) extended this idea to a GARCH-MIDAS model, which combines the ideas of the Spline-GARCH model and of mixed data sampling (MIDAS), to investigate detailed macroeconomic sources of long-run dynamics of stock market volatility. Peitz (2015) developed a spatial semiparametric process to analyze high-frequency data in different dimensions. Amado and Teräsvirta (2017) developed a specification technique for building multiplicative time-varying GARCH models by decomposing the variance into a conditional and an unconditional component, which is smooth over time.

In this chapter, a general class of semiparametric GARCH models is introduced, including the SemiGARCH model as a special case. Similar to the SemiGARCH model, the total volatility is defined as a product of a scale function and a conditional volatility component and the effect of exogenous variables is not con-

sidered. The key difference between the current proposal and the SemiGARCH model is that here the parametric part is not specified beforehand but to be chosen after estimating the scale function. Different specifications will lead to different models. By rewriting the GARCH formulations, it is shown that a semiparametric GARCH model is asymptotically equivalent to the GARCH model used in the parametric part with a time-varying scale parameter, while the other parameters remain constant. This provides us with a deep insight into the current proposal and indicates possible further extensions of it. To estimate the scale function, we propose the use of a constrained non-negative local linear regression, to ensure that the resulting scale function is (at least almost surely) positive. It is shown that the constrained local linear regression defined in this chapter is asymptotically equivalent to the common local linear regression. Note that the data-driven algorithm proposed by Feng (2004) does not apply to the general framework considered in this chapter because of the Box-Cox power transformation in the scale function. MLE and Jarque-Bera (JB) statistics are applied in the selection of the transformation parameter. Hence, the main focus is on the development of a quick, stable data-driven algorithm for the practical implementation of the general semiparametric GARCH approach. For this purpose, a fully data-driven iterative plug-in bandwidth selector algorithm is proposed following Gasser et al. (1991), Herrmann et al. (1992) and Beran and Feng (2002a, b). The application to data examples shows that such bandwidth and transformation parameter selection rules work well. Furthermore, a simple test is introduced to determine, if a semiparametric GARCH or a parametric GARCH model should be used. This test shows that the unconditional volatility during a financial crisis is significantly higher than that in other sub-periods. It seems to be possible to develop a suitable method for detecting the effect of a financial crisis by means of the proposal in this chapter. Further, the estimation and selection of a suitable parametric GARCH model based on the standardized returns are also discussed. Some results in this chapter can be easily adapted to the Spline-GARCH or GARCH-MIDAS models. For instance, both of them are GARCH models with a varying scale parameter determined by the time and other exogenous variables.

The chapter is organized as follows. The model is introduced in Section 3.2. Section 3.3 discusses the semiparametric estimation of the proposed model, the data-driven algorithm and the test method. Data examples in Section 3.4 illustrate the practical usefulness of the proposal. Final remarks in Section 3.5 conclude the chapter. Sketched proofs of some results are given in the appendix.

3.2 The SemiGARCH model with Box-Cox transformation

Let y_t, $t = 0, 1, ..., n$, denote the prices of some stock index and r_t their (log-) returns. To model the slowly changing unconditional variance and conditional heteroskedasticity at the same time, the following semiparametric GARCH class model (Feng, 2004) for the conditional distribution of r_t is introduced:

$$r_t^* = \mu(\tau_t) + s(\tau_t)\sqrt{h_t}\varepsilon_t, \tag{3.1}$$

where $\tau_t = t/n$ is the rescaled time, $\mu(\cdot)$ stands for a smooth trend, $s(\cdot) > 0$ is a smooth scale function and h_t is the conditional variance of the rescaled process $\xi_t = r_t/s(\tau_t) = \sqrt{h_t}\varepsilon_t$ with the centralized returns $r_t = r_t^* - \mu(\tau_t)$. Due to the returns are a.s. distinct from the expectation, the demeaned method (e.g. Harvey et al., 1994) is widely applied to guarantee the centralized returns a.s. positive and it also ensures the possible logarithmic transformation of the non-negative series in the following discussion. It is assumed that ξ_t also has unit variance so that the model is uniquely defined. This implies that the unconditional mean of h_t is 1, i.e. $E(h_t) = 1$. Although our focus is on the estimation of $s^2(\cdot)$ and h_t, a nonparametric trend function is included for modeling possible long-term deterministic changes in the mean of y_t. We will see that the asymptotic properties of $\hat{s}^2(\cdot)$ will not be affected by the estimation errors in $\hat{\mu}(\cdot)$. Model (3.1) defines indeed a sequence of models. The process r_t is nonstationary unless $\mu(\cdot)$ and $s(\cdot)$ are both constant. But r_t is locally stationary following Dahlhaus (1997). The trend function $\mu(\cdot)$ can be recognized as a time-varying function $\mu(\tau_t)$ or in practice, returns may also have

no nonparametric trend function. For simplicity, it will not be considered in the current chapter, because our focus is on the estimation of $s(\cdot)$ and σ_t. Moreover, it is well known that under common regularity conditions the effect of the error in a nonparametric estimator of an unknown trend function on the estimation of $s(\tau_t)$ is asymptotically negligible. Then Model (3.1), without the trend function, is reduced to

$$r_t = s(\tau_t)\sqrt{h_t}\varepsilon_t. \tag{3.2}$$

Model (3.2) is a general SemiMEM (semiparametric multiplicative error model) defined by introducing a smooth scale function into the MEM proposed by Engle (2002). Hence, all of the results given in this chapter hold for a model with a nonparametric trend function provided that the trend function is estimated by another well-developed data-driven algorithm.

The stationary process ξ_t can be analyzed using any suitable GARCH class model and different parametric specifications on h_t will lead to different semiparametric GARCH class models. If it is assumed that ξ_t follows a standard GARCH model, we have

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \xi_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}, \tag{3.3}$$

where $\alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p \geq 0$, $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ and $\alpha_0 = 1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j$. Due to the restriction $E(h_t) = 1$, α_0 is no more a free parameter. Equations (3.2) and (3.3) together define the SemiGARCH model introduced by Feng (2004). See also Feng and McNeil (2008) for an extension of this model to high-frequency financial data.

Nonparametric estimation of variance functions is well known in the literature. Local polynomial estimation of variance functions in nonparametric regression with independent errors is studied e.g. by Ruppert et al. (1997). Kernel and local linear estimators of conditional variance in nonlinear time series are proposed by Feng and Heiler (1998) and Fan and Yao (1998), respectively. A kernel estimator

has the so-called boundary problem, which will not only affect the bias of the estimate at a boundary point but also reduce the convergence rate of the MISE (mean integrated squared error). A modified kernel estimator with second-order bias at a boundary point is proposed by Hall and Presnell (1999). However, their proposal does not apply to the endpoints $\tau = 0$ or $\tau = 1$. A local linear estimator does not share this problem, however, it may be sometimes negative. In this section, we will hence propose the use of local linear estimators with a simple non-negative restriction. Moreover, a consistent estimator of $s^2(\tau_t)$ as a smoother of ξ_t^2 with a given bandwidth requires the existence of the fourth moments of ξ_t. But the selection of the bandwidth, in this case, requires the existence of the eighth moments of ξ_t. This will clearly affect the stability of the estimated scale function. This drawback hinders the application of such non-parametric variance estimators to financial data series, because the marginal distribution of a financial time series may have heavy tails. To solve this problem, we propose to estimate the scale function from the Box-Cox power transformation $|r_t|^\lambda$ with $0 < \lambda \leq 2$. If $\lambda \leq 1$ is used, the existence of the fourth moment of ξ_t is sufficient for developing a convergent bandwidth selector. If the power transformation parameter is regarded as $\lambda = 0$, the Box-Cox transformation of $|r_t|^\lambda$ reduces to a logarithmic form, and obviously, ξ_t will follow a logarithmic process. The logarithmic transform can be applied to some financial variables, such as realized volatility, volatility indexes, etc., so as to convert multiplicative models to additive models. For simplicity, $0 < \lambda \leq 2$ is applied to the Box-Cox transformation without the consideration of logarithmic transformation in the chapter. Note that normally $\lambda = 2$ is used. Now, $s^2(\tau_t)$ is estimated first. $\hat{s}(\tau_t)$ is then obtained by taking the square root of $\hat{s}^2(\tau_t)$. Our proposal is to estimate the local mean of $|r_t|^\lambda$ first and then take the λ-th root of this local mean as an estimator of the scale function. The relationship between this estimate and the classical one is as follows. Define $c_\lambda = E(|\xi_t|^\lambda)$, which is 1 for $\lambda = 2$ following the definition. For $\lambda \neq 2$, we have $c_\lambda \neq 1$ but its concrete value depends on the distribution of ξ_t and will change from case to case. We will see that a nonparametric estimator based on $|r_t|^\lambda$ is indeed an estimator of $g(\tau_t, \lambda) = c_\lambda s^\lambda(\tau_t)$ and not that of $s^\lambda(\tau_t)$. Hence $[\hat{g}(\tau_t, \lambda)]^{1/\lambda} \approx c_\lambda^{1/\lambda} \hat{s}(\tau_t) \neq \hat{s}(\tau_t)$, if $\lambda \neq 2$. To estimate the scale function from r_t^2 is the most natural method.

However, note that the difference between $[g(\tau_t, \lambda)]^{1/\lambda}$ and $s(\tau_t)$ is just a constant factor. Hence the use of $[g(\tau_t, \lambda)]^{1/\lambda}$ as an alternative scale function is equivalent to the use of $s(\tau_t)$. Thus, for given λ, Model (3.2) can be rewritten as

$$r_t = s_\lambda(\tau_t)\xi_{\lambda,t}, \tag{3.4}$$

where $s_\lambda(\tau_t) = c_\lambda^{1/\lambda} s(\tau_t) = [g(\tau_t, \lambda)]^{1/\lambda}$ and $\xi_{\lambda,t} = c_\lambda^{-1/\lambda}\xi_t$ is another stationary process with $E(|\xi_{\lambda,t}|^\lambda) = 1$. Obviously, $\xi_{\lambda,t}$ and ξ_t share the same properties but with a different scale parameter. Hence the resulting estimator based on $|r_t|^\lambda$ can be used to remove the effect of the slowly changing scale in r_t. For $\lambda = 2$ we have $s_2(\tau_t) = s(\tau_t)$. Otherwise, $s_\lambda(\tau_t)$ and $s(\tau_t)$ have different scale parameters. We see $s_\lambda(\tau)$ can also be used as the scale function of the proposed model, which can be estimated consistently from $|r_t|^\lambda$. There are different further transformations which can be used to estimate an equivalent scale function. The power transformations (or equivalently the Box-Cox transformations with non-negative power transformation parameter) are just the simplest examples. Please refer to Eagleson and Müller (1997) for more general description on this point.

It is clear that model (3.4) is an improved alternative of model (3.2) based on the Box-Cox transformation. Model (3.3) and Model (3.4) together can be proposed as the Box-Cox SemiGARCH model, providing a new semiparametric methodology by introducing a power transformation parameter λ into the scale function. Similar to the time-varying GARCH models, any kind of GARCH models can be selected as an extension in the parametric part of generalized Box-Cox SemiGARCH class models. If $\lambda = 2$ and the parametric part is a GARCH process, it is the standard SemiGARCH model proposed by Feng (2004). The Semi-APARCH model (Feng and Sun, 2013) is also another specification, applying the absolute returns and APARCH model in the parametric part.

3.3 The semiparametric estimation procedure

The generalized Box-Cox SemiGARCH class models introduced in the last section can be estimated using a semiparametric procedure. At first, $s_\lambda(\tau_t)$ can be estimated by some nonparametric regression approach consistently without any parametric assumptions on σ_t and ε_t. In the section, local polynomial regression is applied. The slowly changing scale function can be estimated and removed under very weak moment conditions $E(\xi_t^\lambda) < \infty$ for any $\lambda > 0$ based on suitable power transformation of the data. A simple constrained local polynomial regression, which is approximately the same as the standard local polynomial regression, is proposed to ensure that the resulting scale function is always positive. Then the conditional variance can be analyzed further using the GARCH class models based on the standardized returns.

3.3.1 Estimation of $s(\tau_t)$

Let $\zeta_{t,\lambda} = |\xi_{\lambda,t}|^\lambda - 1$ with $E[\zeta_{t,\lambda}] = 0$. Model (3.4) can be expressed as

$$|r_t|^\lambda = g(\tau_t, \lambda) + g(\tau_t, \lambda)\zeta_{t,\lambda}, \tag{3.5}$$

which is a nonparametric regression with heteroskedastic time series errors and $g(\tau_t, \lambda)$ is the trend function and the scale function at the same time. Let $K(u)$ be a kernel function and $b > 0$ be the bandwidth. A local linear estimator of $g(\tau_t, \lambda)$ at $0 \leq \tau_t \leq 1$ is obtained by minimizing

$$Q(\lambda, b) = \sum_{t=1}^{n} \left\{ |r_t|^\lambda - \alpha_0 - \alpha_1(\tau_t - \tau) \right\}^2 K\left(\frac{\tau_t - \tau}{b}\right). \tag{3.6}$$

This results in $\tilde{g}(\tau_t, \lambda) = \hat{\alpha}_0$. The advantage of a local linear estimator is that the bias of it is always of the order $O(b^2)$. This is in particular important for application, because the forecasting of the trend is mainly carried out based on the estimation at the right end point. A problem is that $\tilde{g}(\tau_t, \lambda)$ obtained above is sometimes negative, in particular when the sample size is small and a small bandwidth is used. To ensure the non-negativity, we propose to use the final

estimator $\hat{g}(\tau_t, \lambda) = |\tilde{g}(\tau_t, \lambda)|$, which is almost surely positive. The use of $\hat{g}(\tau_t, \lambda)$ instead of $\tilde{g}(\tau_t, \lambda)$ is reasonable. Firstly, it can be shown that $[\hat{g}(\tau_t, \lambda) - g(\tau_t, \lambda)]^2 \leq [\tilde{g}(\tau_t, \lambda) - g(\tau_t, \lambda)]^2$. That is the performance of $\hat{g}(\tau_t, \lambda)$ is not worse than that of $\tilde{g}(\tau_t, \lambda)$ following the MSE (mean squared error). Moreover, negative values of $\tilde{g}(\tau_t, \lambda)$ are just a limited sample problem, because the probability that $|\hat{g}(\tau_t, \lambda) - \tilde{g}(\tau_t, \lambda)| > \Delta$ for any $\Delta > 0$ tends to zero in an exponential rate. This is shown in the following lemma, where Assumptions A1 to A4 are described in the appendix.

Lemma 3.1 *Suppose that a bandwidth of the order $b = n^{-\lambda}$ with $0 < \lambda < 1$ is used and $\tilde{g}(\tau_t, \lambda)$ is consistent, asymptotically normal with bias $B[\tilde{g}(\tau_t, \lambda)] = O(n^{-\eta_1})$ and variance $Var[(\tilde{g}(\tau_t, \lambda))] = O(n^{-\eta_2})$, where $\eta_1, \eta_2 > 0$. If the assumptions A1 to A4 hold, then we have*

$$nP[\hat{g}(\tau_t, \lambda) \neq \tilde{g}(\tau_t, \lambda)] = nP[\tilde{g}(\tau_t, \lambda) < 0] \to 0, \text{ as } n \to \infty. \tag{3.7}$$

The result of the lemma also holds if n is replaced by n^k for any $k > 1$, e.g. $k = 2$. Hence, when n and b are both large, then $\tilde{g}(\tau_t, \lambda) < 0$ will practically never happen, if b is large enough. Also, there is no difference between the asymptotic properties of $\tilde{g}(\tau_t, \lambda)$ and $\hat{g}(\tau_t, \lambda)$. Note that r_t are uncorrelated. The scale function $\hat{g}(\cdot)$ defined above has the same asymptotic properties as those for a nonparametric regression estimator with independent errors and a non-constant scale function. For more theoretical discussions on these topics, we refer the reader to Beran et al. (2015), where the estimation of the scale function in a semiparametric ACD model for daily average transaction duration is considered. The authors also obtained detailed asymptotic results of the constrained local linear estimator in that context. According to the similarity between the ACD and the GARCH models, asymptotic results of $\hat{g}(\tau_t, \lambda)$ can be derived based on their results by replacing the average durations there with $|r_t|^\lambda$.

The key idea behind our proposal is that although $\hat{g}^{1/\lambda}(\tau_t, \lambda)$ is not a consistent estimator of $s_\lambda(\tau_t)$, it can be directly used to remove the non-stationarity in returns, because $\hat{\xi}_{\lambda,t} = r_t / \hat{g}^{1/\lambda}(\tau_t, \lambda)$ is also an approximately stationary process.

Comparing the general formula (3.5) with the special case with $\lambda = 2$, we can see that, instead of the estimation of the scale function $s(\tau_t)$ in r_t, here the scale function in the process $r_{\lambda,t}$ is indeed directly estimated, where $r_{\lambda,t} = \text{sign}(r_t)|r_t|^{\lambda/2}$ with $r_{2,t} = r_t$. As far as we know, there is still no study in the literature on the estimation of the scale function in a SemiGARCH framework based on the power transformation $|r_t|^\lambda$. The main purpose is to develop a consistent data-driven estimator of the nonparametric scale function $s_\lambda(\tau_t)$. If higher robustness is of interest, $\lambda \leq 1$ can be used and the assumptions that $E(\xi_t^4) < \infty$ together with further regularity conditions is sufficient for developing a convergent bandwidth selector. This is the same moment condition required for estimating the GARCH parameters using conditional QMLE (quasi maximum likelihood estimation). In this section, we will still consider the use of $\lambda \leq 2$ and in Fig. 3.2, we can see the selected λ are obviously smaller than 1, which means the stricter robustness requirement can be fulfilled.

Model (3.5) is an extension of Model (4) in Feng (2004), where only the special case with $\lambda = 2$ is considered. Asymptotic properties of $\hat{g}(\tau_t, \lambda)$ can hence be proved analogously. The following summarizes and compares the asymptotic behavior of $\hat{s}_\lambda(\tau_t)$ and $\hat{g}(\tau_t, \lambda)$, where $\text{MSE}[\hat{s}_\lambda(\tau_t), b]$ and $\text{MSE}[\hat{g}(\tau_t, \lambda), b]$ denote the mean squared error of the two estimators obtained with the bandwidth b. Assume now that $E(\xi_t^4) < \infty$, a consistent estimate of $s_\lambda(\tau_t)$ can be obtained as follows. Note that $\hat{\xi}_{\lambda,t} \approx c_\lambda^{-1/\lambda}\hat{\xi}_t$ and that $E(\xi_t^2) = 1$. This leads to a consistent estimate of c_λ

$$\hat{c}_\lambda = \left[\frac{1}{n}\sum_{t=1}^{n}\hat{\xi}_{\lambda,t}^2\right]^{-\lambda/2} \tag{3.8}$$

and

$$\hat{c}_\lambda^{-1/\lambda} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\hat{\xi}_{\lambda,t}^2}. \tag{3.9}$$

The scale function is obtained as

$$\hat{s}_\lambda(\tau_t) = \hat{c}_\lambda^{-1/\lambda}[\hat{g}(\tau_t, \lambda)]^{1/\lambda}, \tag{3.10}$$

through rescaling the sample variance of the standardized returns to be one. Note that, so long as $\hat{g}(\tau_t, \lambda)$ is consistent, the effect of the error in it on $\hat{c}_\lambda$ is asymptot-

ically negligible. Hence both of $\hat{c}_\lambda$ and $\hat{c}_\lambda^{-1/\lambda}$ are still $\sqrt{n}$-consistent. It leads to the conclusion that the MSE of $\hat{s}_\lambda(\tau_t)$ in this way is approximately $c_\lambda^{-2/\lambda}\text{MSE}[\hat{g}(\tau_t,\lambda)]$, which is still of the order $O(n^{-4/5})$, provided that $\hat{g}(\tau_t,\lambda)$ is obtained by a suitable data-driven algorithm. However, $\hat{s}_\lambda(\tau_t)$ is not an efficient estimate if $\lambda \neq 2$ is used, because the optimal bandwidth for estimating $g(\tau_t,\lambda)$ is different to that for estimating $s_\lambda(\tau_t)$. Moreover, we see that to obtain a consistent nonparametric estimate of $s_\lambda(\tau_t)$, the condition $E(\xi_t^4) < \infty$ is also necessary. To avoid possible confusion, we propose to estimate $g(\tau_t,\lambda)$ for some chosen λ and to calculate $\hat{\xi}_{\lambda,t}$ at first. Then we can obtain $\hat{c}_\lambda$ and $\hat{s}_\lambda(\tau_t)$ following (3.8) through (3.10). Finally, we will calculate the standardized returns $\hat{\xi}_t = r_t/\hat{s}_\lambda(\tau_t)$ again, which are approximately independent of the choice of λ and will be used for further analysis.

Suppose that $E(\xi_t^\lambda) < \infty$ and the Assumptions A1 to A4 stated in the appendix hold. For any $0 < \lambda \leq 2$ and $0 \leq \tau_t \leq 1$, both $\hat{s}_\lambda(\tau_t)$ and $\hat{g}(\tau_t,\lambda)$ are consistent estimators of $s_\lambda(\tau_t)$ and $g(\tau_t,\lambda)$, respectively. Further, it also holds that $\text{MSE}[\hat{s}_\lambda(\tau_t),b] \approx c_\lambda^{-2/\lambda}\text{MSE}[\hat{g}(\tau_t,\lambda),b]$, and $\hat{s}_\lambda(\tau_t)$ and $\hat{g}(\tau_t,\lambda)$ have the same asymptotically optimal bandwidth. The finding of particular interest is that $\hat{s}_\lambda(\tau_t)$ and $\hat{g}(\tau_t,\lambda)$ have the same asymptotically optimal bandwidth. Note that our aim is to estimate $s_\lambda(\tau_t)$. However, it is straight-forward to select the bandwidth for $\hat{g}(\tau_t,\lambda)$. The bandwidth is just what we need for an optimal estimate of $s_\lambda(\tau_t)$. So the problem is solved well. There is no need to develop a separate bandwidth selection procedure for estimating $s_\lambda(\tau_t)$. Asymptotic properties of $\hat{g}(\cdot)$ can be obtained following known results in nonparametric regression with dependent errors (see e.g. Altman, 1990 and Hart, 1991). In the sequel, some necessary results are summarized. For a kernel function K, define $R(K) = \int K^2(u)du$ and $I(K) = \int u^2K(u)du$. Let $S_{|\xi|^\lambda} = \sum_{k=-\infty}^{\infty} \gamma_{|\xi|^\lambda}(k)$, where $\gamma_{|\xi|^\lambda}(k) = \text{cov}\,(|\xi_0|^\lambda, |\xi_k|^\lambda)$, then the following holds.

Theorem 3.1 *Under the Assumptions A1 through A5 stated in the appendix, the following holds,*

i) The bias of $\hat{g}(\tau_t,\lambda)$ *is* $B[\hat{g}(\tau_t,\lambda)] = E[\hat{g}(\tau_t,\lambda)] - g(\tau_t,\lambda) = \frac{1}{2}b^2[g(\tau_t,\lambda)]''I(K) + o(b^2)$.

ii) The variance of $\hat{g}(\tau_t, \lambda)$ *is given by*

$$Var\left(\hat{g}(\tau_t, \lambda)\right) = \frac{V}{nb} + o\left(\frac{1}{nb}\right), \tag{3.11}$$

where $V = S_{|\xi|^\lambda}[g(\tau_t, \lambda)]^2 R(K)$.

iii) If a bandwidth $b = o(b_A)$ *is used, the bias is asymptotically negligible and*

$$\sqrt{nb}[\hat{g}(\tau_t, \lambda) - g(\tau_t, \lambda)] \xrightarrow{\mathcal{D}} N(0, V), \tag{3.12}$$

where V *is as defined in (3.11).*

In Theorem 3.1, the asymptotic bias and variance of $\hat{g}(\tau_t, \lambda)$ are obviously similar to those in nonparametric regression with some specific GARCH or ACD class model, because of the similarity in calculating the sum of the autocovariance. The result of Theorem 3.1 *iii*) also indicates that $\hat{g}(\tau_t, \lambda)$ is asymptotically unbiased and asymptotically normal with a bandwidth of a smaller order than b_{A}.

Theorem 3.2 *Suppose that Assumptions A1 to A4 hold, we have:*
i) At any point $0 < \tau_t < 1$, *the local asymptotically optimal bandwidth, which minimizes the dominating part of the* MSE *(mean squared error) of* $\hat{g}(\tau_t, d)$, *is given by*

$$b_A(\tau_t) = \left(S_{|\xi|^\lambda} \frac{R(K)}{I^2(K)} \frac{g^2(\tau_t, \lambda)}{\{[g(\tau_t, \lambda)]''\}^2}\right)^{1/5} n^{-1/5}. \tag{3.13}$$

ii) Let MISE$=\int_0^1 \{\mathrm{MSE}[\hat{g}(\tau_t, d)]\} d\tau$ *be the mean integrated squared error. Then the (global) asymptotically optimal bandwidth minimizing the dominating part of the* MISE *is given by*

$$b_{\mathrm{A}} = \left(S_{|\xi|^\lambda} \frac{R(K)}{I^2(K)} \frac{\int g^2(\tau, \lambda) d\tau}{\int \{[g(\tau, \lambda)]''\}^2 d\tau}\right)^{1/5} n^{-1/5}. \tag{3.14}$$

Results in Theorem 3.2 are closely related to those for a local linear estimator of the mean function with heteroskedastic time series errors. Furthermore, note that the result in the first part of Theorem 3.2 does not hold at a boundary point, because the kernel constants in the asymptotic variance and asymptotic bias of

$\hat{g}(\tau_t, \lambda)$ at the boundary change from point to point. But this does not affect the asymptotic MISE so that the global bandwidth can be calculated over $\tau_t \in [0, 1]$.

3.3.2 Semiparametric estimation of a given model

The unknown parameters of chosen GARCH class models can be estimated from $\hat{\xi}_t$ by approximate (conditional) QMLE method proposed in the literature. A suitable model can also be selected using e.g. the BIC.[2]

Denote the true unknown parameter vector of a chosen GARCH model by θ_0. Let $\hat{\theta}$ be the estimate of θ_0 obtained from $\hat{\xi}_t$ and $\tilde{\theta}$ denotes the standard QMLE obtained under the assumption that ξ_t is observable. It is well known that under suitable regularity conditions $\tilde{\theta}$ is $\sqrt{n}$-consistent and asymptotically normal. The additional variance caused by the errors in $\hat{\xi}_t$ is asymptotically negligible. The $O(b^2)$ term in B_θ is due to $E[\hat{s}_\lambda(\tau_t) - s_\lambda(\tau_t)]$ and the $O[(nb)^{-1}]$ term due to $\mathrm{Cov}\,[\xi_t^2, \hat{s}_\lambda(\tau_t)]$. If a bandwidth $O(n^{-1/2}) < b < O(n^{-1/4})$ is used, B_θ is asymptotically negligible. Now $\hat{\theta}$ is also $\sqrt{n}$ consistent and asymptotically normal. If the data-driven algorithm proposed in the next section is used, the bias term B_θ will be of the order $O(n^{-2/5})$. We see that in the general SemiGARCH models $\sqrt{n}$-consistent parametric estimation is no longer possible if the scale function changes over time. In the special case, when r_t follows a stationary GARCH class model, a bandwidth of the order $O_p(1)$ will be selected by the proposed the data-driven algorithm in the next section. Now, the parametric estimation is still $\sqrt{n}$-consistent but is inefficient. This means that some efficiency will be lost if a generalized semiparametric GARCH class model is fitted to some stationary GARCH class process. In the next section, a simple stationary test is proposed based on the selected bandwidth. If this test is significant, the proposed semiparametric model will be used. Otherwise, stationary generalized GARCH class models should be employed.

[2] In the chapter, for fitting GARCH models, the R packages *fGarch* and *rugarch* are applied and the data-driven algorithm to be proposed in the next section is also carried out in R.

3.3.3 The bandwidth estimation algorithm

Numerous criteria for selecting the bandwidth in nonparametric regression are proposed. One bandwidth selection rule which works well in different contexts, is the iterative plug-in (IPI) idea (Gasser et al., 1991). This approach will also be used in the current section. Note that the estimation of $g(\tau_t, \lambda)$ is just the estimation of the scale function in $|r_t|^\lambda$.

The IPI algorithm is developed based on the formula of the asymptotically optimal bandwidth for estimating $\tilde{g}(\tau_t, \lambda)$, b_{A} say, which can be obtained by adapting those known results properly. In the sequel, this formula will be given without proof. Under regularity assumptions, in particular the assumption that $\gamma_{|\xi|^\lambda}(k)$ are absolutely summable, the asymptotically optimal bandwidth minimizing the dominating part of the MISE is given by

$$b_{\mathrm{A}} = \left(S_{|\xi|^\lambda} \frac{R(K)}{I^2(K)} \frac{\int [g(\tau,\lambda)]^2 d\tau}{\int \{[g(\tau,\lambda)]''\}^2 d\tau} \right)^{1/5} n^{-1/5}. \tag{3.15}$$

To select b, $S_{|\xi|^\lambda}$ has to be estimated. And the IPI idea is successfully applied to select bandwidth in different contexts (see e.g. Herrmann et al., 1992, Brockmann et al., 1993, Beran and Feng, 2002, and Ghosh and Draghicescu, 2002). An IPI bandwidth selector is calculated as Eq. (3.15). The IPI procedure is started with an initial bandwidth b_0. In Gasser et al. (1991), Herrmann et al. (1992) and Brockmann et al. (1993), the starting bandwidth $b_0 = n^{-1}$ is used. Beran and Feng (2002) proposed to use $b_0 = n^{-5/7}$ so that the starting bandwidth satisfies $b_0 \to 0$ and $nb_0 \to \infty$. The bandwidth $b_0 = 0.5n^{-1/5}$ is used by Feng (2004), which is of the optimal order $O(n^{-1/5})$. In this section, we proposed to select the starting bandwidth from a set of given bandwidths using the CV (cross-validation, Wahba and Wold, 1975) criterion, so that the algorithm is fully data-driven. It is well known that the choice of the starting bandwidth only has a clear effect on the required number of iterations but not on the finally selected bandwidth.

In an IPI algorithm, a bandwidth $b_{\lambda,j}$ for estimating the second derivative $[g(\tau,\lambda)]''$ is calculated from $\hat{b}_{j-1}$ using some inflation method. The choice of the inflation method is very important because the rate of convergence of an IPI

bandwidth selector depends on this choice. The original proposal of Gasser et al. (1991), applied a multiplicative inflation method (MIM), where $b_{\lambda,j} = \hat{b}_{j-1} \cdot n^{\alpha}$ with $\alpha = 1/10$. Now, we have $b_{\lambda,j} = O(n^{-1/10})$, once convergence is reached. This ensures that the variance of $\hat{b}/b_A$ has the fastest rate of convergence $O(n^{-1/2})$ but the bias of $\hat{b}$ is relatively large, where $\hat{b}$ denotes the finally selected bandwidth. An exponential inflation method (EIM), $b_{\lambda,j} = (\hat{b}_{j-1})^{\alpha}$, was proposed by Beran and Feng (2002). The authors proposed to use the optimal choice $\alpha = 5/7$, which minimizes the MSE of $\int [g(\tau, \lambda)]'' d\tau$. Numerical experiments show that sometimes the MIM method does not work well, because the inflation factor $n^{1/10}$ depends strongly on n, and the range of the sample size considered in the current context is very large. The EIM method with $\alpha = 5/7$ works well in different contexts and will be used.

Ghosh and Draghicescu (2002) proposed to estimate some unknown functions in bandwidth selection for quantile regression with time series errors directly from the data. Following their idea, it is proposed to estimate the unknown quantity $S_{|\xi|^{\lambda}}$ non-parametrically by the sum of the sample autocovariance $\hat{\gamma}_{|\xi|^{\lambda}}(k)$ of the residuals until some lag M, where M satisfies $M \to \infty$ and $M/n \to 0$. Bühlmann (1996) proposed the optimal window selection of Bartlett window and C^2-window with IPI. Bartlett window is selected as the lag window and in the following $M = [3n^{1/5}] = O(n^{1/5})$ will be used, where $[\cdot]$ denotes the integer part. Here, the cc_f is chosen as a constant 3 and the optimal c_f selection is neglected. Under this choice, the effect of the error in $\hat{S}_{|\xi|^{\lambda}}$ on the finally selected bandwidth is asymptotically negligible. Note that $\hat{\gamma}_{|\xi|^{\lambda}}(k)$ tends to zero very fast. Hence, the finally selected bandwidth will not be changed clearly, if a larger M, e.g. $M = [4n^{1/5}]$, is used. Also, note that bandwidths $\hat{b}_j$ obtained in several iterations at the beginning are usually inconsistent. It is not good to use those bandwidths to estimate $S_{|\xi|^{\lambda}}$. Following Herrmann et al. (1992), we select the bandwidth first by ignoring the correlation and scale change. In this stage a simple difference-based variance estimator $\hat{g}_{|\xi|^{\lambda}} = \frac{1}{2(n-1)} \sum_{t=2}^{n} (|\xi_t|^{\lambda} - |\xi_{t-1}|^{\lambda})^2$ will be used. The bandwidth selected at the end of this stage will be used as a new starting point for selecting the bandwidth under correlated errors with a smooth scale function. From now

on $\hat{S}_{|\xi|^\lambda}$ will be estimated and adapted in each iteration. The detailed bandwidth selection algorithm is discussed with the selection of the power transformation parameter in the next subsection.

3.3.4 The power transformation parameter estimation algorithm

Let b_0 denote the starting bandwidth, depending on an initial λ_0 value input. In the application, the starting input values $\lambda_0 = 2, 1, 0.5$ and 0.1 will be considered. In the algorithm, $\lambda_0 = 1$ is applied and the results remain with the other initial λ values.

The proposed data-driven algorithm is as follows.

1. Obtain $\hat{\mu}(\tau_t)$ using an IPI algorithm and let $\hat{r}_t^* = |r_t - \hat{\mu}(\tau_t)|$.

2. Select $\hat{b}_0$ from $b_{0,i} = c_{0,i} n^{-1/5}$ with $c_{0,i} = 0.05, 0.10, 0.15, 0.20, 0.25$ using the CV criterion and the starting power transformation parameter input $\lambda_0 = 1$. Then put $j = 1$.

3. Select a bandwidth by ignoring the correlation and scale change.

4. In the m-th iteration for $m \geq 1$:

   4a) Let $\Delta\lambda = 0.001$ be the interval of λ and $\lambda_n = n \cdot \Delta\lambda$, where $5 \leq n \leq 1000$.

   4b) Determine the $\hat{\lambda}_m = \lambda_n$, where λ_n is the power parameter maximizing the MLE or minimizing the JB statistic.

   4c) Increase m by one and repeat 4b) until $\hat{\lambda}_{m+1} \approx \hat{\lambda}_m$ and let $\hat{\lambda} = \hat{\lambda}_m$.

5. Let J_1 be the number of iterations in the last stage. In the j-th iteration with $j > J_1$:

   5a) Estimate $\hat{g}(\tau_t, \hat{\lambda})$ with b_{j-1}. Let $\hat{\xi}_t = \hat{r}_t^* / \hat{g}(\tau_t, \hat{\lambda})$ and $\hat{S}_{|\xi|^{\hat{\lambda}}} = \sum_{|k|<M} \hat{\gamma}_{|\xi|^{\hat{\lambda}}}(k)$.

5b) Let $[\hat{g}(\tau_t, \hat{\lambda})]''$ denote the estimate of $[g(\tau_t, \hat{\lambda})]''$ obtained by using $b_{\lambda,j} = b_{j-1}^{5/7}$.

5c) Improve b_{j-1} by

$$b_j = \left(\hat{S}_{|\xi|^{\hat{\lambda}}} \frac{R(K)}{I^2(K)} \frac{\int \hat{g}^2(\tau, \hat{\lambda}) d\tau}{\int \{[\hat{g}(\tau, \hat{\lambda})]''\}^2 d\tau} \right)^{1/5} n^{-1/5}. \tag{3.16}$$

5d) Increase j by one and repeatedly carry out 5b) and 5c) until convergence is reached or until a given maximal number of iterations has been done.

The finally selected power transformation parameter $\hat{\lambda}$ and bandwidth $\hat{b}_A$ are obtained in the last iteration of Step 4 and Step 5, respectively. In Step 1, the scale change is also ignored to save computing time. The condition $|b_j - b_{j-1}| < 1/n$ is used as a convergence criterion of $\hat{b}$ since such a difference is negligible. The bandwidth $\hat{b}_0$ used in Step 2 provides an object starting point of the algorithm. The maximal number of iterations, which indeed does not play any role in a common case, is 20 in Steps 1 and 3 and 30 in Step 5. The $\hat{\lambda}$, in Step 4, is a stable global power transformation parameter of the Box-Cox transformation. It means that now the scale function is estimated from the λ-th power of the absolute returns instead of the squared returns. Note that both the estimated scale function with the selected power transformation parameter λ and the scale function applied during the descaled process is $\hat{g}(\tau_t, \lambda)$. Obviously, the convert parameter $\hat{c}_{\lambda}^{-1/\lambda}$ in Equation (3.9) can not be neglected.

3.3.5 A simple stationary test

The proposed semiparametric models should be used, only if the underlying process is nonstationary in the mean and/or nonstationary in the variance. In the sequel, a simple method is proposed to test whether the variance of the process is constant. Similarly, a test of the stationarity in the mean can also be carried out, if this is of interest.

Assume that the acf (autocorrelation function) of $|\xi_t|^{\lambda}$ is absolutely summable for any $0 < \lambda \leq 2$. Our null-hypothesis (H_0) assumes that the process r_t is

stationary with constant standard deviation $s(\tau) \equiv s_0$ and finite fourth moments. Let $\hat{g}(\tau_t, \lambda)$ be the estimator of $g(\tau_t, \lambda)$ defined above with a bandwidth b such that $b \to 0$ and $nb \to \infty$ as $n \to \infty$. Under H_0 it is clear that $g(\tau_t, \lambda)$ is also a constant $g_0(\lambda) \equiv c_\lambda s_0^\lambda$. Under H_0 and corresponding regularity assumptions, we have

$$\sqrt{n\hat{b}_{\mathrm{A}}}[\hat{g}(\tau_t, \lambda) - g_0(\lambda)] \to N[0, R(K)V_{|\xi|^\lambda}], \tag{3.17}$$

where $R(K) = \int K^2(u)du$ is the kernel constant in the asymptotic variance of $\hat{g}(\tau_t, \lambda)$ and $V_{|\xi|^\lambda} = S^0_{|\xi|^\lambda} g_0^2(\lambda)$, where $S^0_{|\xi|^\lambda}$ is similar to $S_{|\xi|^\lambda}$ in Theorem 3.1 but defined for the process $\xi_t^0 = r_t^*/g_0(\lambda)$. The overall variance $g_0(\lambda)$ can be estimated by the sample variance of $\{\hat{r}_t^*\}$, $\hat{g}_0(\lambda)$ say. The quantity $S^0_{|\xi|^\lambda}$ can be estimated from $\hat{\xi}_t^0 = \hat{r}_t^*/\hat{g}_0(\lambda)$ following the idea in the last subsection. Let $\widehat{\mathrm{SD}}_\sigma = [\hat{V}_{|\xi|^\lambda} R(K)/(n\hat{b}_A)]^{1/2}$ and $Z_{\alpha/2}$ be the $\alpha/2$ upper quantile of the standard normal distribution for given confidence level α. Then $\hat{g}_0(\lambda) \pm Z_{\alpha/2}\widehat{\mathrm{SD}}_\sigma$ provide the approximate $(1-\alpha)*100\%$ confident bounds of $g_0(\lambda)$ under the stationary assumption on $\{r_t^*\}$. If more than $\alpha*100\%$ of the estimates $\hat{g}(\tau, \lambda)$ are clearly outside these confidence bounds, it indicates that $\{r_t^*\}$ is nonstationary in the variance and a semiparametric model should be used. Otherwise, generalized parametric GARCH models will be preferable.

3.4 Applications

Several major stock market indexes are selected to carry out the algorithm. In the following empirical research, Standard & Poor's 500 Index (S&P) and Deutscher Aktienindex 30 (DAX) from January 1996 to December 2015 are employed as the sample data to fit the general SemiGARCH models.

3.4.1 The estimation of λ

An IPI algorithm is carried out to calculate the Box-Cox power transformation parameter λ, calculated also as the power in the scale function of the general SemiGARCH models. In the algorithm, we set different initial λ inputs, which are

2, 1, 0.5 and 0.1, respectively. Due to the consideration of the absolute returns, the initial λ values are always positive. For the discussion on the negative and zero values, please refer to Feng et al. (2017). In the λ selection, we developed a six-step IPI algorithm and it is obvious that most of the examples quickly reach the $\hat{\lambda}$ after the second IPI procedure, which is also discovered by Herrmann and Gasser (1994), Beran and Feng (2002) and Feng (2013). To ensure the positivity of the input series, the absolute centralized returns should be considered, then the Box-Cox transformation can be carried out searching the $\hat{\lambda}$ in the IPI process. According to Lemma 2 in Feng (2013), we found also the λ monotonically increasing or decreasing in probability, depending on the starting input λ.

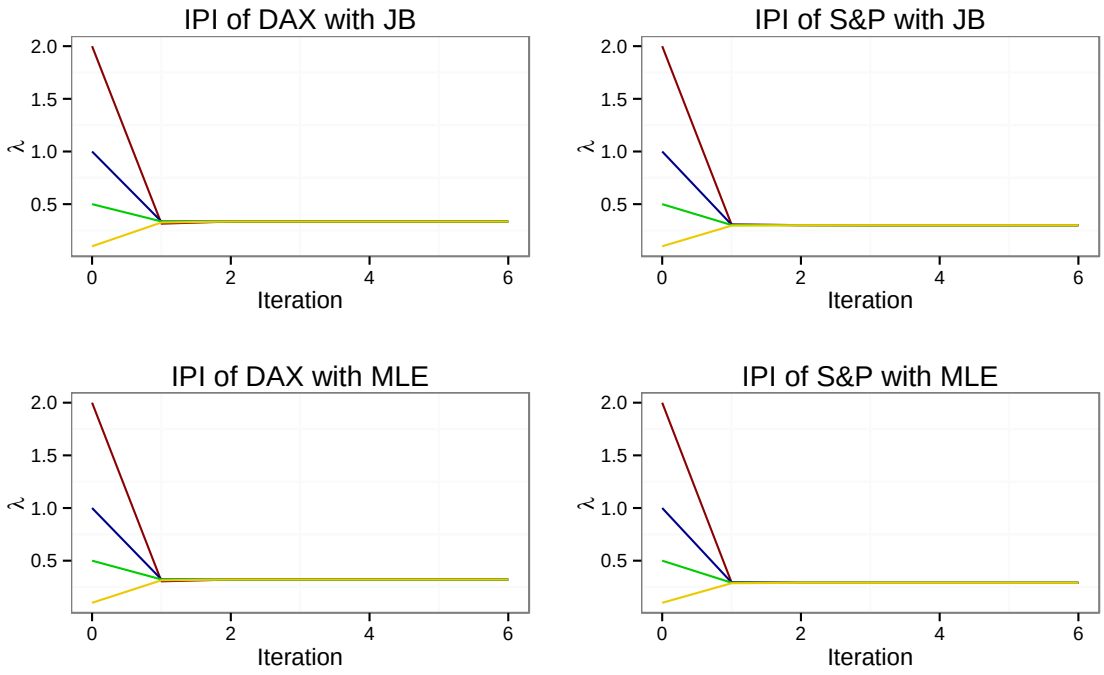

Figure 3.1: The IPI process with JB and MLE

As shown in Fig. 3.1, the λ values of both DAX and S&P with the JB as well as the MLE methods tend to the $\hat{\lambda}$ very fast, leading to coincide lines with different outsets, while the $\hat{\lambda}$ seems to be independent with the initial λ inputs.[3]

[3]The R packages *tseries* and *MASS* are used to select λ with JB and MLE.

Obviously, the $\hat{\lambda}$ for both of the MLE and JB methods can be reached in a few iteration procedures. There is no significant difference between the $\hat{\lambda}$ and the optimal λ calculated until the second IPI procedure, e.g. the optimal λ in the second IPI procedure of S&P with JB (starting $\lambda = 0.1$) is 0.3 and the obtained $\hat{\lambda}$ after all the iteration procedures is also 0.3. However, the result of λ may also reach the selected value in the first IPI step, e.g. for DAX with JB, when the initial $\lambda = 1$, the λ values are always 0.335 in the IPI process. Further, we can conclude, the convergence rate of the λ tending to the fixed $\hat{\lambda}$ depends on the initial inputs. If the starting input λ is far above or below the $\hat{\lambda}$, the convergence rate is dramatically greater than that with a relatively close distance from $\hat{\lambda}$, bringing to the power parameter decreases or increases quickly to the fixed $\hat{\lambda}$.

In the MEM model, the conditional variance of the returns is always considered to follow a squared power transformation, most likely expanding the positive skewness and affecting the least-square estimation quality. To overcome the drawback, a weaker moment power transformation of the absolute returns is considered. Taylor (1986), Ding et al. (1993) and Granger and Ding (1995) considered the absolute returns because the autocorrelation of the long-term dependent absolute financial returns is the maximized, which is recognized as the Taylor effect. Further, Ding and Granger (1996) also indicated that a fourth root transformation is preferred to the absolute returns but for the exchange rate. Noguchi et al. (2016) developed a quantile-matching technique to determine the power parameter λ by minimizing the distance between the lower and upper percentile from the sample median with an extensive range of the quantiles and with their criteria, a cube-root transformation seems to be optimal to model the absolute returns. In the section, we developed optimal λ searching methodologies by maximizing the likelihood estimation and minimizing the JB statistic. Please note that, in order to get rid of the possible influence of the time-varying trend, it is removed in a standardized process before carrying out the λ searching methodology.

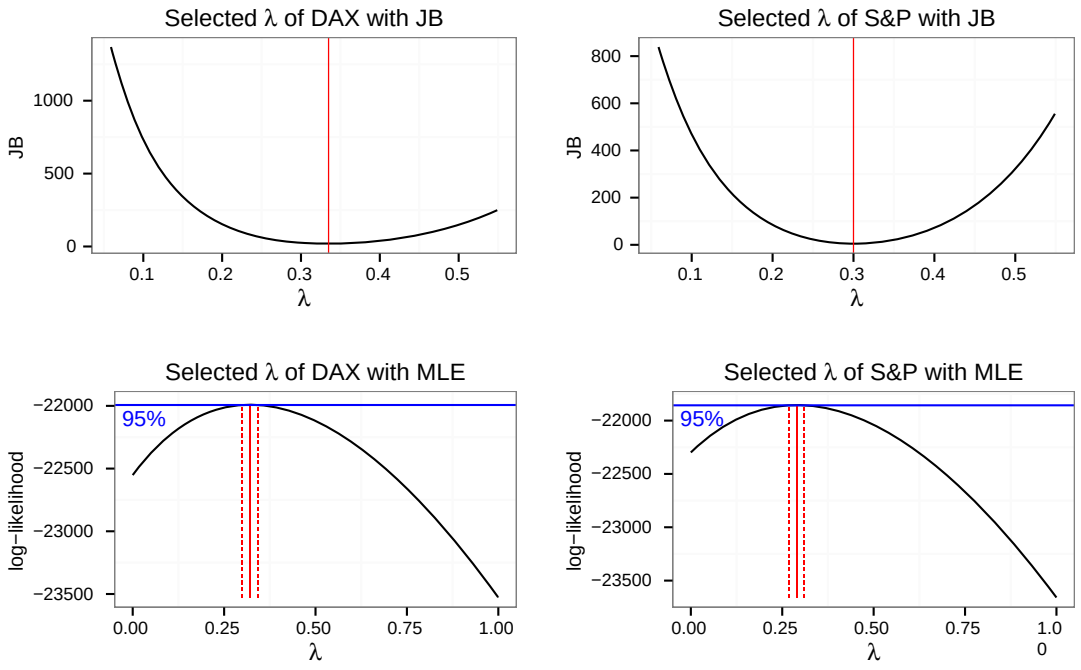

Figure 3.2: The $\hat{\lambda}$ with JB and MLE

In Fig. 3.2, the power parameters appear a U-shape curve. The $\hat{\lambda}$ is definitely where it minimizes the JB statistic or maximizes the MLE with a fixed bandwidth $\hat{b}$. The 95% confidence intervals of MLE can be calculated based on the χ^2 distribution, displayed as a vertical dashed line in the figure and the $\hat{\lambda}$ values are around one third, which are exactly 0.335, 0.300 and 0.321, 0.290 with JB and MLE of DAX and S&P, respectively. Obviously, the differences of the $\hat{\lambda}$ between the two methods are tiny. However, the $\hat{\lambda}$ values selected by both iterative algorithms are dramatically distinct from the ordinary ones, 1 (absolute returns) and 2 (squared returns). The histograms of the transformed examples with selected λ discussed above are displayed in Fig. 3.3.

In financial markets, a precise scale function with smaller λ is required to reveal the trend of the returns with various extreme observations, while, the consideration of smoothness of the scale function leads to the relatively larger λ values. Further, if smaller λ is applied, the requirement of higher-order moments of the returns does

not exist, leading to the scale function estimation under a possible weak moment condition, such as the existence of variance. In addition, at the left boundary, the JB value decreases at an extremely exponential speed, i.e. the JB statistic values are significantly increased if the approaching-zero λ values are considered. Following the definition of the Box-Cox transformation, there is no doubt that a logarithmic transform has to be considered if the λ is tending to 0. In the interesting case, the descaled series follows a logarithmic process and the model is also additive rather than multiplicative, i.e. the additive error model is an alternative, if a close zero λ is detected. Further, if the robustness is considered, the squared returns are still the optimal choice, although the normality statistics and MLE are disappointing.

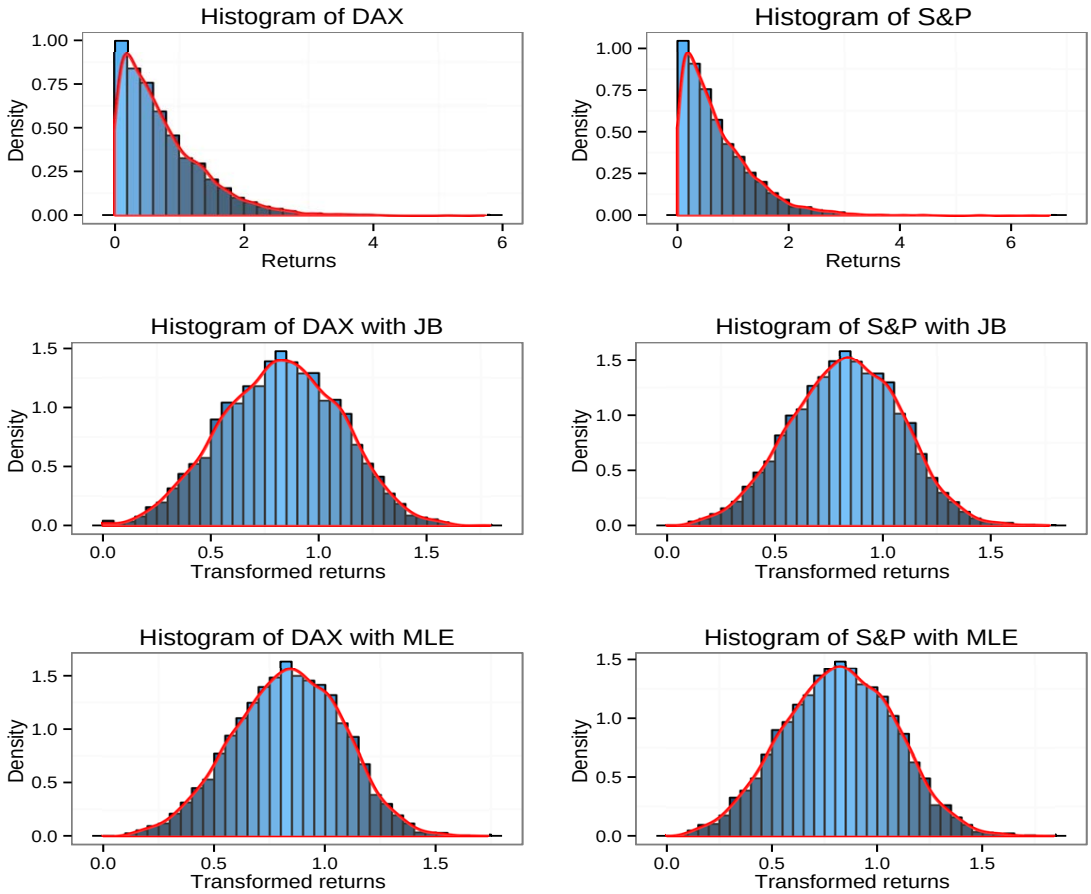

Figure 3.3: The histogram of DAX and S&P with JB and MLE

3.4.2 The selection of the parametric models

In the section, the power transformed absolute returns do not follow the squared ($\lambda = 2$) or the absolute ($\lambda = 1$) patterns by manual, but rather that of the λ-th power, which is selected by means of the iterative λ selection algorithm.

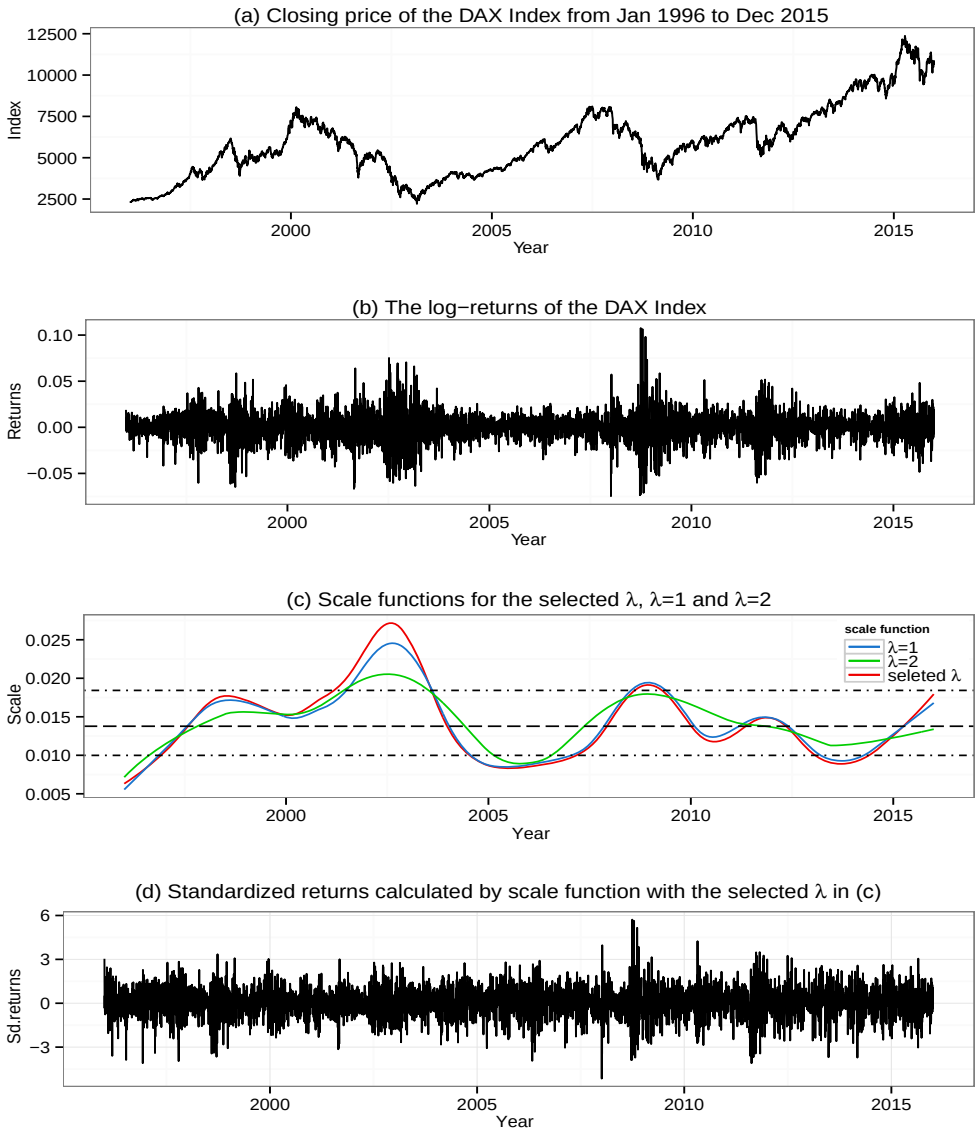

Figure 3.4: The smoothing results of DAX Index from Jan 1996 to Dec 2015

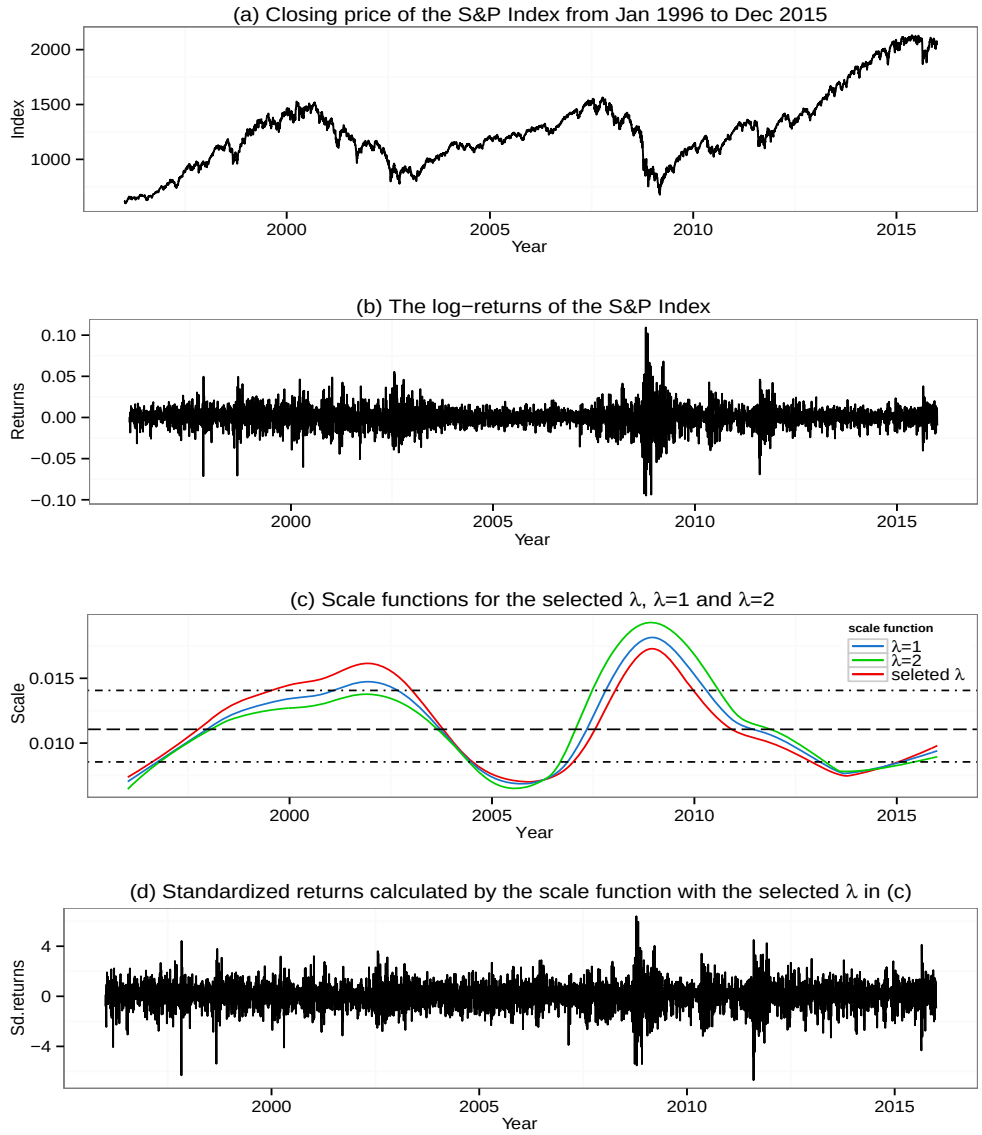

Figure 3.5: The smoothing results of S&P Index from Jan 1996 to Dec 2015

The smoothing results are displayed in Fig. 3.4 and Fig. 3.5. The $\hat{\lambda}$ values selected by the both MLE and JB are similar, so for simplicity, the $\hat{\lambda}$ via MLE is used in the parametric models fitting. Also, the stationary test in the variance is based on the $\hat{\lambda}$ by means of MLE. The returns seem to be more stationary after

removing the scale function, regarded as the long term component. In addition, clear GARCH cluster effects can still be observed, because the short term component displayed in a GARCH class process is barely affected by removing the long term component. In other words, the financial returns can be divided into the long and short components, which can be described by the scale function using Box-Cox transformation and the descaled process using the GARCH class process.

The GARCH, APARCH, EGARCH and CGARCH models of order (1,1), (1,2), (2,1) and (2,2) are chosen to analyze the conditional heteroskedasticity in the stationary standardized returns. It is also discovered that there is no significance with the mean function of the return series and it will not be considered in the model fitting. The innovations in the models are assumed to follow a normal- and t-distribution. In Table 3.1, the BIC of models with $\lambda = \hat{\lambda}, 1, 2$ are provided and it is discovered that the EGARCH(2, 1) models with t-distribution are always selected because of the minimum BIC values.

Table 3.1: BIC of the parametric models with $\lambda = \hat{\lambda}, 1$ and 2

Index	Model	Order	$\lambda = \hat{\lambda}$		$\lambda = 1$		$\lambda = 2$	
			Normal	t	Normal	t	Normal	t
DAX	GARCH	(1,1)	2.7472	2.7323	2.7483	2.7340	2.6981	2.6556
		(1,2)	2.7489	2.7340	2.7501	2.7358	2.7032	2.6596
		(2,1)	2.7452	2.7291	2.7461	2.7308	2.6960	2.6525
		(2,2)	2.7469	2.7308	2.7477	2.7324	2.6977	2.6542
	APARCH	(1,1)	2.7220	2.7091	2.7221	2.7098	2.6699	2.6362
		(1,2)	2.7236	2.7107	2.7238	2.7114	2.6749	2.6396
		(2,1)	2.7229	2.7103	2.7231	2.7109	2.6703	2.6355
		(2,2)	2.7246	2.7119	2.7248	2.7126	2.6720	2.6372
	EGARCH	(1,1)	2.7214	2.7082	2.7210	2.7087	2.6767	2.6351
		(1,2)	2.7229	2.7097	2.7225	2.7101	2.6803	2.6377
		(2,1)	2.7158	2.6993	2.7152	2.6998	2.6674	2.6238
		(2,2)	2.7168	2.7005	2.7162	2.7010	2.6686	2.6251
	CGARCH	(1,1)	2.7500	2.7352	2.7507	2.7368	2.6772	2.6578
		(1,2)	2.7516	2.7368	2.7523	2.7384	2.6778	2.6607
		(2,1)	2.7480	2.7321	2.7486	2.7336	2.6736	2.6545
		(2,2)	2.7497	2.7338	2.7502	2.7353	2.6753	2.6562
S&P	GARCH	(1,1)	2.7243	2.6988	2.7310	2.7060	2.7002	2.6753
		(1,2)	2.7259	2.7005	2.7326	2.7076	2.7018	2.6769
		(2,1)	2.7207	2.6956	2.7278	2.7030	2.6975	2.6726
		(2,2)	2.7208	2.6971	2.7278	2.7044	2.6976	2.6740
	APARCH	(1,1)	2.6745	2.6590	2.6797	2.6651	2.6481	2.6339
		(1,2)	2.6761	2.6606	2.6813	2.6667	2.6496	2.6355
		(2,1)	2.6769	2.6623	2.6824	2.6684	2.6510	2.6368
		(2,2)	2.6738	2.6605	2.6796	2.6669	2.6488	2.6362
	EGARCH	(1,1)	2.6725	2.6565	2.6776	2.6626	2.6472	2.6325
		(1,2)	2.6740	2.6581	2.6791	2.6642	2.6487	2.6341
		(2,1)	2.6651	2.6487	2.6708	2.6551	2.6404	2.6248
		(2,2)	2.6662	2.6499	2.6720	2.6563	2.6416	2.6261
	CGARCH	(1,1)	2.7275	2.7021	2.7341	2.7092	2.7032	2.6782
		(1,2)	2.7292	2.7038	2.7358	2.7109	2.7049	2.6798
		(2,1)	2.7237	2.6988	2.7308	2.7061	2.6997	2.6750
		(2,2)	2.7238	2.7003	2.7308	2.7076	2.7004	2.6758

From Table 3.2, the shape parameter, also known as the degree of freedom of the innovation distribution in all cases are significantly greater than 8, which means that the eighth moment of ξ_t exists and little heavy tails of the distribution of the innovations in the six research markets. Meanwhile, the degree of freedom of S&P is obviously lower than that of DAX, which means that the possibility of extreme returns in the US market is much higher than that in the German market. For another, in EGARCH models, γ_1 indicates no longer the leverage effect but the size effect of the past returns on volatility, which is typically a cluster effect. Besides, the leverage effect is denoted as α_1, being always negative to reveal the aggravation of past negative returns. In the cases of S&P and DAX, it is discovered that the leverage parameters are obviously determined by the negative sum of α_1 and α_2. It seems that the leverage effect in EGARCH models is weaker than that if other parametric models are applied, such as APARCH models.

Table 3.2: Fitting results of the SemiEGARCH(2, 1)-t models with selected λ

Data	Stat.	μ	ω	α_1	α_2	β_1	γ_1	γ_2	shape
	coeff.	0.0378	-0.0083	-0.2189	0.0991	0.9454	-0.0865	0.2515	10.4867
DAX	s.e.	0.0122	0.0033	0.0225	0.0233	0.0081	0.0338	0.0344	1.3717
	t	3.1076	-2.5237	-9.7398	4.2581	116.3966	-2.5593	7.3100	7.6448
	coeff.	0.0286	-0.0111	-0.2567	0.0858	0.9461	-0.1336	0.2591	9.6926
S&P	s.e.	0.0116	0.0034	0.0234	0.0243	0.0068	0.0346	0.0356	1.2381
	t	2.4659	-3.3061	-10.9928	3.5363	139.9253	-3.8667	7.2810	7.8288

3.5 Final remarks

We put forward a wide class of SemiGARCH models with Box-Cox transformation. A data-driven algorithm is also carried out in the transformation parameter selection, which is a great improvement in the scale function estimation of SemiGARCH models. The parameter λ we applied in the scale function estimation is obtained after several IPI procedures until it converges, also a supplement in

displaying the behavior of the long-term component in SemiGARCH models. In the parametric part, general GARCH models can be selected to describe the performance of the returns after removing the long-run trend. GARCH class models are discussed as the cluster models to show the short-run behaviors in some major financial markets of the world. It is found, if more extreme values are in a market, the transformation parameter λ tends to be smaller, for example, the $\hat{\lambda}$ of both DAX and S&P are only about a quarter, indicating the stability in the two stock markets. It is also proven from the distribution of innovations that the innovation of DAX follows a $\varepsilon_t \sim t(10.3433)$ distribution, exhibiting the existence of the eighth moment and little heavy tails.

The framework of general SemiGARCH models is set up, however, some open questions still have to be discussed further. e.g. the statistical properties of $\hat{\lambda}$ have not been fully explored yet. The optimal selection of the constant value at the zero point of spectral density in the IPI procedures is also of great interest.

Chapter 4

Value at Risk and Expected Shortfall under general Semiparametric GARCH models[1]

Risk management has been emphasized by financial institutions and the Basel Committee on Banking Supervision (BCBS). The core issue in risk management is the measurement of the risks. Value at Risk (VaR) and Expected Shortfall (ES) are the widely used tools in quantitative risk management. Due to the ineptitude of VaR on tail risk performances, ES is recommended as the financial risk management metrics by BCBS. In this section, we generate general SemiGARCH class models with a time-varying scale function. GARCH class models, based on the conditional t-distribution, are parametric extensions. Besides, backtesting with the semiparametric approach is also discussed. Following Basel III, the traffic light tests are applied in the model validation. Finally, we propose the loss functions with the views from regulators and firms, combing a power transformation in the model selection and it is shown that semiparametric models are a necessary option in practical financial risk management.

[1] Chapter 4 is based on the working paper: *Value at Risk and Expected Shortfall under General Semiparametric GARCH models* (Zhang, 2019b), CIE, 2019–06.

4.1 Introduction

Value at risk is the most popular metrics for financial risk management since the late 1980s. The BCBS first introduced VaR as the basic tool of risk measure and capital requirement in the supervisory framework of the BCBS (1996). However, VaR possesses some theoretical shortcomings. Artzner et al. (1999) pointed out that VaR is not subadditive and it does not reveal the risk well if extreme loss tail behaviors happen. He suggested considering the loss under the Expected shortfall level, which is proved to be sub-additive but not elicitable. Due to a more sensitive tail loss measure, the regulator recommended a risk metric shift from a 10-day 99% VaR to 97.5% ES (BCBS, 2012) and the incoming modified standard will be applied soon.

Backtesting is introduced as a method that applies historical data to predict the (out-sample) losses from actual realized (in-sample) losses within a fixed time interval, such as 250 days required by BCBS. Obviously, backtesting helps to detect the relationship between the expected VaR/ES and estimated losses. The VaR backtesting standards are explicit, such as the traffic light test (BCBS, 2012, 2016), the Kupiec's POF (proportion of failures) test (Kupiec, 1995), the TUFF test (Time until first failure, Kupiec, 1995), the Christoffersen test (1998), the joint test of coverage and independence (Haas, 2001) and so on. Although ES is about to be carried out in the very near future, the related backtesting rules are yet not found by BCBS. Recently, Gneiting (2012) has argued that, due to the elicitability, the direct backtesting method of ES can not be achieved. Meanwhile, Christoffersen (2003) has pointed out that, due to the ES conditional elicitability, it is feasible to evaluate the forecast and allow for tests, but not feasible for direct comparison and ranking the performance of prediction methods. Acerbi and Szekely (2014) suggested that the elicitability does not affect the risk model backtesting but just its comparison, leading to a failure in the evaluation function and a direct backtesting algorithm is also studied. McNeil et al. (2015) proposed a one-sample t-test to do the ES backtesting, checking if the mean of the excess loss is zero. Besides, Costanzino and Curran (2018) put forward a traffic light test for ES, which is similar to that of VaR proposed by BCBS, introducing breach

values to determine the colour zones by cumulative probabilities. On the other hand, to evaluate the risk measure forecast, Lopez (1999), Sarma et al. (2003), Angelidis and Degiannakis (2007) and Abad et al. (2015) discussed a two-stage evaluation approach, i.e. first test the violations of the risk models and then rank the models by the calculated statistics. In the rank process, we have introduced some loss function, defined from the view of different agents, such as the regulator and the firms.

In this chapter, we consider risk management in real markets with a semiparametric process. A time-varying scale function is introduced to decompose the long term risk component. After descaled the long term component, we discuss the VaR and ES with a stationary GARCH class process, which is defined as a model free class by using any GARCH type model. Besides, in the parametric process, we imply a power transformation of the returns, reducing the moment requirement of the GARCH models. Further, a two-stage method is carried out in the model evaluation, checking firstly the violations of the models by coverage and independence tests and then ranking the risk models by different loss functions with different power parameters. Following the requirements of Basel III, in this chapter, the confidence level for VaR and ES are 99% and 97.5% respectively and the forecast out-sample range is 250 days. Due to the robustness, the traffic light tests for both VaR and ES are also posed and in the ES traffic light test, breach values are applied as statistics to determine the zone's colour.

The chapter is organized as follows. Section 4.2 proposes the model. The statistical tests and loss function are the topics of Section 4.3. In Section 4.4, we discuss empirical implications with the two-stage method. Finally, Section 4.5 concludes.

4.2 VaR and ES with semiparametric processes

VaR and ES can be calculated based on the marginal distribution and the conditional distribution. In this section, we consider the VaR and ES in a semipara-

metric process, using a localized conditional distribution.

4.2.1 VaR and ES

VaR is the most important risk measurement tool based on the loss distribution and it is widely used in the financial institutions. In Basel III, VaR is a standard tool to measure the market risk. Generally, VaR is considered to be the maximum expected loss of a portfolio over a given time interval with a certain confidence level. For a given α, VaR_α is the up-α-quantile of the loss distribution,

$$VaR_\alpha = \inf\{l \in R : P(L > l) \leq 1 - \alpha\}, \tag{4.1}$$

where $\alpha \in (0, 1)$ is the given confidence level and L stands for the loss, defined as the negative returns $L_t = -r_t$. From Eq. (4.1), VaR at the confidence level α is given by the smallest number of l to ensure that the probability of the loss L exceeds l is not greater than $1 - \alpha$. Eq. (4.1) is also the non-parametric approach based on the marginal distribution and in this approach, no distribution assumption is required. The value of VaR depends strongly on the distribution and it is a constant for some marginal distribution, which does not depend on t. Thus, the VaR is a quantile of the loss distribution function.

If the loss L follows a t-distribution $t(\nu, \mu, \Sigma^2)$, suppose that $(L - \mu)/\Sigma$ is a standard t-distribution with ν degrees of freedom, VaR can be calculated as:

$$VaR_\alpha = \mu + \sigma t_\nu^{-1}(\alpha), \tag{4.2}$$

where $t_\nu(\cdot)$ and $t_\nu^{-1}(\cdot)$ are the density function and the quantile function of standard t-distribution, respectively. Please note that σ is not the standard deviation in the loss distribution, but that multiplicated by a constant related to the degree of freedom. Besides the loss distribution, the choices of time horizon and confidence level are also important for VaR calculation. The time horizon depends on the liquidity of the portfolios and the frequency they are traded. Less liquid means a longer time horizon and more liquid means a shorter horizon. Generally, VaR

in a short time horizon is less than the VaR in a long time horizon, due to the potential unexpected risk in the future.

Although VaR is a widely used risk management tool, it is also criticized by different views, especially its non-subadditivity. Artzner et al. (1999) showed that VaR is not a coherent risk measure due to the lack of sub-additivity and the poor tail risk capture, causing the risk of the merged portfolios is less than the sum of risk of individual portfolios. ES, on the other hand, is employed to overcome the shortcomings and the risk measure is changed from the VaR currently in use to a new metric, the so-called ES (BCBS, 2012).

For the loss L, the ES at the confidence level $\alpha \in (0,1)$ can be defined as:

$$\begin{aligned} ES_\alpha(L) &= E(L|L \geq VaR_\alpha(L)) \\ &= \frac{1}{1-\alpha}\int_\alpha^1 VaR_\gamma(L)d\gamma, \end{aligned} \tag{4.3}$$

where $VaR_\gamma(L)$ is the quantile function of the loss distribution. It is obviously that the ES can be assumed to be the expected loss given that L exceeds $VaR_\alpha(L)$. Thus, the expected shortfall indicates not only the information about frequency but also the size of large losses.

Similar to VaR, we consider the loss distribution for $\nu > 2$ and in this chapter the t-distribution is standard, so the variance of the distribution Σ is obviously 1. Then, ES with a parametric process can be expressed as (McNeil et. al, 2015),

$$ES_\alpha = \mu + \sigma ES_\alpha(L), \tag{4.4}$$

$$ES_\alpha(L) = \frac{g_\nu(t_\nu^{-1}(\alpha))}{1-\alpha}\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu - 1}. \tag{4.5}$$

Fig. 4.1 is the simulated example of VaR and ES with the t marginal loss distribution at the 95% confidence level. In general, the area below x axis is negative and indicates the profit whereas the above part is positive and indicates the loss. Assume that the loss value at α level is l_α, it is obviously to obtain $P(L > l_\alpha) = 1 - \alpha$, so the VaR value is l_α and it indicates that the possibility of

the maximum possible loss exceeding the l_α during the considered period is not greater than α. Besides, it is also found that ES is a VaR integration for level in $(1-\alpha, 1)$, which is reorganized as the conditional VaR (CVaR). It is obvious that, under the same confidence level, ES is always higher than VaR, meaning that ES a more sensitive metric than VaR in risk management.

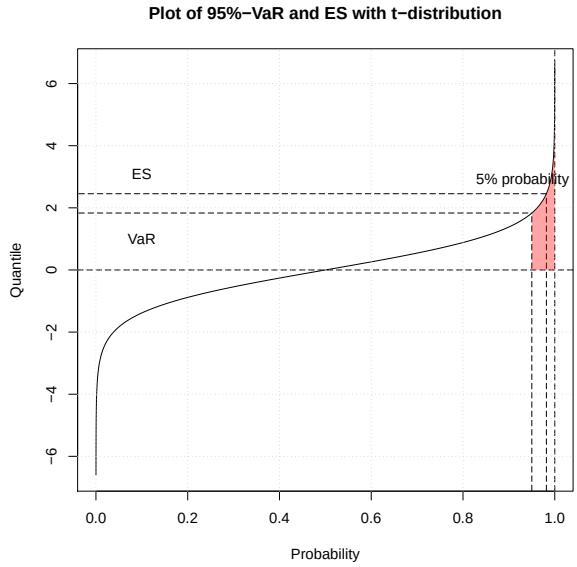

Figure 4.1: Plot of 95%-VaR and ES with t-distribution

4.2.2 The semiparametric models

To model the market risk with long and short risk decompositions, a semiparametric model with the time-varying scale should be considered. Let r_t, $t = 1, ..., n$, denote the logarithmic returns from an asset. In the following we propose to analyze r_t using a general SemiMEM (semiparametric multiplicative error model) defined by introducing a smooth scale function into the MEM proposed by Engle (2002).

$$r_t = s(\tau_t)\sqrt{h_t}\varepsilon_t, \tag{4.6}$$

where $s(\tau) > 0$ is a time-varying smooth scale function, $\tau_t = t/n$ is the rescaled time, h_t is the conditional variance of the re-scaled process $\xi_t = r_t/s(\tau_t) = \sqrt{h_t}\varepsilon_t$ and $\varepsilon_t \sim t(n)$ are standardized margins (zero mean and unit variance) i.i.d. ran-

dom variables. Generally, there is a nonparametric drift function in the returns series, however, it is not discussed in the section for simplicity, because the mean function varies around zero and it does not affect the following model estimation and the risk measures calculation.

The stationary process ξ_t can be analyzed using any suitable GARCH-type model. The parametric model has no affection to the trend estimation, so the semiparametric model is indeed a parametric-free model. In this section, we will consider modeling σ_t by the parametric approaches, such as the GARCH, the APARCH and the EGARCH models.

Besides, we consider the estimation of the time-varying scale change based on the power transformation r_t^λ for $\lambda \in [0, 1]$ here. Then, the power transform SemiGARCH models are expressed as

$$\text{sgn}(r_t) \cdot |r_t| = s_\lambda(\tau_t) r_{t,\lambda}, \tag{4.7}$$

where $\text{sgn}(r_t)$ is the sign of the returns, $s_\lambda(\tau_t)$ is the power transformed scale function and $r_{t,\lambda}$ is stationary with $E(r_{t,\lambda}^\lambda) = 1$. Obviously, the stationary process $r_{t,\lambda}$ is the product of the returns and a power transformed constant as $r_{t,\lambda} = C_\lambda^{-1/\lambda} r_t$. However, the original scale function can not be directly estimated and an equivalent scale function has to be applied in the trend estimation. The equivalent scale function $\tilde{s}(\tau_t)$ reads as

$$\tilde{s}(\tau_t) = s_\lambda(\tau_t) = C_\lambda^{1/\lambda} \cdot s(\tau_t), \tag{4.8}$$

The value of C_λ is determined by λ and the marginal distribution of r_t. Because $E(r_t) = 1$, we know that for the original series without transformation, $C_1 \equiv 1$, meaning that the first order $m(\cdot)$ is always the scale function of the commonly proposed estimator based on r_t. Considering the λ_0-th moments of the process and $\lambda_0 \in (0, 1]$, suppose that $E(r_t^{\lambda_0}) = 1$ and $E(\xi_t^{\lambda_0}) = 1$, we can conclude that $m_{\lambda_0}(\tau_t)$ is the scale function of $r_t^{\lambda_0}$, indicating that for the power transformed series, the constant $C_{\lambda_0} \equiv 1$. If the power $\lambda = 0$, the SemiGARCH model is transformed to an additive model with a logarithmic form. The details on the power transform,

please refer to Zhang (2019a).

We apply the local linear method to estimate the scale function estimator $\hat{s}_\lambda(\tau_t) = |\tilde{s}_\lambda(\tau_t)|$. $\tilde{s}_\lambda(\tau_t) = \hat{a}_0(x)$ is a local linear estimate obtained by minimizing

$$Q(a_0, a_1) = \sum_{t=1}^{T} \left\{ r_t^\lambda - a_0(\tau) - a_1(\tau)(\tau_t - \tau) \right\}^2 K\left(\frac{\tau_t - \tau}{b} \right), \tag{4.9}$$

where K is a symmetric kernel function. Feng (2019) propose a kernel regression method to prediction the risk measure. Besides, the simulation methods (Acerbi and Szekely, 2014) can also be considered here.

The semiparametric models can be also applied to the risk measurement of VaR and ES. A descaled process should be considered first by removing the estimated trend in the in-sample data. The out-sample conditional variances are calculated through the fitted unit GARCH models based on the in-sample descaled returns data. Suppose that the loss follows a t-distribution, for $k = 1, \ldots, K$, VaR and ES in a semiparametric process of the out-sample data should be considered as

$$VaR_\alpha^{t+k} = \mu(\tau_{t+k}) + \sigma(\tau_{t_0}) h_{t+k}^{1/2} t_\nu^{-1}(\alpha) \tag{4.10}$$

and

$$ES_\alpha^{t+k} = \mu(\tau_{t+k}) + \sigma(\tau_{t_0}) h_{t+k}^{1/2} ES_\alpha(L), \tag{4.11}$$

where $\mu(\cdot)$ is the local mean, $\sigma(\cdot)$ is the local variance and $ES_\alpha(L)$ is the quantile in (4.5). Obviously, the quantiles are depending on the innovations distribution. The local variance $\sigma(\cdot)$ is almost the same for t within a small period and is determined by observations within a time period around the observation point, which changes slowly over time and stands for middle term effect. In the out-sample risk prediction, we treat the local variance as a constant and it is approximately defined by the last local variance in in-sample series at $t = t_0$ as $\sigma(\cdot) = \sigma(\tau_{t_0})$. The conditional standard deviation in the out-sample should be $h_{t+k}^{1/2}$, estimated by unit GARCH models with descaled returns of the in-sample. For simplicity, the time-varying local drift function can also be treated as zero.

4.3 The Backtesting of VaR and ES

Backtesting is a necessary model validation method to check the model performance on the risk prediction with historical data. The model is accepted only if it can satisfy some statistical tests and predict robustly. In this section, a two-stage (Sarma et al., 2003) evaluation procedure is carried out. In the first stage, some tests, such as coverage test (Kupiec's POF test, Kupiec, 1995), independence test (Christoffersen, 1998), joint test (mixed Kupiec test, Haas, 2001) and traffic light test (BCBS, 2006, 2012, 2016) are put forward to test the statistical accuracy. Next, for the selected surviving models in the second, we rank their performance with the loss function (Lopez, 1998, Sarma et al., 2003, Caporin, 2008, Abad et al., 2015), such as the regulator loss function (RLF) and the firm loss function (FLF), respectively.

4.3.1 The backtesting of VAR

Different methodologies can be applied in the backtesting of VaR, such coverage tests, distribution tests and independence tests. In this section, we discuss the coverage test (the Kupiec's POF test), the independence test (Christoffersen's independence test) and a joint test, considering the coverage and independence together.

4.3.1.1 The Kupiec's POF test

The Kupiec POF (POF) test is based on the failure rate provided by Kupiec (1995). The aim of this test is to test the frequency of VaR exceeding over a given time interval. The null hypothesis of the POF test is

$$H_0 : p = \frac{x}{n}; \quad H_1 : p \neq \frac{x}{n}, \tag{4.12}$$

where x is the number of violation. The log-likelihood statistic LR_{UC} is in the form as

$$LR_{POF} = -2\ln\left[\frac{(1-p)^{n-x}p^x}{(1-\frac{x}{n})^{n-x}(\frac{x}{n})^{n1}}\right]. \tag{4.13}$$

Under the null hypothesis, LR_{POF} asymptotically follows χ^2-distribution as $LR_{POF} \sim \chi^2(1)$. However, there are two drawbacks to the POF test. As mentioned by Kupiec (1995), the test is not robust with relatively short out-sample interval, such as 250 days required by BCBS. For another, the test only considers the failure rate but neglects the time between failures.

4.3.1.2 The independence test

Christoffersen's independence test (Ind) (1998), which is the first test for independence of violations, is a likelihood ratio test that looks for aberrant frequent consecutive violations, i.e. this test examines if the probability of violations of the risk measures depends on the previous observation. He estimates the one-step-ahead transition probabilities $Pr(I_{t+1}|I_t)$ with a first-order Markov process,

$$\Pi_1 = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \\ \pi_{10} & \pi_{11} \end{bmatrix}, \tag{4.14}$$

where $\pi_{ij} = Pr(I_{t+1}|I_t)$. Let n_{ij} is the number of the observations with value i followed by j, then Matrix (4.14) can be estimated as

$$\hat{\Pi}_1 = \begin{bmatrix} \frac{n_{00}}{n_{00}+n_{01}} & \frac{n_{01}}{n_{00}+n_{01}} \\ \\ \frac{n_{10}}{n_{00}+n_{01}} & \frac{n_{11}}{n_{00}+n_{01}} \end{bmatrix}. \tag{4.15}$$

The null hypothesis of the independence test is

$$H_0 : \pi_0 = \pi_1; \quad H_1 : \pi_0 \neq \pi_1, \tag{4.16}$$

and the test statistic for independence of violations is defined as

$$LR_{Ind} = -2\ln\left[\frac{(1-\pi)^{n_{00}+n_{10}}(\pi)^{n_{01}+n_{11}}}{(1-\pi_0)^{n_{00}}\pi_0^{n_{01}}(1-\pi_1)^{n_{10}}\pi_1^{n_{11}}}\right] \sim \chi^2(1). \tag{4.17}$$

4.3.1.3 The joint test

The joint test (Mixed Kupiec test, Mix, Haas, 2001) combines the Kupiec's POF test and the Christoffersen's independence test, examining the coverage and independence together. As pointed out by Christoffersen (1998), if the first observation is conditioned on in the independence test, i.e. ignore the first observation, the statistics of the joint tests are exactly equal to the sum of those of the coverage test and the independence test. If the first observation is considered, then the approximate sign should be used.

$$LR_{Mix} = LR_{POF} + LR_{Ind} \sim \chi^2(2). \tag{4.18}$$

4.3.2 The backtesting of ES

The ES backtesting is now of great interest. Gneiting (2012) discussed that the ES is not available for backtesting due to the elicitability problem, which is proved not affected indeed. McNeil et al. (2015), Acerbi and Szekely (2014) and Costanzino and Curran (2018) provided different ES backtesting methods. In this section, the violation-based test and the traffic light test are mainly discussed.

4.3.2.1 The violation-based test

In McNeil et al. (2015), an indirect ES backtesting method is contributed with the consideration of the VaR violations. They consider ES can be expressed as the sum of VaR and an excess loss process as

$$ES_\alpha = VaR_\alpha + (ES_\alpha - VaR_\alpha), \tag{4.19}$$

where $ES_\alpha - VaR_\alpha$ is the excess loss. So, the ES backtesting can be separated into two individual components backtesting, one for the VaR component, the other for the excess loss component. The VaR backtesting methods discussed above are still feasible to the backtesting of the VaR component in Eq. (4.19). The null

hypothesis is that the excess loss, when VaR is violated, is i.i.d. and has an expectation zero, while, the excess loss has a mean greater than zero, leading to an underestimation of the conditional shortfall, as an alternative hypothesis. Please note that the test applied here is a one-sided t-test.

Suppose that the VaR component has passed the backtesting, the excess loss component can be tested by

$$K_t = \left(\frac{L_t - ES_\alpha^t}{ES_\alpha^t} \right) I_{\{L_t > VaR_\alpha^t\}}. \tag{4.20}$$

For simplicity, we ignore the expectation of the loss. Obviously, if there is no exceed VaR violation, the violation residual is definitely zero.

4.3.2.2 The traffic light test

In Basel III, the VaR backtesting must be based on a VaR measure calibrated at a 99th percentile confidence level and a prediction interval based on a sample of 250 observations. The BCBS provided a methodology for backtesting proprietary VaR measures and in this methodology, based on the number of violations in the out-sample data, VaR is categorized as one of three colored zones: green, amber and red. For all sample sizes, the amber zone lower boundary is from the cumulative probability equals or exceeds 95% and the red zone starts at the point where the cumulative probability equals or exceeds 99.99%. Following the instructions, the boundaries of the backtesting zone at 99% and 97.5% confidence levels are as follow.

If a model validation falls into the green zone, there is little worried about the concerned model's accuracy in this range. From Table 4.1, we see that in the amber zone, the models produce more exceptions, indicating that there is a higher probability for inaccurate models than for accurate models. Obviously, the model inaccurate grows by the increasing of the exception number.

Although the VaR traffic light coverage test is clearly illustrated by BCBS, a similar ES test is not yet discussed. Costanzino and Curran (2018) proposed an ES traffic light backtesting approach, which is analogous to the VaR traffic light

Table 4.1: Traffic light backtesting boundaries at 99% and 97.5% confidence levels

Zone	No. of violations		Cumulative prob.
	99%	97.50%	
Green	[0, 4]	[0, 10]	$< 95\%$
Amber	[5, 9]	[11, 16]	$< 99.99\%$
Red	10 or more	16 or more	$\geq 99.99\%$

backtesting approach proposed by BCBS, however, the probability information on the random loss cumulative distribution cannot be neglected.

The ES traffic light test introduces a new breach indicator, calculating the severity of the breach when the losses go beyond the related VaR confidence level. Different from the discrete exception numbers in the VaR traffic light test, the breach value in the ES traffic light backtesting is continuous.

Following the ES definition in Eq. (4.3), the ES generalized breach indicator ω_E for $\alpha \in (0, 1)$ is defined as

$$\begin{aligned} \omega_E^{(i)}(\alpha) &= \frac{1}{1-\alpha} \int_\alpha^1 \mathbb{I}_{\{L_i \geq VaR_i(p)\}} dp \\ &= \left(1 - \frac{1 - F_L(L_i)}{1-\alpha}\right) \mathbb{I}_{\{L_i \geq VaR_i(\alpha)\}} \\ &= \delta^{(i)}(\alpha) \cdot \omega_V^{(i)}(\alpha), \end{aligned} \tag{4.21}$$

where $F_L(L_i)$ is the cumulative distribution of the random loss L and $\omega_V^{(i)}(\alpha)$ is the breach indicator of VaR defined as

$$\begin{aligned} \omega_V^{(i)}(\alpha) &= \mathbb{I}_{\{L_i \geq VaR_i(\alpha)\}} \\ &= \begin{cases} 1, & \text{if } L_i \geq VaR_i(\alpha), \\ 0, & \text{if } L_i < VaR_i(\alpha). \end{cases} \end{aligned} \tag{4.22}$$

Compare the VaR and ES breach indicators, we can find that the ES breach indicator $\omega_E^{(i)}(\alpha)$ is indeed the VaR breach indicator $\omega_{VaR}^{(i)}(\alpha)$ multiplicated by an extra factor $\delta^{(i)}(\alpha)$ and the factor is not continuous, leading to the ES breach value is discrete. From the formula, if the loss L tends to VaR, leading to the cumulative distribution tends to α, then $\delta^{(i)}(\alpha) \to 1$. So, $\omega_E^{(i)}(\alpha)$ is close to zero. On the other hand, if $L \to +\infty$, the cumulative distribution $F_L(L_i) = 1$ and $\delta^{(i)}(\alpha) = 0$, so that the ES breach indicator $\omega_E^{(i)}(\alpha) = 1$. Here, we understand that if no VaR violation happens, the ES breach indicator is of course zero, while, if the violation is extremely large, the ES breach indicator will reach its maximum value 1. However, for the VaR breach indicator, no matter how large the violation is, the breach value is always 1, meaning that VaR is not as good as ES to reveal the extreme performance in the tail.

Then, the total ES breaches values for N transaction days is expressed as

$$\begin{aligned}\omega_E^N(\alpha) &= \sum_{i=1}^{N} \frac{1}{1-\alpha} \int_{\alpha}^{1} \mathbb{I}_{\{L_i \geq VaR_i(p)\}} dp \\ &= \sum_{i=1}^{N} \left(1 - \frac{1 - F_L(L_i)}{1-\alpha}\right) \mathbb{I}_{\{L_i \geq VaR_i(\alpha)\}} \\ &= \sum_{i=1}^{N} \delta^{(i)}(\alpha) \cdot \omega_V^{(i)}(\alpha). \end{aligned} \tag{4.23}$$

The boundary of ES traffic light test is fixed as

$$\sup_{x \in R_0^+} \{P(\omega_E^N(\alpha) \leq x) < q\}. \tag{4.24}$$

If $q < 0.95$, then x is the minimum green zone upper limit of the ES traffic light test. Similarly, the boundaries for the amber and red zone is taken into account with $0.95 \leq q < 0.9999$ and $q > 0.9999$, respectively.

The table below (Costanzino and Curran, 2018) is the ES traffic light test zone boundaries with $\alpha = 2.5\%$ and $N = 250$. It is showed in the table that the breach values of the 97.5% ES traffic light test seem to be more sensitive than those of VaR under the same confidence level, which are quite close to the VaR boundaries with $\alpha = 1\%$. It means that, in practice, the application of ES traffic light test

may be a strict and robust backtesting method than the others.

Table 4.2: The ES traffic light test boundaries under 97.5% confidence level

Zone	Breach value	Cumulative prob.
	0	0.18%
	1.3929	10%
	2.1131	25%
Green	3.0276	50%
	4.052	75%
	5.0622	90%
	5.7049	95%
	6.9844	99%
Amber	8.5285	99.90%
	9.8833	99.99%
Red	9.8833 more	$> 99.99\%$

4.4 The loss function

In the first stage backtesting, only the structure of the violation is detected, ignoring the amount of each exceedance. So, in the second stage, we are going to apply the loss function to analyze and rank the accurate of the risk prediction with different models and describe how well the models reveal the market risk by some numerical scores. In literature, two categories of the loss function are discussed, the RLF and the FLF. Lopez (1999) first proposed a general form loss function and in his RLF function, the loss, exceeding the VaR estimation, contributes to a more aggressive penalization. However, the difference between the loss and the estimated VaR is normally not so dramatic and it seems to exist an extreme magnitude gap (the constant one and the square of the difference as mentioned above). Then, the value one will be the dominant part in the RLF, meaning that it may

overstate some loss by minor exceedance. Sarma et al. (2003) defined a similar RLF with no constant factor left and the Sarma's RLF considers the exceedance contribution based on its real magnitude. Further, an FLF is also discussed in their paper and it is found that the FLF is exactly the same as the RLF when the loss exceeds the estimated risk measure, however, the opportunity cost of the reserved capital should be included, if no violation happens. Obviously, the cost here is the interest of the reserved capital held by the firms. Further, Feng (2019) proposed an FLF, considering the scenario that the minor violation happens. If the positive loss is no larger than the estimated risk measure, it should not be treated as the reserved capital. In this chapter, the RLF (Loss1) and FLF (Loss2) by Sarma et al. and the FLF (Loss3) by Feng et al. are applied in the empirical analysis. The loss functions for VaR and ES share the same formula but just the different risk measures. Let VaR be an example and the loss $L_t = -r_t$, then the Loss1 RLF is as

$$\text{Loss1} = \begin{cases} (L_t - VaR_t)^2, & \text{if } L_t > VaR_t, \\ 0, & \text{if } L_t \leq VaR_t. \end{cases} \tag{4.25}$$

The Loss 2 FLF reads as

$$\text{Loss2} = \begin{cases} (L_t - VaR_t)^2, & \text{if } L_t > VaR_t, \\ \beta(VaR_t - L_t), & \text{if } L_t \leq VaR_t, \end{cases} \tag{4.26}$$

where β is the daily interest rate.

Finally, the Loss3 FLF is as below

$$\text{Loss3} = \begin{cases} (L_t - VaR_t)^2, & \text{if } L_t > VaR_t, \\ \beta(VaR_t - L_t), & \text{if } 0 < L_t \leq VaR_t, \\ \beta VaR_t, & \text{if } L_t \leq 0. \end{cases} \tag{4.27}$$

Actually, the loss function represents the different view of market agents to backtesting. The regulator is willing to see the firms' capital for the trading book-

ing is linked to some risk measure, which is managed to be estimated, satisfying the regulatory requirements. For firms, the goal is to estimate and control the risk measures at a possible low level under the conditions that the backtesting results can be accepted by the regulator. Further, both sides are not concerned about the prediction accuracy, however, for the purpose of internal risk management, it would be an advantage for firms in seeking model improvement. The distinct views bring the different loss function criterion. The RLF, standing for the regulatory, always leads to an overestimated risk measure, making sure that the market risk is under control. While, for the firms, an underestimate-trend FLF is preferred, so as to avoid unnecessary large risk reserve capital and its opportunity cost. Therefore, to meet the requirements of both agents, the risk models selected should generally possess suitable exceedance number in backtesting, e.g. around the boundaries between the green and the amber zones.

4.5 The empirical study

In this section, some practical examples are discussed with the provided algorithm. We apply the DAX 30 index (DAX), the FTSE 100 index (FTSE), the Euro STOXX 50 index (EST), the Russell 2000 index (RUT), the S&P BSE SENSEX index (BSN) and the Brent Crude Oil Futures (BRO). The BSN is collected, ranging from July 1997 to September 2018 and the rest are from January 1988 to September 2018. For the models, we check the parametric (CS-), semiparametric with different selected power transformation parameters (LL1-, LL2- and LL3-) and log-transformed (LC-) models with the specific stationary process, such as GARCH (-OG), APARCH (-AP) and EGARCH (-EG).

In Table 4.3, the VaR backtesting results are listed. All the parametric models can pass the unconditional coverage test and the mixed test smoothly (p-value greater than 5%), however, for some semiparametric models, the p-values of the coverage and mixed are relatively lower, even some of the model cannot pass the tests, such as the LCAP and LCEG for FTSE, which the p-values are only 1.87%, far below the required 5%. If the models are able to pass the test, the

semiparametric models seem to perform better than the parametric models in the second stage of the backtesting. Obviously, for all the selected cases and the models (in bold in both tables below), displaying the smallest loss function values, are semiparametric models. Besides, we have to still pay attention to the peaks over threshold (POT, also applied by Peitz, 2015), showed in the figures for each example. Generally, the statistical test required a large sample size, in the backtesting however we do a backtesting based on only 250 days, which seems to be insufficient to support the statistical robustness. In this case, the test results by the BCBS traffic light test should be considered in advantage. If a model is capable to pass the backtesting, it should satisfy that the POT values are in the green zone of the traffic light test and all the statistical tests should be passed.

In the results table, we can understand that if a model can pass the traffic light test, then it passes also the statistical tests, but not vice versa. Further, from the view of different agents, the parametric models should be a regulator-prefer type. With parametric models, the estimated risk tends to be overestimated, leading to the minor POT value but the relatively larger loss function. The regulator is willing to accept the parametric models as the tools, however negative to the firms. In our study, the semiparametric models seem to be a trade-off result and satisfy the interest of both sides. For the regulator, the semiparametric models are able to pass all the related tests, meet the regulatory requirements and supervise the market risk, meanwhile, the estimated risk measures are not too large to be accepted by firms, so as to avoid the unnecessary cost of the risk capital. Specifically, firms are the modeler but regulators are the supervisor, meaning that the semiparametric models are sure to be practical and benefit of the firms.

Table 4.3: Coverage, independence tests and loss function values of 99% VaR

	Model		p-POF	P-mix	VPOT	Loss1	Loss2	Loss3	λ
DAX	CS	OG	0.2806	0.5564	1	0.0046	5.5739	4.8148	-
		AP	0.2806	0.5564	1	0.0114	5.8350	5.0760	-
		EG	0.2806	0.5564	1	0.0760	5.8293	5.0703	-
	LL1	OG	0.7530	0.9364	2	0.0198	-	-	0.79
		AP	0.7466	0.9453	1	0.0417	-	-	0.56
		EG	0.7466	0.9453	1	0.0927	-	-	0.47
	LL2	OG	0.7530	0.9175	3	-	5.2686	-	0.05
		AP	0.7466	0.9339	2	-	5.3671	-	0.05
		EG	0.7466	0.9339	2	-	5.3952	-	0.05
	LL3	OG	0.7530	0.9175	3	-	-	4.5096	0.05
		AP	0.7466	0.9339	2	-	-	4.6081	0.05
		EG	0.7466	0.9339	2	-	-	4.6362	0.05
	LC	OG	0.7530	0.9175	3	0.0824	5.2005	**4.4415**	0.00
		AP	0.3767	0.6443	4	0.1067	5.2698	4.5108	0.00
		EG	0.7530	0.9175	3	0.1438	5.2821	4.5231	0.00
FTSE	CS	OG	0.7530	0.9175	3	0.6102	5.0837	4.4607	-
		AP	0.7530	0.9175	3	0.1703	4.7139	4.0909	-
		EG	0.3767	0.6338	4	0.3004	4.7196	4.0966	-
	LL1	OG	0.3767	0.6338	4	0.7508	-	-	0.06
		AP	0.1597	0.3359	5	0.3152	-	-	0.06
		EG	0.0583	0.1437	6	0.4299	-	-	0.06
	LL2	OG	0.3767	0.6338	4	-	4.9806	-	0.06
		AP	0.1597	0.3359	5	-	4.5705	-	0.05
		EG	0.0583	0.1437	6	-	4.5952	-	0.06
	LL3	OG	0.3767	0.6338	4	-	-	4.3577	0.06
		AP	0.1597	0.3359	5	-	-	**3.9475**	0.05
		EG	0.0583	0.1437	6	-	-	3.9723	0.06
	LC	OG	0.0583	0.1437	6	1.0184	5.0075	4.3845	0.00
		AP	0.0187	0.0513	7	0.6090	4.5731	3.9501	0.00
		EG	0.0187	0.0513	7	0.7397	4.6137	3.9907	0.00

Table 4.3 to be continued

	Model		p-POF	P-mix	VPOT	Loss1	Loss2	Loss3	λ
EST	CS	OG	0.7466	0.9339	2	0.0837	5.0064	4.3405	-
		AP	0.2806	0.5564	1	0.0071	5.1902	4.5243	-
		EG	0.2806	0.5564	1	0.0037	5.1059	4.4399	-
	LL1	OG	0.7466	0.9339	2	0.0668	-	-	1.00
		AP	0.2806	0.5564	1	0.0045	-	-	1.00
		EG	0.2806	0.5564	1	0.0013	-	-	1.00
	LL2	OG	0.7466	0.9339	2	-	4.9336	-	0.93
		AP	0.2806	0.5564	1	-	5.0205	-	0.93
		EG	0.2806	0.5564	1	-	4.9796	-	0.93
	LL3	OG	0.7466	0.9339	2	-	-	4.2677	0.93
		AP	0.2806	0.5564	1	-	-	4.3545	0.93
		EG	0.2806	0.5564	1	-	-	4.3136	0.93
	LC	OG	0.7466	0.9339	2	0.1105	4.8997	**4.2337**	0.00
		AP	0.2806	0.5564	1	0.0337	4.9679	4.3020	0.00
		EG	0.2806	0.5564	1	0.0220	4.9288	4.2629	0.00
RUT	CS	OG	0.3767	0.0868	4	2.0614	7.3645	6.5774	-
		AP	0.3767	0.0868	4	1.5761	6.7511	5.9640	-
		EG	0.3767	0.0868	4	1.7798	6.9343	6.1472	-
	LL1	OG	0.1597	0.0477	4	1.6865	-	-	0.18
		AP	0.1597	0.0477	4	1.2207	-	-	0.21
		EG	0.1597	0.0769	5	1.2819	-	-	0.24
	LL2	OG	0.1597	0.0477	4	-	7.0337	-	0.21
		AP	0.1597	0.0477	4	-	6.3817	-	0.25
		EG	0.1597	0.0769	5	-	6.4244	-	0.25
	LL3	OG	0.1597	0.0477	4	-	-	6.2466	0.21
		AP	0.1597	0.0477	4	-	-	5.5946	0.25
		EG	0.1597	0.0769	5	-	-	5.6374	0.25
	LC	OG	0.3767	0.0868	4	1.5856	7.0082	6.2211	0.00
		AP	0.3767	0.0868	4	1.1205	6.3620	**5.5749**	0.00
		EG	0.3767	0.0868	4	1.1871	6.4060	5.6189	0.00

Table 4.3 to be continued

	Model		p-POF	P-mix	VPOT	Loss1	Loss2	Loss3	λ
BSN	CS	OG	0.2806	0.5564	1	0.4544	5.3788	4.7029	-
		AP	0.2806	0.5564	1	0.4899	5.5983	4.9225	-
		EG	0.2806	0.5564	1	0.6632	5.6264	4.9505	-
	LL1	OG	0.7466	0.9339	2	0.9348	-	-	0.05
		AP	0.7466	0.9339	2	1.1237	-	-	0.05
		EG	0.7466	0.9339	2	1.0621	-	-	0.05
	LL2	OG	0.7466	0.9339	2	-	5.1356	-	0.25
		AP	0.7466	0.9339	2	-	5.2781	-	0.85
		EG	0.7466	0.9339	2	-	5.2755	-	0.85
	LL3	OG	0.7466	0.9339	2	-	-	**4.4597**	0.25
		AP	0.7466	0.9339	2	-	-	4.6023	0.85
		EG	0.7466	0.9339	2	-	-	4.5997	0.85
	LC	OG	0.7466	0.9339	2	1.3396	5.1648	4.4890	0.00
		AP	0.3767	0.6338	4	1.7979	5.3967	4.7209	0.00
		EG	0.3767	0.6338	4	1.6566	5.2584	4.5826	0.00
BRO	CS	OG	0.2806	0.5564	1	8.8457	19.1354	17.5963	-
		AP	0.2806	0.5564	1	8.1441	18.3043	16.7651	-
		EG	0.2806	0.5564	1	8.0478	18.2725	16.7333	-
	LL1	OG	0.2806	0.5564	1	6.3465	-	-	0.10
		AP	0.2806	0.5564	1	6.3642	-	-	0.10
		EG	0.2806	0.5564	1	6.5822	-	-	0.10
	LL2	OG	0.2806	0.5564	1	-	17.9651	-	0.09
		AP	0.2806	0.5564	1	-	17.5176	-	0.07
		EG	0.2806	0.5564	1	-	17.5155	-	0.07
	LL3	OG	0.2806	0.5564	1	-	-	16.4259	0.09
		AP	0.2806	0.5564	1	-	-	15.9785	0.07
		EG	0.2806	0.5564	1	-	-	**15.9764**	0.07
	LC	OG	0.7466	0.9339	2	10.3922	20.1867	18.6476	0.00
		AP	0.7466	0.9339	2	9.7506	19.4522	17.9131	0.00
		EG	0.7466	0.9339	2	9.4725	19.2278	17.6887	0.00

The results of the ES backtesting tests are found in Table 4.4. For ES, not only the 97.5% ES backtesting is required, but also the VaR backtesting at the same significant level. Similarly, the unconditional coverage test, the mixed test, the excess loss t test are carried out. In most cases, the parametric model performance still satisfies the regulator. The relatively higher risk measure estimation brings stricter risk control in the market. However, please note that in the examples of EST, RUT and BRO, the POT of VaR (VPOT) and the POT of ES (EPOT) increase dramatically, therefore leading to some test failures, such as the CSEG of RUT. From the table, it is also indicated that the beach value increases following by the increase of the VPOT and the VaR traffic light test at 97.5% level seems to be too relax to guarantee the model staying in the green zone. In the VaR 97.5% traffic light test the green zone range is $[0, 10]$, however, if the VPOT value is around 10, the model is still simply able to pass the ES traffic light test, which indeed requires a reduce of the VaR backtesting green zone at this level, such as the cases of LCEG of BRO with VPOT value 10 and breach value 4.0003 and LL1OG of DAX with also VPOT 10 and breach value only 3.9375.

Finally, the power transformation and the semiparametric models are also necessary to estimate the risk measure and rank the loss function. It is found that the minimum loss function value always happens when the power parameter $\lambda \in [0, 1)$, e.g. 4.4963 with LCOG of BSN, 15.5369 with LL3EG of BRO ($\lambda = 0.07$), etc., which suggesting a consideration on the model class selection with multiplicative models or additive models. Further, it is interesting, the minimum loss values are all obtained with semiparametric models and the loss function values of the semiparametric models seem to be smaller than that of parametric models in most cases. In the study, we have examined the risk measures prediction and backtesting with the parametric models and the semiparametric models. Some positive cases are found and prove that the semiparametric models are able to characterize the market risk and provide the necessary regulatory information to both the regulators and firms. It is concluded that the semiparametric models are definitely well-performed risk management tools and treated as a supplement of the parametric models.

Table 4.4: Coverage, independence tests, breach and loss function values of 97.5% ES

	Model		p-POF	P-mix	VPOT	EPOT	Breach	t-test	Loss1	Loss2	Loss3	λ
DAX	CS	OG	0.2904	0.3155	9	1	3.1225	0.9994	0.0000	5.7104	4.9514	-
		AP	0.6068	0.8072	5	1	2.0462	0.9677	0.0054	5.9663	5.2073	-
		EG	0.7577	0.8019	7	1	2.1775	0.9607	0.0606	5.9519	5.1928	-
	LL1	OG	0.1584	0.0094	10	1	3.9375	0.9986	0.0053	-	-	0.93
		AP	0.2904	0.4535	8	1	3.1786	0.9668	0.0301	-	-	0.56
		EG	0.2904	0.0137	9	1	3.3195	0.9653	0.0760	-	-	0.47
	LL2	OG	0.1584	0.0094	10	3	4.6304	0.9920	-	5.3691	-	0.05
		AP	0.2904	0.0137	9	1	3.5898	0.9694	-	5.4787	-	0.05
		EG	0.2904	0.0137	9	1	3.7232	0.9422	-	5.5059	-	0.05
	LL3	OG	0.1584	0.0094	10	3	4.6304	0.9920	-	-	4.6101	0.05
		AP	0.2904	0.0137	9	1	3.5898	0.9694	-	-	4.7196	0.05
		EG	0.2904	0.0137	9	1	3.7232	0.9422	-	-	4.7469	0.05
	LC	OG	0.0161	0.0181	13	3	5.2456	0.9975	0.0396	5.2872	4.5282	0.00
		AP	0.2904	0.3155	9	3	4.1412	0.8616	0.0806	5.3644	**4.6054**	0.00
		EG	0.2904	0.3155	9	2	4.3050	0.8184	0.1204	5.3814	4.6224	0.00
FTSE	CS	OG	0.4898	0.3950	8	3	4.1544	0.5839	0.5394	5.1018	4.4788	-
		AP	0.7577	0.7787	7	3	3.9854	0.6452	0.1181	4.7519	4.1290	-
		EG	0.4898	0.6041	8	3	4.5022	0.4956	0.2430	4.7352	4.1123	-
	LL1	OG	0.1584	0.0097	9	4	5.1385	0.4017	0.6604	-	-	0.06
		AP	0.2904	0.4081	9	4	5.0830	0.5100	0.2420	-	-	0.06
		EG	0.2904	0.4081	9	6	5.4977	0.2674	0.3587	-	-	0.06
	LL2	OG	0.1584	0.0101	10	4	5.1921	0.5339	-	4.9734	-	0.05
		AP	0.2904	0.4081	9	5	5.3097	0.3934	-	4.5799	-	0.07
		EG	0.2904	0.4081	9	6	5.5368	0.2514	-	4.5901	-	0.05
	LL3	OG	0.1584	0.0101	10	4	5.1921	0.5339	-	-	4.3504	0.05
		AP	0.2904	0.4081	9	5	5.3097	0.3934	-	-	**3.9570**	0.07
		EG	0.2904	0.4081	9	6	5.5368	0.2514	-	-	3.9672	0.05
	LC	OG	0.0798	0.1706	11	6	6.6618	0.2357	0.9000	4.9675	4.3445	0.00
		AP	0.0798	0.1296	11	7	6.6091	0.2356	0.5029	4.5447	3.9217	0.00
		EG	0.0372	0.0622	12	7	6.9838	0.2250	0.6439	4.5785	3.9555	0.00

Table 4.4 to be continued

	Model		p-POF	P-mix	VPOT	EPOT	Breach	t-test	Loss1	Loss2	Loss3	λ
EST	CS	OG	0.0798	0.1359	11	2	4.1351	0.9987	0.0450	5.1151	4.4491	-
		AP	0.4898	0.6248	8	1	2.8359	0.9994	0.0011	5.3241	4.6582	-
		EG	0.2904	0.4239	9	1	3.1623	0.9994	0.0001	5.2416	4.5756	-
	LL1	OG	0.2904	0.3927	10	1	3.2805	0.9988	0.0213	-	-	1.00
		AP	0.4898	0.6248	8	1	2.5699	0.9997	0.0002	-	-	1.00
		EG	0.4898	0.6248	8	0	2.8332	0.9996	0.0000	-	-	1.00
	LL2	OG	0.2904	0.3605	11	2	4.3630	0.9967	-	5.0340	-	0.93
		AP	0.4898	0.6248	8	1	3.4716	0.9963	-	5.1424	-	0.93
		EG	0.4898	0.5840	9	1	3.6344	0.9988	-	5.1052	-	0.93
	LŁ3	OG	0.2904	0.3605	11	2	4.3630	0.9967	-	-	4.3680	0.93
		AP	0.4898	0.6248	8	1	3.4716	0.9963	-	-	4.4765	0.93
		EG	0.4898	0.5840	9	1	3.6344	0.9988	-	-	4.4393	0.93
	LC	OG	0.0798	0.1359	11	2	4.5054	0.9942	0.0634	4.9940	**4.3281**	0.00
		AP	0.4898	0.6248	8	1	3.6521	0.9929	0.0186	5.0850	4.4190	0.00
		EG	0.1584	0.2540	10	1	3.8275	0.9990	0.0099	5.0496	4.3837	0.00
RUT	CS	OG	0.1584	0.2599	10	4	5.1923	0.4714	1.8581	7.2858	6.4987	-
		AP	0.0798	0.1706	11	4	5.2040	0.6728	1.4131	6.6994	5.9123	-
		EG	0.0372	0.0990	12	4	5.6299	0.6944	1.5884	6.8640	6.0769	-
	LL1	OG	0.2904	0.0350	8	4	4.1838	0.4386	1.4887	-	-	0.18
		AP	0.2904	0.0350	8	4	4.1217	0.5260	1.0616	-	-	0.18
		EG	0.0372	0.0572	8	4	4.1881	0.5027	1.1204	-	-	0.21
	LL2	OG	0.2904	0.0350	8	4	4.1895	0.4376	-	6.9618	-	0.22
		AP	0.2904	0.0350	8	4	4.1621	0.5142	-	6.3356	-	0.27
		EG	0.0372	0.0113	9	4	4.2829	0.6138	-	6.3762	-	0.42
	LŁ3	OG	0.2904	0.0350	8	4	4.1895	0.4376	-	-	6.1747	0.22
		AP	0.2904	0.0350	8	4	4.1621	0.5142	-	-	5.5485	0.27
		EG	0.0372	0.0113	9	4	4.2829	0.6138	-	-	5.5891	0.42
	LC	OG	0.7577	0.0329	7	4	4.0124	0.3559	1.3987	6.9482	6.1611	0.00
		AP	0.9268	0.2965	6	4	3.9306	0.2477	0.9742	6.3298	**5.5427**	0.00
		EG	0.7577	0.3790	7	4	3.9695	0.4181	1.0368	6.3708	5.5837	0.00

Table 4.4 to be continued

Model			p-POF	P-mix	VPOT	EPOT	Breach	t-test	Loss1	Loss2	Loss3	λ
BSN	CS	OG	0.0460	0.1344	2	1	1.1984	0.3830	0.3936	5.4491	4.7733	-
		AP	0.0460	0.1344	2	1	1.2390	0.3660	0.4318	5.6668	4.9909	-
		EG	0.0460	0.1344	2	1	1.5182	0.2787	0.5995	5.6860	5.0102	-
	LL1	OG	0.3343	0.0108	4	2	2.3270	0.4213	0.8582	-	-	0.05
		AP	0.9268	0.8587	6	2	2.7894	0.4022	1.0503	-	-	0.05
		EG	0.6068	0.7907	5	2	2.5662	0.3358	0.9868	-	-	0.05
	LL2	OG	0.3343	0.0108	4	2	2.3968	0.4063	-	5.1643	-	0.25
		AP	0.9268	0.8587	6	2	2.8742	0.3796	-	5.2984	-	0.85
		EG	0.6068	0.7907	5	2	2.6744	0.3133	-	5.2967	-	0.92
	LL3	OG	0.3343	0.0108	4	2	2.3968	0.4063	-	-	4.4884	0.25
		AP	0.9268	0.8587	6	2	2.8742	0.4022	-	-	4.6226	0.85
		EG	0.6068	0.7907	5	2	2.6744	0.3133	-	-	4.6209	0.92
	LC	OG	0.7577	0.4266	7	2	3.8484	0.3918	1.2535	5.1722	**4.4963**	0.00
		AP	0.4898	0.4347	8	4	4.8706	0.2217	1.7069	5.3839	4.7080	0.00
		EG	0.7577	0.4266	7	4	4.6645	0.2006	1.5669	5.2478	4.5720	0.00
BRO	CS	OG	0.4898	0.6041	8	1	3.3241	0.6622	8.0018	18.6323	17.0931	-
		AP	0.4898	0.6041	8	1	3.2744	0.7093	7.2966	17.7998	16.2606	-
		EG	0.4898	0.6041	8	1	3.2126	0.7206	7.2059	17.7743	16.2351	-
	LL1	OG	0.1461	0.3352	3	1	1.3104	0.4517	5.5662	-	-	0.10
		AP	0.9268	0.8587	6	1	1.6908	0.7633	5.5667	-	-	0.09
		EG	0.9268	0.8587	6	1	2.0530	0.7203	5.7719	-	-	0.10
	LL2	OG	0.1461	0.3352	3	1	1.3120	0.4514	-	17.5684	-	0.09
		AP	0.9268	0.8587	6	1	1.7022	0.7622	-	17.0956	-	0.07
		EG	0.9268	0.8587	6	1	2.0594	0.7194	-	17.0761	-	0.07
	LL3	OG	0.1461	0.3352	3	1	1.3120	0.4514	-	-	16.0292	0.09
		AP	0.9268	0.8587	6	1	1.7022	0.7622	-	-	15.5564	0.07
		EG	0.9268	0.8587	6	1	2.0594	0.7194	-	-	**15.5369**	0.07
	LC	OG	0.4898	0.6041	8	1	4.0137	0.4967	9.5255	19.6396	18.1004	0.00
		AP	0.1584	0.2434	10	2	4.1432	0.7198	8.8662	18.8904	17.3513	0.00
		EG	0.1584	0.2434	10	1	4.0003	0.7487	8.5918	18.6755	17.1364	0.00

The VPOT and EPOT of DAX are plotted in Fig. 4.2 and please find the other plots in Appendix B. In the figures, VaR is estimated at the 95% confidence lever and ES is at 97.5%. The out-sample size, required by Basel III, is about one year (250 transaction days). Besides, even though the VPOT (blue point) and EPOT (green point) values are very small, the coverage test fails also by CSOG, CSAP and CSEG of BSN with a p-value only 4.6%, comparing the perfect performance with the semiparametric models. On the other hand, a violation indicator breach values for ES is introduced for the ES backtesting. Unlike those of VaR, a series of discrete numbers, the ES breach value is continuous, lying around half of the VPOT values. Following Costanzino and Curran's ES traffic light test (2018), we can see that the boundary between the green and amber zones is only 5.7049, almost half of the VaR breach indicator 10 at 97.5% level in theory, however close to the boundary at 99% level, which is 4. In other words, the ES breach value indicates a more sensitive risk measure than that of VaR at the same confidence level.

In the studied cases, we found that the ES breach indicator is also a strict test rule. Even passing all the other tests, the ES breach values still possibly locate in the amber zone, e.g. the breach values of the three LC models of FTSE are beyond the green zone limitation, while the t-test still indicates the models should be accepted with a reasonable statistic over 5%. Besides, the t-test seems to be not capable to reveal the quality of the risk measure estimation, especially when the risk measures are underestimated. In the LCOG case of DAX, the VPOT is as high as 13 and the breach value is 5.2456, almost reaching its green zone upper limitation, comparing the high t-test value as 0.9975. From the above cases, we can conclude that the t-test is not robust due to the limited observations in backtesting and the traffic light test, combing with the POF test and independence test, is reliable to both VaR and ES backtesting.

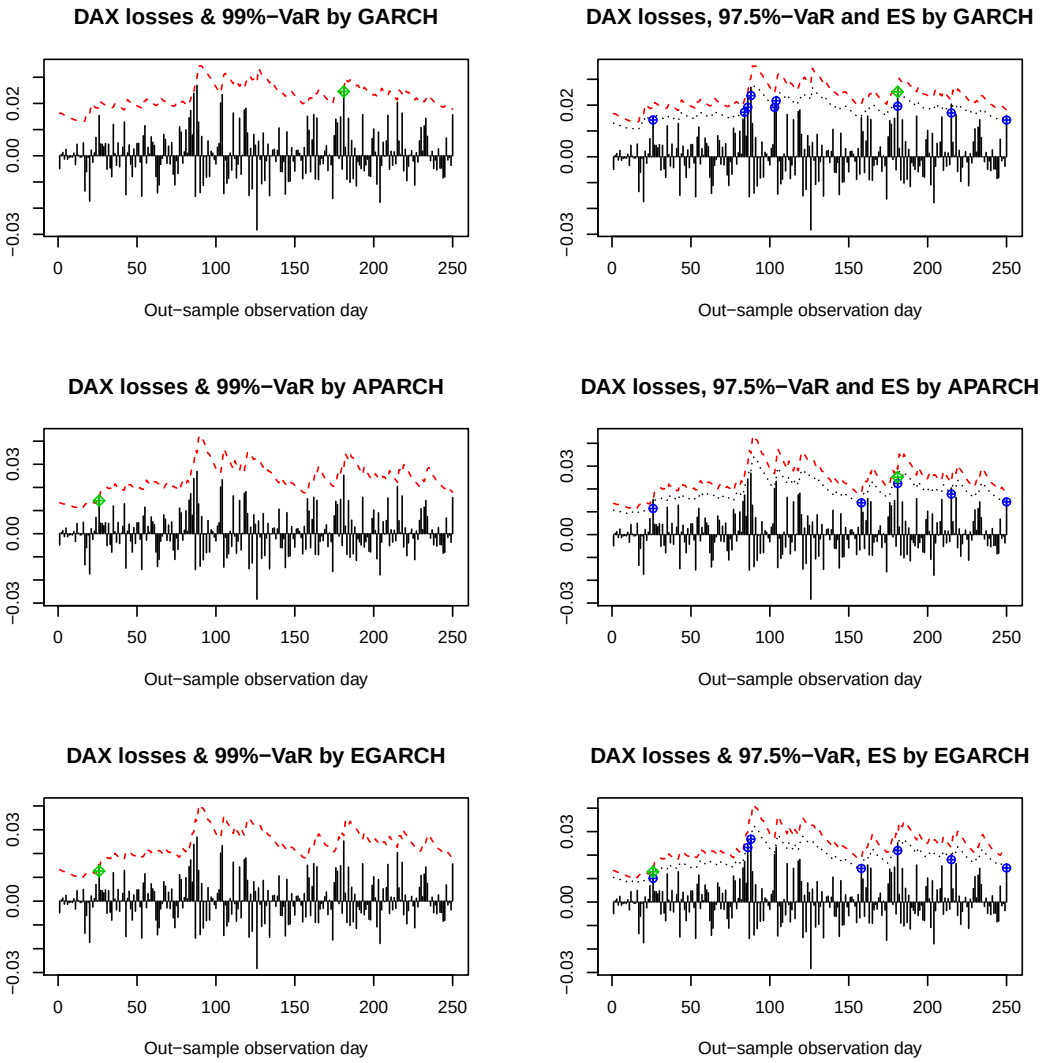

Figure 4.2: DAX POT of VaR and ES with parametric models

4.6 Final remarks

In Basel III and its coming finalization, ES is required to be a basic risk management tool, however, the backtesting of ES is not as clear as that of VaR. In the study, we have examined the risk measures prediction and backtesting with the parametric models and the semiparametric models. Besides, a simple traffic light test of ES is also discussed by introducing breach values. Some positive cases are found and prove that the semiparametric models are able to characterize the market risk and provide the necessary regulatory information to both the regulators and firms. It is concluded that the semiparametric models are definitely well-performed risk management tools and treated as a supplement of the parametric models.

Chapter 5

Modeling high-frequency returns at fixed trading time points using a general SemiGARCH model

---

The use of the GARCH model is widely observed in the empirical literature. However, this model may cause misclassification and assumes that the unconditional variance of the time series is constant. The recently proposed semiparametric GARCH model, which composes of the conditional heteroskedasticity and scale functions, can improve the GARCH model. In this section, the definitions, the features and the estimation of the GARCH model, the SemiGARCH model and their extensions are investigated. Based on the SemiGARCH model the SemiEGARCH and the SemiCGARCH models (Peitz, 2015) are introduced in this work. In the empirical example the SemiAPARCH, the SemiEGARCH and the SemiCGARCH models are applied to the returns of Allianz and BMW at fixed trading time points. It is found that the semiparametric models have a more correct theoretical basis. They can model the conditional heteroskedasticity and the scale change at the same time. Furthermore, the semiparametric models work well with the returns at fixed trading time points.

5.1 Introduction

Trading off risks against returns appears to be essential and vital for making a financial decision. Hence the econometric analysis of risk (volatility) becomes

an important part in forecasting market tendency and supporting making financial decisions, such as portfolio diversification, risk management and derivative pricing. In the last 20 years volatility was a research hot-spot in the financial industry. Volatility is regarded as a parameter for evaluating the risk of assets return. Generally, the stronger the volatility is, the higher the risk is.

In the classical financial models, the variance of the time series is always assumed as constant. However, it is found that the volatilities of financial time series have always the features of clustering and fat tails (Mandelbrot, 1963 and Fama, 1965). These features are not consistent with the assumption of constant, so the classical econometric methods cannot analyze the financial time series efficiently in practice. To overcome this problem, several economists have carried out studies on researching and developing frameworks for evaluating volatility. Since Engle introduced the autoregressive conditional heteroskedasticity (ARCH) model (Engle, 1982), the extensions of the ARCH model appeared and spread rapidly. Among the carried out researches, the Generalized ARCH (GARCH) model and its derivatives are most widely used (Bollerslev, 1986).

According to many studies (Gourieroux and Monfort, 1992 and Eubank, 1993), in parametric volatility models, the preselected model might be too restricted or too low-dimensional, which may not fit unexpected features and cause the misspecification. However, in nonparametric models, the parameters of the model cannot be estimated and the model cannot be explained due to a lack of specific functions. Instead of parametric volatility models or nonparametric models, the recently proposed semiparametric volatility model will be introduced in detail in this section, which introduces a smooth scale function into the standard GARCH model. This model does not need a prespecified function and is less sensitive to model misspecification. At the same time, the model can be also explained (Di and Gangopadhyay, 2011).

The definition, estimation, some properties of the semiparametric model and the methods of bandwidth selection are discussed. Furthermore, based on the study of the semiparametric GARCH model and the semiparametric asymmetric power ARCH model, which are introduced by Feng (2004) and Feng and Sun

(2013), the semiparametric exponential GARCH and component GARCH models are originally defined. Then the discussed semiparametric models, i.e. the SemiAPARCH, the SemiEGARCH and the SemiCGARCH models, are applied to the returns of BMW and Allianz from January 2006 to September 2014. Different from other algorithms, to get the more exact analyzing results the fixed trading time points are used here.

The scope of this chapter is as follows. In section 5.2, the parametric volatility models are introduced. The semiparametric volatility models are described in section 5.3. Section 5.4 reports the application of the semiparametric volatility models to the returns of BMW and Allianz and the empirical results on the volatility of the selected data sets. Finally, this chapter is concluded in section 5.5.

5.2 The semiparametric volatility model

The SemiGARCH model is a general framework if the nonstationary trend is removed. In this section, the APARCH, EGARCH and CGARCH models are considered as the parametric part to analyze stationary processes.

5.2.1 The SemiGARCH model

The SemiGARCH model combines a smooth scale function with the standard GARCH model:

$$Y_t = \mu + s(\tau_t)\varepsilon_t, \tag{5.1}$$

where μ is an unknown constant, $\tau_t = t/n$, $s(\cdot) > 0$ is the nonparametric component, a smooth, bounded scale function and $\{\varepsilon_t\}$ is a parametric component. The conditional variance of $\{\varepsilon_t\}$ is assumed to follow a GARCH(p, q) process:

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{5.2}$$

where $h_t^{1/2}$ is the conditional standard deviations of the standardized process ε_t, $\omega > 0$, $\alpha_1, \ldots, \alpha_p \geq 0$ and $\beta_1, \ldots, \beta_q \geq 0$. To estimate the scale function, $E(\varepsilon_t^8) < \infty$ is assumed to ensure model 5.2 strictly stationary, which implies in particular that $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ (Feng, 2004).

The SemiGARCH model provides us a tool to decompose financial risk into an unconditional component $s(\tau_t)$, a conditional component $h_t^{1/2}$ and the i.i.d. innovations η_t.

The estimation of the SemiGARCH model can combine the nonparametric estimation of the local variance $\nu(\tau_t) = \sigma^2(\tau_t)$, with parametric estimation of the unknown parameter vectors $\theta = (\alpha_0, \alpha_1, \ldots, \alpha_p; \beta_1, \ldots, \beta_q)$.

At first, the scale function can be estimated by some nonparametric regression approaches without any parametric assumptions. In this model, the kernel estimation will be used. If the constant mean μ is replaced by a smooth function $\mu(\tau_t)$, we can get a nonparametric regression with a time-varying mean as

$$Y_t = \mu(\tau_t) + s(\tau_t)\varepsilon_t, \tag{5.3}$$

where ε_t is a zero mean stationary process.

Eq. (5.1) can be transformed into a general nonparametric regression problem. Letting $r_t = Y_t - \mu$, $Z_t = r_t^2$ and $\xi_t = \varepsilon_t^2 - 1$, which are zero mean stationary time series errors. Then Model 5.1 can be rewritten as

$$Z_t = g(\tau_t) + g(\tau_t)\xi_t. \tag{5.4}$$

Letting $\hat{\mu} = \bar{y}$ and $\hat{z}_t = \hat{r}_t^2$, in which $\hat{r}_t$ is then defined by $\hat{r}_t = y_t - \bar{y}$. A Nadaraya-Watson kernel regression is

$$\hat{v}(\tau) = \frac{\sum_{t=1}^{n} K(\frac{\tau_t - \tau}{b})\hat{z}_t}{\sum_{t=1}^{n} K(\frac{\tau_t - \tau}{b})} =: \sum_{t=1}^{n} w_t \hat{z}_t, \tag{5.5}$$

where w_t is the weighting function $w_t = \frac{K(\frac{\tau_t - \tau}{b})}{\sum_{t=1}^{n} K(\frac{\tau_t - \tau}{b})}$, $K(u)$ is a second order kernel function with a compact support [-1,1] and b is the bandwidth, the size of the weights (Fan, 1993).

According to the above assumptions, the estimator ε_t is now replaced by the standardized residuals

$$\hat{\varepsilon_t} = \hat{r_t}/\hat{s}(\tau_t) = (y_t - \bar{y})/\hat{s}(\tau_t). \tag{5.6}$$

Then the estimator of parametric vector θ can be obtained by the standard maximum likelihood method, which has been introduced in Chapter 2. A suitable model can also be selected by using other methods, e.g. the Akaike information criterion (AIC), the Bayesian information criterion (BIC), etc. (Feng, 2004).

To calculate the asymptotic optimal bandwidth, the assumptions in the appendix should be followed. Define $R(K) = \int K^2(u)du$ and $I(K) = \int u^2 K(u)du$. The asymptotic bias B of $\hat{g}(\tau)$ is

$$B[\hat{g}(\tau)] = E[\hat{g}(\tau) - g(\tau)] = \frac{I(K)g''(\tau)}{2} b^2 + o(b^2). \tag{5.7}$$

The asymptotic variance of $\hat{g}(\tau)$ can be expressed as

$$V[\hat{g}(\tau)] = var(\hat{g}(\tau)) = \frac{2\pi c_f g^2(\tau) R(K)}{nb} + o\left(\frac{1}{nb}\right). \tag{5.8}$$

where c_f is a constant factor in the asymptotic variance. From Eq. (5.7) and (5.8) we can see that the bias and the variance are asymptotically dominated by b^2 and $(nb)^{-1}$, respectively. Then, the mean integrated squared error (MISE) can be calculated by

$$\text{MISE}(\hat{g}) = \frac{b^4 I^2(K)}{4} \int [g''(\tau)]^2 d\tau + \frac{2\pi c_f R(K)}{nb} \int g^2(\tau) d\tau + \max\left\{o(b^4), o\left(\frac{1}{nb}\right)\right\}. \tag{5.9}$$

By minimizing the dominant part of the MISE, the asymptotically optimal bandwidth of $\hat{g}$ is

$$b_A = \left(2\pi c_f \frac{R(K)}{I^2(K)} \frac{\int g^2(\tau) d\tau}{\int [g''(\tau)]^2 d\tau} \right)^{1/5} n^{-1/5}. \tag{5.10}$$

Applying the estimator $\hat{g}$ requires the specification of kernels and bandwidth. Optimal kernels have been obtained analytically. The selection of bandwidth becomes the most important problem when applying nonparametric regression estimators such as kernel estimators. The regression works well, only if the bandwidth is suitable. Because the kernel estimation uses the points around x_0 to estimate the scale function, a kernel regression is usually biased. The larger the bandwidth, the larger the square bias because further points from x_0 are used, but the smaller the variance because more observations are used for estimation. The bandwidth should be optimized to balance the variance and bias. The optimal bandwidth is the one, which can minimize the mean squared error (MSE) or the mean integrated squared error (MISE) (Gasser, 1991). Various methodologies can be applied to select the optimal bandwidth, such as the Cross-Validation (CV), the Generalized CV (GCV) and the iterative plug-in (IPI) methods.

5.2.2 The extensions of SemiGARCH model

Let $r_t = Y_t - \mu, t = 1, \ldots, n$ be the returns from an asset. The SemiAPARCH model is defined as follows

$$r_t = s(\tau_t)\varepsilon_t. \tag{5.11}$$

The conditional variance of the rescaled process h_t follows a parametric APARCH process

$$h_t^{\delta/2} = \omega + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^{\delta} + \sum_{j=1}^{q} \beta_j h_{t-j}^{\delta/2}, \tag{5.12}$$

where $\omega > 0$, $\delta \geq 0$, $\alpha_i \geq 0$, $i = 1, \ldots, p$, $-1 < \gamma_i < 1$, $i = 1, \ldots, p$, $\beta_j \geq 0$, $j = 1, \ldots, q$, γ is the leverage parameter and δ is the parameter for the power term.

The reason to use the SemiAPARCH model rather than the APARCH is that, if the scale function changes over time, the parametric component cannot be estimated consistently from the data, when the nonstationary scale function is not estimated. However, after estimating the nonstationary scale function an approximate stationary process for further analysis can be obtained. When the process follows a parametric model, the semiparametric framework still works but with some loss of efficiency (Feng and Sun, 2013).

From the discussion of kernel regression in Section 5.2.1 we can see that, in order to get an exact estimation, the sample volume should be large enough. However, the bias is of the order b^2. Therefore, the estimation at the boundary has a large bias, which may result in a larger selected bandwidth. This is the so-called boundary problem. In this case, a constrained local linear regression is applied. The scale function $g(\tau_t)$ is estimated from the absolute returns, instead of the squared returns, so as to reduce the moment condition requirement. The estimation of the scale function with absolute returns requires just the existence of the fourth order moment of ε_t. For details, please refer to Chapter 6.

The local linear estimator $\hat{g}(\tau_t) = \hat{a}_0$ at $0 \leq x \leq 1$ can be obtained by minimizing

$$Q = \sum_{t=1}^{n} [|r_t| - a_0 - a_1(\tau_t - \tau)]^2 K(\frac{\tau_t - \tau}{b}). \tag{5.13}$$

The optimal bandwidth for estimating $g(\tau)$ is different from the one for estimating $g^2(\tau)$, due to a constant factor. In this method a fully data-driven algorithm is carried out by adapting an iterative plug-in idea with a starting bandwidth selected by the CV method. If the sample size is limit and a relatively small bandwidth is used, the local linear estimator may be estimated as a negative value. To ensure the non-negativity, the estimator $\hat{g}(\tau)$ is assumed to take its absolute value as $\hat{g}(\tau) = |\hat{g}(\tau)|$. Then the estimation of ε_t can defined as $\hat{\varepsilon}_t = r_t/\hat{g}(\tau_t)$. The unknown parameters of a chosen APARCH model can be estimated by an approximate conditional maximum likelihood estimation. AIC or BIC can be applied to select a suitable parametric model (Feng and Sun, 2013).

Besides, in the semiparametric EGARCH process, the conditional variance follows

$$\ln h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i f(\varepsilon_{t-i}) + \sum_{j=1}^{q} \beta_j \ln h_{t-j}, \tag{5.14}$$

where

$$f(\varepsilon_{t-i}) = \vartheta\varepsilon_{t-j} + \kappa(|\varepsilon_{t-i}| - E|\varepsilon_{t-i}|), i = 1, \ldots, p,$$

$\alpha_0, \alpha_i, \beta_i, \vartheta$ and κ are as defined in Chapter 2.

Similarly, for SemiCGARCH model, the conditional variance h_t is assumed to follow the CGARCH model, which is introduced by Engle and Lee (1999),

$$h_t = q_t + \sum_{i=1}^{p} \alpha_i(\varepsilon_{t-i}^2 - q_{t-i}) + \sum_{j=1}^{q} \beta_j(h_{t-j} - q_{t-j}), \tag{5.15}$$

and

$$q_t = \omega + \rho q_{t-1} + \varphi(\varepsilon_{t-1}^2 - h_{t-1}), \tag{5.16}$$

where q_t is the permanent component of the conditional variance and $(h_{t-j} - q_{t-j})$ is the transitory component of the conditional variance. Obviously, the semiparametric model discussed in this chapter is general and model free, i.e. the stationary conditional variance can be fitted to any GARCH type model.

5.3 The empirical study

In this section, the semiparametric models are applied to ten high-frequency financial data sets, which are the returns of BMW and Allianz at five given fixed trading time points from January 2006 to September 2014, respectively. Usually, in literature, the daily observations are applied, which are composed of the average value or the close price of a trading day. If daily data is considered, the analysis is not accurate enough and the characteristics of the returns at different time points in transaction days cannot be revealed. Therefore, the observations at five specific fixed time points, which are 09:30, 11:00, 12:30, 14:00 and 15:30, are selected. In each data set, the interval of the observations are 24 hours so as to keep the data in a daily frame. Due to the overnight effect, the open price is not

chosen, avoiding that the open price may be obviously unusual from that at the other time points. Besides, the market information release time is able to affect the open price. Furthermore, due to the time differences, abroad information may also cause an abnormal fluctuation in the open price (Tsai et al., 2012). So, in order to avoid the possible risk, stockholders are willing to change their equity or stock holding at the closing time, leading to an abnormal fluctuation in the close price. Therefore, the close price is also excluded.

In the empirical examples, the characters of the semiparametric volatility models, the comparison between the semiparametric volatility models and the parametric volatility models, the daily volatility of returns at the different trading times and the financial crisis influence will be discussed. The estimation of the trend function and the parametric model fitting are carried out in R.

Firstly, the scale function is estimated. The applied method to estimate the semiparametric trend is as discussed above. Here, the bandwidth is automatically selected. An initial bandwidth is given according to CV. An IPI process is carried out until the bandwidth is stable. The standardized returns are calculated by means of removing the estimated scale function and fitted to the parametric APARCH, EGARCH and CGARCH models with the orders (1,1), (1,2), (2,1) and (2,2). In the parametric models, the innovations are assumed to follow a t-distribution. In Table 5.1, we can see that the models with t-distribution have smaller BIC values than the ones with a normal distribution.

5.3.1 The empirical result of Allianz

In this subsection, the proposed algorithm is applied to the returns of Allianz at five given trading time points. The long-term risk of these five data sets is analyzed through the estimated scale function. The short-term risk is estimated by using the APARCH, EGARCH and CGARCH models with the selected orders.

From Fig. 5.1 to Fig. 5.5, the observations, the returns series, the estimated scale functions with $d = 1$ (solid line) and $d = 2$ (dashed line) and the standardized returns, calculated by means of the estimated scale function are plotted.

Table 5.1: BIC of selected models with normal and t-distribution of ALV

	APARCH(1,1)		EGARCH(1,1)		CGARCH(1,1)	
	norm	std	norm	std	norm	std
T = 09:30	2.7459	2.7117	2.7473	2.7108	2.7689	2.7283
T = 11:00	2.7816	2.7542	2.7815	2.7528	2.7996	2.7671
T = 12:30	2.7828	2.7543	2.7826	2.7524	2.8041	2.7690
T = 14:00	2.7738	2.7424	2.7713	2.7396	2.8001	2.7592
T = 15:30	2.7921	2.7558	2.7923	2.7545	2.8084	2.7715

According to the plots in each figure, we can find two dramatic volatility changes in the series, indicating the global financial crisis and the Euro debt crisis, which happened in August 2011. The global financial crisis in 2008 caused serious negative influences on the economy in many European countries. In order to save the banks and the other financial institutions, the sovereign debt increased sharply and exceeded the solvency in several countries. Following, the Euro debt crisis started from the Greece debt crisis in 2010 and then nearly the whole of Europe was involved until September 2011. The high peaks of scale function show that Allianz has extremely high long-term risk during the financial crisis. From these figures, it also can be seen that the volatility at 09:30 is strongest in all the given trading time points because of the overnight effect.

From the plots we can see that, corresponding to the volatility of returns, there are two sub-periods in the scale function with high peaks during the two financial crises. Also, it is shown than the first peak is higher than the second one, indicating that Allianz had higher risk during the global financial crisis than that in the Euro debt crisis. Further, except for the periods of the two financial crises, the scale functions stay at a relatively low level, usually within the confidence intervals, suggesting a stable development of the financial market in the considered periods. However, the stock prices between the two financial crises stay at a relatively low level. We can conclude that the negative influence of the global financial crisis is

continuous and it takes a time to restore investors' confidence and willingness to the financial market.

Comparing the scale functions at the given trading time points, estimated with $d = 1$ and $d = 2$ respectively, we can see that they are almost approaching in the stable periods. The differences often happen at the boundary and during the financial crisis. For example, the estimated scale functions at 12:30 and 14:00 with $d = 1$ are below the ones with $d = 2$ at the outbreak of the two financial crises and are clearly up at the beginning of the crises. Furthermore, the estimated scale functions at 09:30, 11:00 and 15:30 display more differences between with $d = 1$ and $d = 2$. The scale function with $d = 1$ at 09:30 is clearly over the one with $d = 2$ at the beginning of the global financial crisis. Then, the increasing speed of the scale function with $d = 1$ becomes gradually slow. In the worst several months of the financial crisis, the scale function with $d = 1$ stays below the one with $d = 2$. In the Euro debt crisis, the scale function with $d = 1$ is also up to the one with $d = 2$ at the beginning and then the two lines tend to be overlapped.

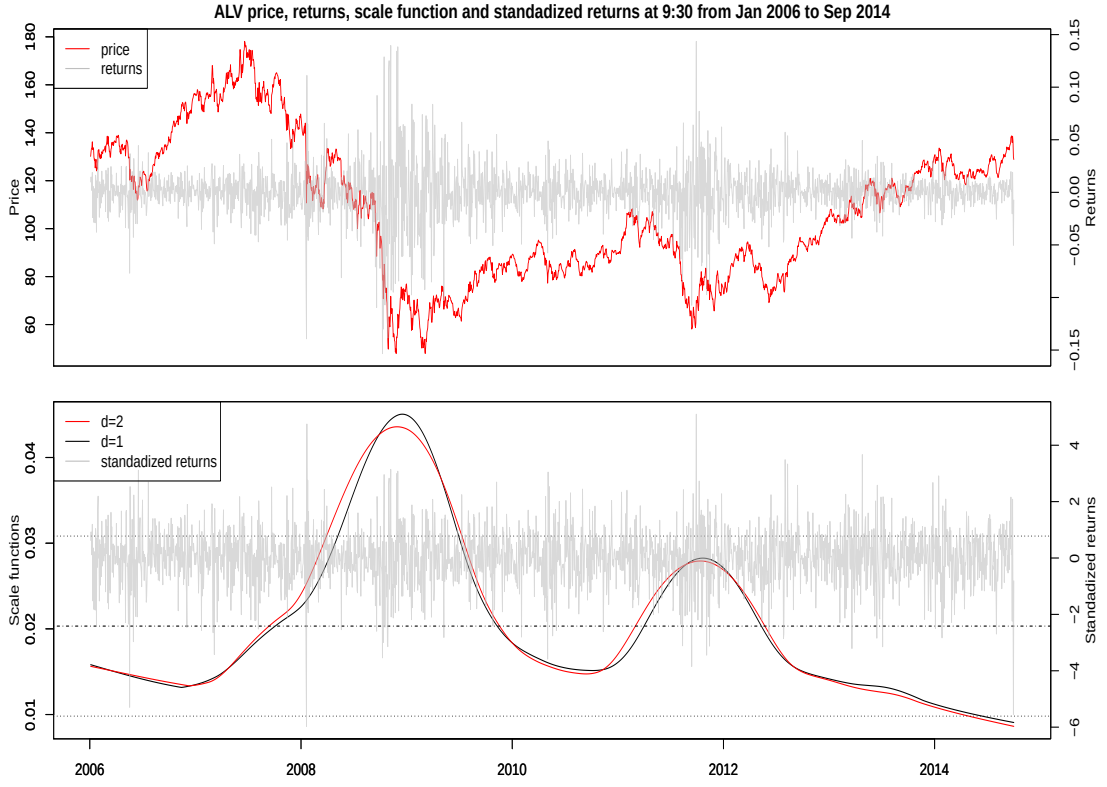

Figure 5.1: The smoothing results of ALV at 09:30

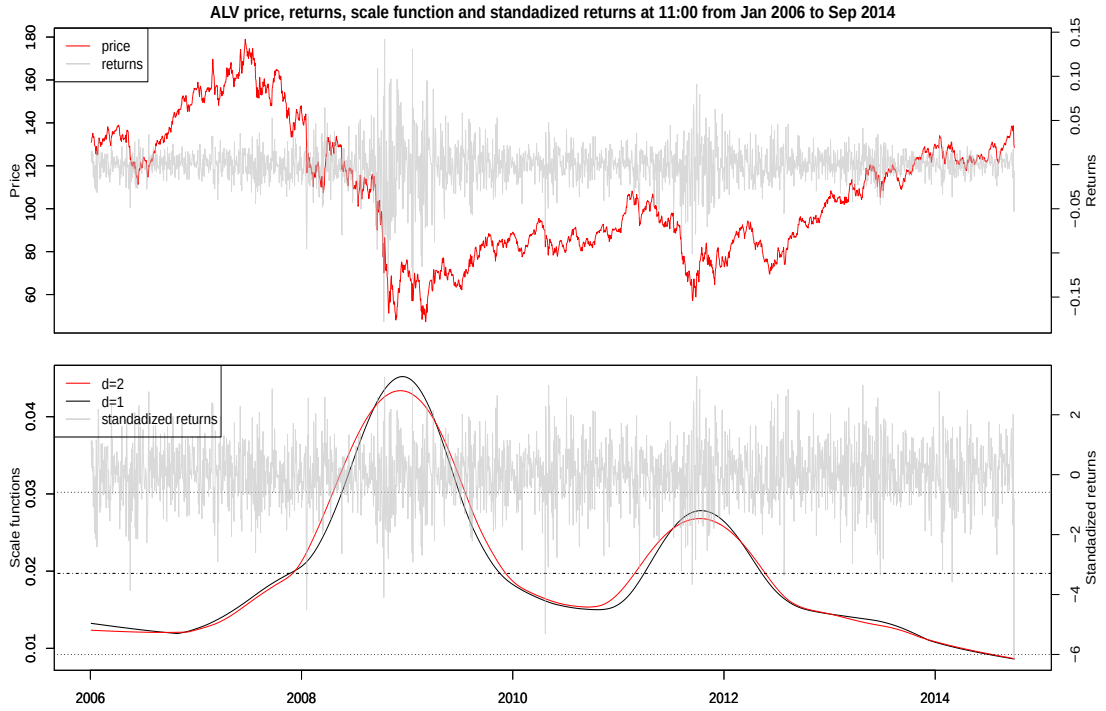

Figure 5.2: The smoothing results of ALV at 11:00

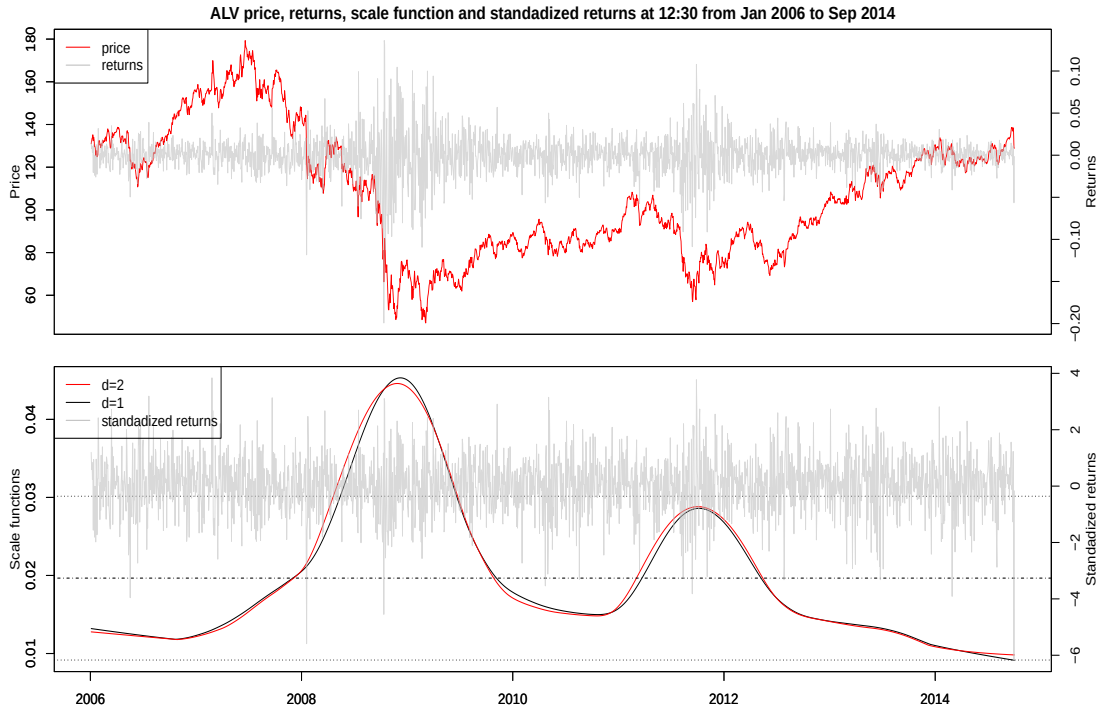

Figure 5.3: The smoothing results of ALV at 12:30

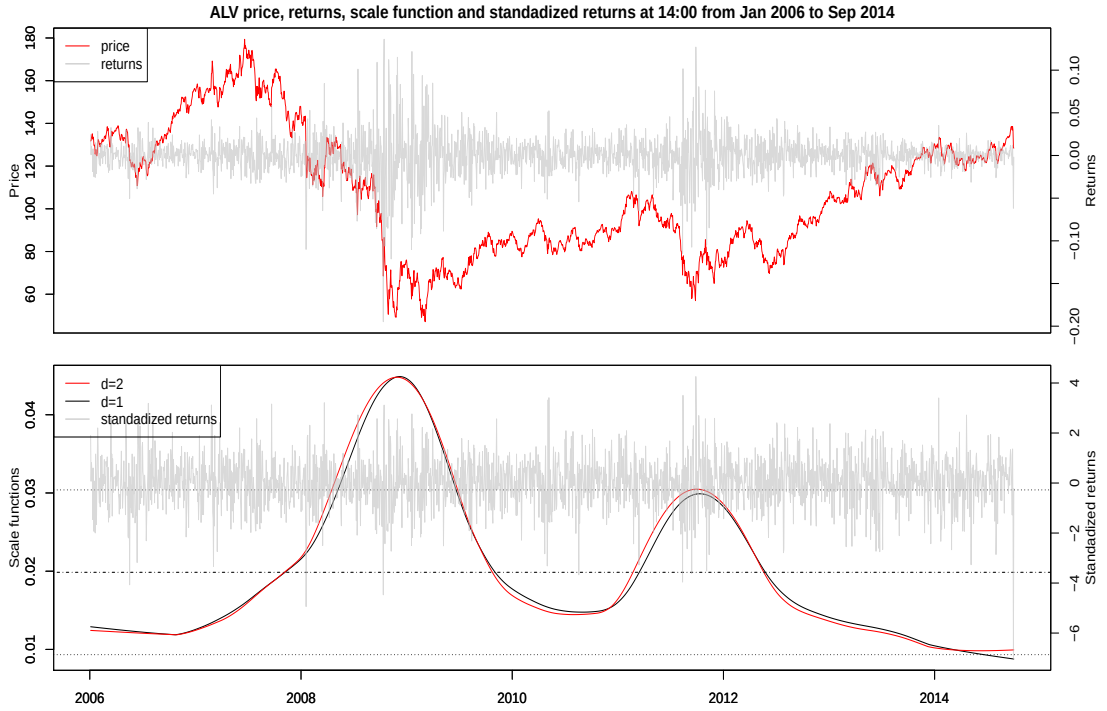

Figure 5.4: The smoothing results of ALV at 14:00

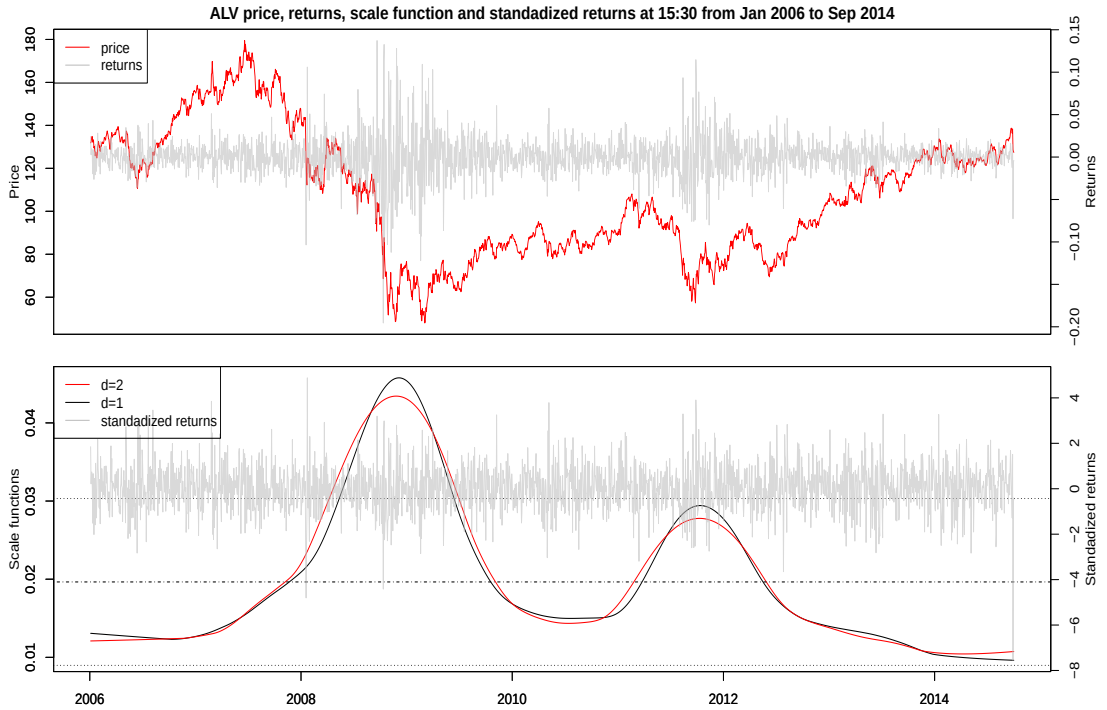

Figure 5.5: The smoothing results of ALV at 15:30

At 11:00 and 15:30, the differences between the scale functions with $d = 1$ and $d = 2$ are more obvious. The scale function with $d = 1$ is higher than the one with $d = 2$ at the beginning of each financial crisis. When the scale functions reach to the highest level, the scale function with $d = 1$ stays at a significantly lower level than the one with $d = 2$. Then it becomes higher than $d = 2$ again as the decline of the financial crisis. Consequently, the scale function with $d = 1$ can make the data relatively stable, especially during the financial crisis. According to figures, the standardized return series are quite stable. However, because the nonparametric and parametric components are almost orthogonal to each other, the series still clearly exhibit the influence of market changes that are not affected by estimating or removing the nonparametric component.

Table 5.2: Selected bandwidths at fixed time points of ALV

T	09:30	11:00	12:30	14:00	15:30
$d = 1$	0.0985	0.0933	0.0912	0.0921	0.1018
$d = 2$	0.1069	0.1046	0.0931	0.1070	0.0878

In Table 5.2, the bandwidths are selected by using the scale function with $d = 1$ and $d = 2$. The selected bandwidths with $d = 1$ are generally smaller than that for $d = 2$ and smaller bandwidths describe changes in data more accurately. Besides, if the scale function with $d = 1$ is applied, the returns are estimated in the absolute form and it only requires the existence of the fourth moment, which is much weaker than the eighth moment requirement with $d = 2$.

The fitted APARCH, EGARCH and CGARCH models with the student-t distribution can be obtained based on the standardized returns. The best order in all cases is the order (1,1) by comparing the BIC in Table 5.3. Therefore, we only discuss the APARCH(1,1), EGARCH(1,1) and CGARCH(1,1) models in this work. If the best model is selected only by the BIC, the EGARCH models are the best model in all five cases and the CGARCH models are the worst. However, different models can be used in different economic situations, i.e. the APARCH and EGARCH models can show the leverage effect, while the CGARCH models

exhibit the persistence of the shocks in the long-term and the short-term.

Table 5.3: BIC of all selected models of ALV

T	09:30	11:00	12:30	14:00	15:30
APARCH-t(1,1)	2.7117	2.7542	2.7543	2.7424	2.7558
APARCH-t(1,2)	2.7151	2.7577	2.7577	2.7458	2.7593
APARCH-t(2,1)	2.7183	2.7611	2.7606	2.7493	2.7622
APARCH-t(2,2)	2.7218	2.7645	2.7641	2.7524	2.7656
EGARCH-t(1,1)	2.7108	2.7528	2.7524	2.7396	2.7545
EGARCH-t(1,2)	2.7142	2.7562	2.7558	2.7430	2.7580
EGARCH-t(2,1)	2.7176	2.7597	2.7579	2.7458	2.7609
EGARCH-t(2,2)	2.7204	2.7631	2.7614	2.7492	2.7646
CGARCH-t(1,1)	2.7283	2.7671	2.7690	2.7592	2.7715
CGARCH-t(1,2)	2.7309	2.7706	2.7725	2.7627	2.7750
CGARCH-t(2,1)	2.7317	2.7706	2.7721	2.7625	2.7742
CGARCH-t(2,2)	2.7342	2.7740	2.7756	2.7660	2.7777

From Table 5.4 to 5.8 display the estimated coefficients for Allianz at the five given trading time points, respectively. From the value of shapes, we can see that the APARCH model usually has the largest degree of freedom of the distribution in each case, while the CGARCH model has the smallest. At 09:30, 11:00 and 14:00 the degrees of freedom are all between 6 and 7,5 and in all models at 12:30 and in the EGARCH and CGARCH models at 15:30 the degrees of freedom are almost equal to 8. It means that the distributions in these trading time points are nearly heavy-tailed and the eighth moment of ε_t does not exist, but the fourth moment. In the APARCH model at 15:30 the degree of freedom is 8.13. Now the distribution is also nearly heavy-tailed but the eighth moment of ε_t exists. The possibility of an extreme return at 09:30, 11:00 and 14:00 is higher than that at 12:30 and 15:30. From the above discussion, the eighth moment in most cases does not exist. This also expresses that the selection of the scale function with $d = 1$ is better.

Table 5.4: Estimated coefficients of the selected models of ALV at 09:30

	APARCH(1,1)		EGARCH(1,1)		CGARCH(1,1)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
μ	0.0308	0.0182	0.0235	0.0181	0.0462	0.0180
ω	0.0721	0.0172	-0.0085	0.0060	0.0047	0.0006
α_1	0.0656	0.0309	-0.1215	0.0208	0.1018	0.0191
β_1	0.8411	0.0288	0.9281	0.0167	0.8351	0.0300
γ_1	0.7207	0.3454	0.1729	0.0304	-	-
δ	1.8156	0.3597	-	-	-	-
η_{11}	-	-	-	-	0.9956	0.0000
η_{21}	-	-	-	-	0.0000	0.0000
shape	6.5064	0.8851	6.3915	0.8530	6.0607	0.7792

Table 5.5: Estimated coefficients of the selected models of ALV at 11:00

	APARCH(1,1)		EGARCH(1,1)		CGARCH(1,1)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
μ	0.0279	0.0189	0.0282	0.0191	0.0442	0.0187
ω	0.0868	0.0219	-0.0068	0.0059	0.0054	0.0006
α_1	0.0717	0.0226	-0.1202	0.0215	0.0915	0.0188
β_1	0.8336	0.0341	0.9160	0.0210	0.8412	0.0347
γ_1	0.6554	0.2800	0.1427	0.0304	-	-
δ	1.6610	0.3892	-	-	-	-
η_{11}	-	-	-	-	0.9949	0.0000
η_{21}	-	-	-	-	0.0000	0.0000
shape	7.3037	1.1341	7.1364	1.0881	6.9448	1.0259

Table 5.6: Estimated coefficients of the selected models of ALV at 12:30

	APARCH(1,1)		EGARCH(1,1)		CGARCH(1,1)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
μ	0.0242	0.0190	0.0201	0.0190	0.0438	0.0188
ω	0.0858	0.0204	-0.0068	0.0057	0.0034	0.0003
α_1	0.0724	0.0215	-0.1192	0.0208	0.0921	0.0181
β_1	0.8398	0.0311	0.9174	0.0200	0.8330	0.0324
γ_1	0.7439	0.2797	0.1439	0.0307	-	-
δ	1.4690	0.2978	-	-	-	-
η_{11}	-	-	-	-	0.9967	0.0000
η_{21}	-	-	-	-	0.0000	0.0000
shape	7.9157	1.2338	7.7971	1.1984	7.2398	1.0469

Table 5.7: Estimated coefficients of the selected models of ALV at 14:00

	APARCH(1,1)		EGARCH(1,1)		CGARCH(1,1)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
μ	0.0237	0.0187	0.0231	0.0180	0.0428	0.0186
ω	0.0957	0.0204	-0.0094	0.0067	0.0038	0.0004
α_1	0.0884	0.0199	-0.1339	0.0220	0.1000	0.0191
β_1	0.8292	0.0291	0.9039	0.0210	0.8207	0.0327
γ_1	0.8207	0.2096	0.1647	0.0324	-	-
δ	1.1473	0.2595	-	-	-	-
η_{11}	-	-	-	-	0.9963	0.0000
η_{21}	-	-	-	-	0.0000	0.0000
shape	7.4317	1.0993	7.4128	1.0943	6.8385	0.9355

Table 5.8: Estimated coefficients of the Selected models at of ALV 15:30

	APARCH(1,1)		EGARCH(1,1)		CGARCH(1,1)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
μ	0.0119	0.0190	0.0085	0.0189	0.0295	0.0190
ω	0.0703	0.0189	-0.0052	0.0052	0.0041	0.0004
α_1	0.0636	0.0221	-0.1108	0.0197	0.0830	0.0173
β_1	0.8570	0.0309	0.9281	0.0192	0.8462	0.0333
γ_1	0.6649	0.2724	0.1378	0.0317	-	-
δ	1.7361	0.4075	-	-	-	-
η_{11}	-	-	-	-	0.9960	0.0000
η_{21}	-	-	-	-	0.0000	0.0000
shape	8.1277	1.2303	7.9807	1.1877	7.6925	1.1261

In the APARCH models, the leverage parameters γ for the five given trading time points are 0.72, 0.83, 0.74, 0.82 and 0.66, respectively. This means that the leverage effect of Allianz is always strong and the contribution of a negative return is more than the contribution of a positive return. In the EGARCH models, α is the sign effect and γ is the size effect. According to the tables, the estimated α is smaller than -0.1 in all cases. So, we can say that this model is also able to expresses the leverage effect. From the estimated coefficients of the CGARCH model, all the ρ values are larger than 0.99, and φ are equal to 0, leading to a smaller immediate impact of shocks on the long-run component than that on the short-run component. Because the ρ value is close to one, the shock cannot only cause the change of short-term volatility but also keep this abnormal volatility in the long term. The value of $(\alpha+\beta)$ is between 0.9 and 1, and $0 < (\alpha+\beta) < \rho < 1$. It indicates that volatility can reflect the shock immediately and the persistence is long. The impact of volatility on the short-run component will diminish as well but will be more persistent on the long-run component.

From Fig. 5.6 to 5.10, all of the APARCH, EGARCH and CGARCH models' volatility can express the financial crisis. However, when there is positive news, the APARCH and EGARCH models have lower volatility than the CGARCH model, like the marked areas by square.

In the circle marked areas, there is higher volatility for negative news and obviously, the negative news causes a larger change than positive one. For another, the fitted volatility of CGARCH is extremely high due to the missing earlier returns, required in the first several points estimation. Generally, the first several estimated points are not considered in the discussion. Although the extreme points appear in the volatility, they are ignored without any influence in the following fitted volatility.

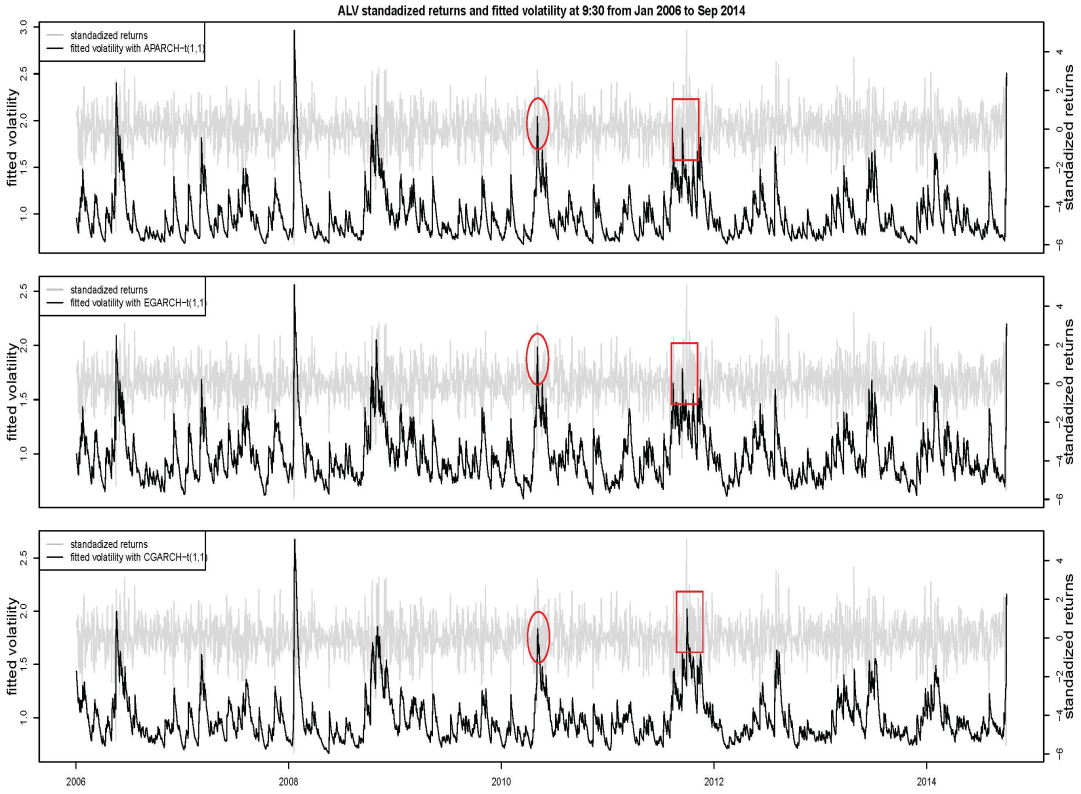

Figure 5.6: The volatility series of different models of ALV at 09:30

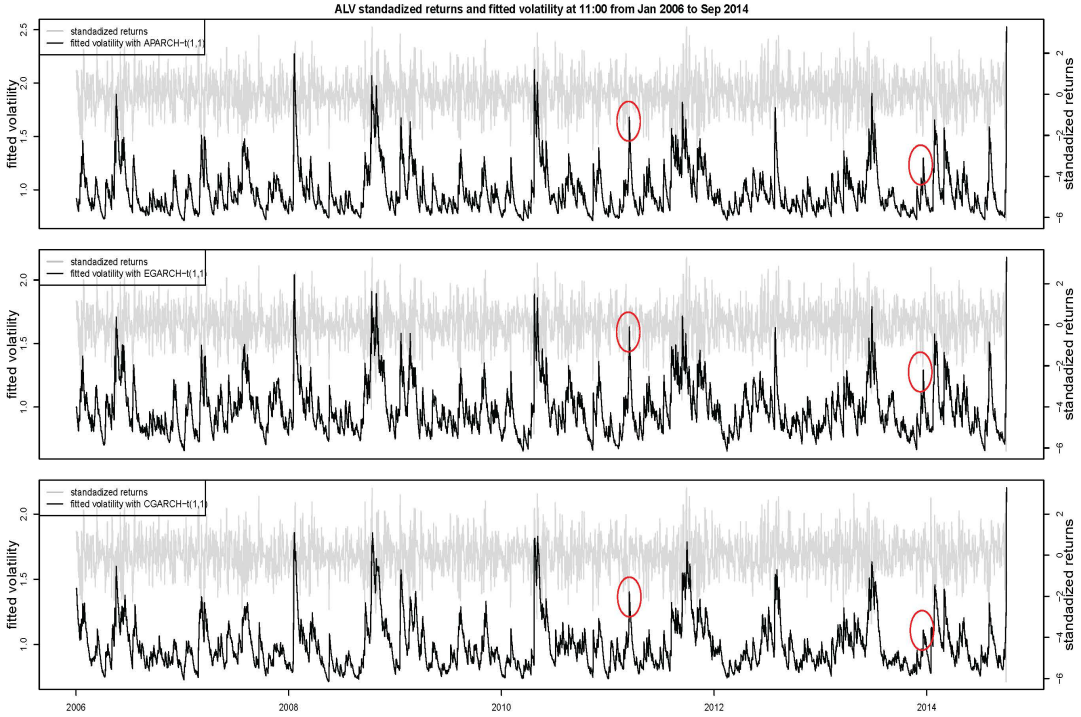

Figure 5.7: The volatility series of different models of ALV at 11:00

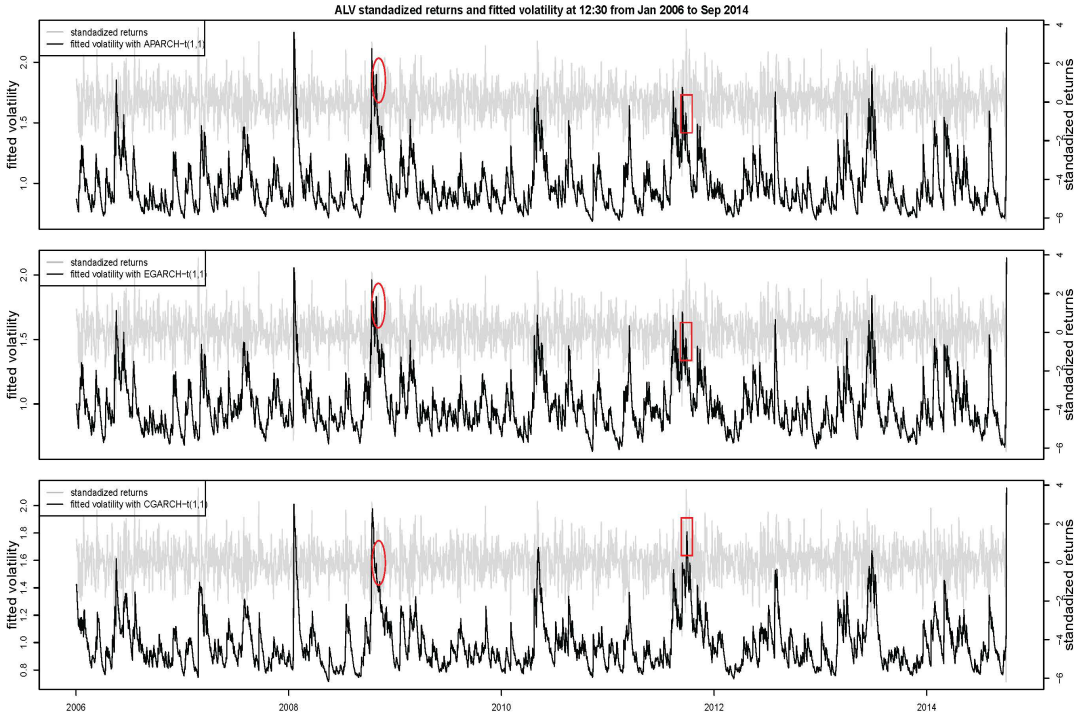

Figure 5.8: The volatility series of different models of ALV at 12:30

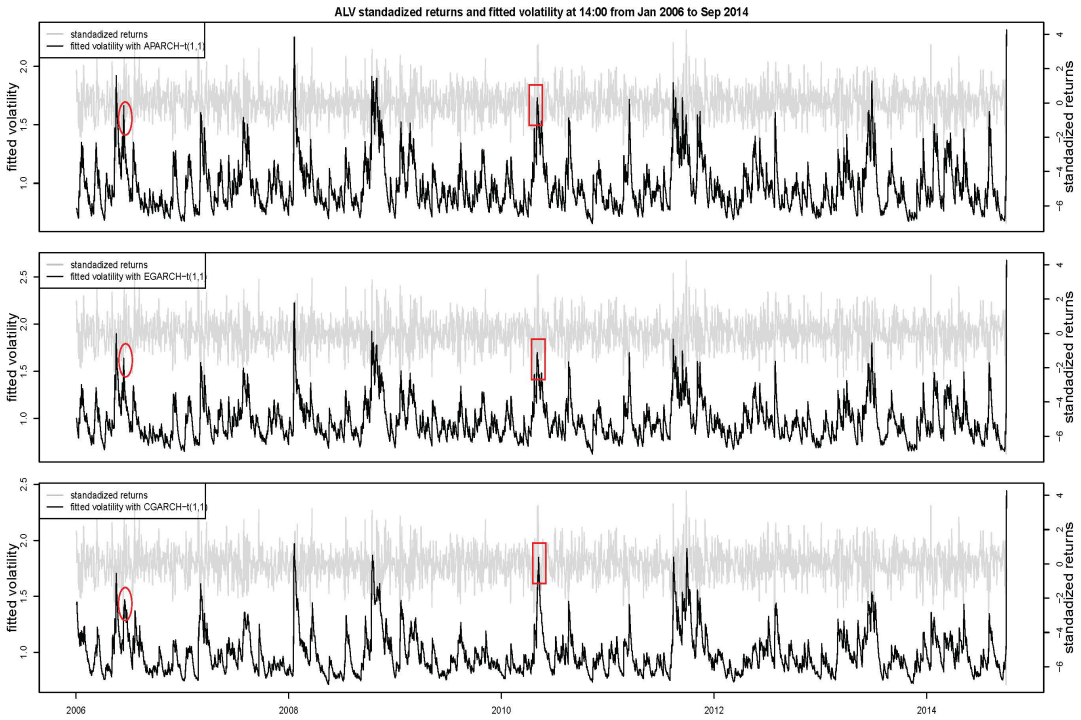

Figure 5.9: The volatility series of different models of ALV at 14:00

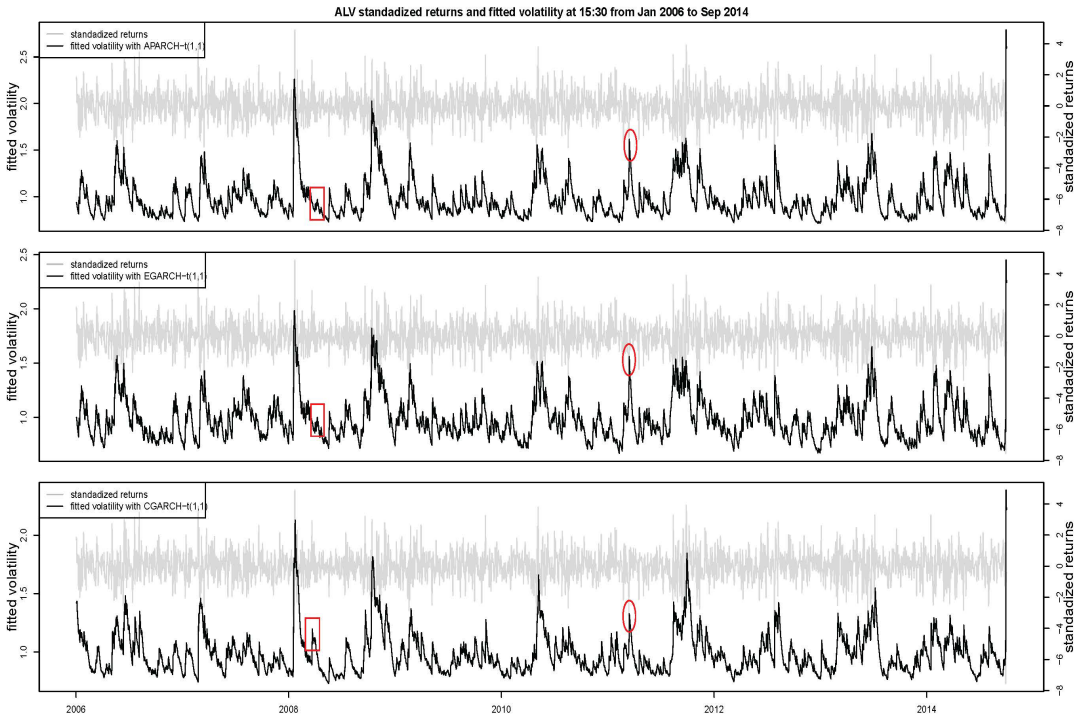

Figure 5.10: The volatility series of different models of ALV at 15:30

5.3.2 The empirical result of BMW

In this section, the semiparametric volatility models are applied to the stock price of BMW at five given trading time points, i.e. 09:30, 11:00, 12:30, 14:00 and 15:30 from January 2006 to September 2014. Similar to the analysis of Allianz, the semiparametric volatility models are also used to discuss the risk and the leverage effect by the estimated scale function and the conditional heteroskedasticity.

The observations, the log-returns, the estimated scale functions with $d = 1$ and $d = 2$ and the standardized returns are displayed from Fig. 5.11 to 5.15. In the figures, the returns change dramatically in two sub-periods, which corresponding to the global financial crisis and the Euro debt crisis. Although the volatility change between the two financial crisis is relatively stable, it fluctuates still stronger than before and after the financial crisis, showing the relatively higher risk. Besides, the volatilities at 09:30 and 11:00 are stronger than that at 12:30, 14:00 and 15:30, possibly due to the overnight effect. The volatility near close time is weaker than that at open time, but stronger than that at the other transaction time. Correspondingly, it is found that the estimated scale functions have two peaks. Furthermore, the scale functions stay at an extremely high level (out of the confidence interval) in the global financial crisis and a relatively high level (within the confidence interval) in the Euro debt crisis. Obviously, BMW has a higher risk in the global financial crisis.

The estimated scale functions at each given time is calculated with $d = 1$ and $d = 2$. In this case, at the boundary there is no significant difference between the scale functions with $d = 1$ and $d = 2$ at 09:30, 11:00, 12:30 and 14:00. However, the scale function with $d = 2$ at 15:30 is below that with $d = 1$ after the year 2014. In the two financial crises, the two lines show up clear differences. From the beginning of the global financial crisis, the estimated scale function with $d = 2$ is up to the one with $d = 1$ at the five given trading time points. At 12:30 and 14:00, the level of the scale functions with $d = 1$ are higher than the others, so the difference between the two estimated scale functions is limited at 12:30 and 14:00. Only the scale functions at 09:30 tend to gradually overlap after the drop

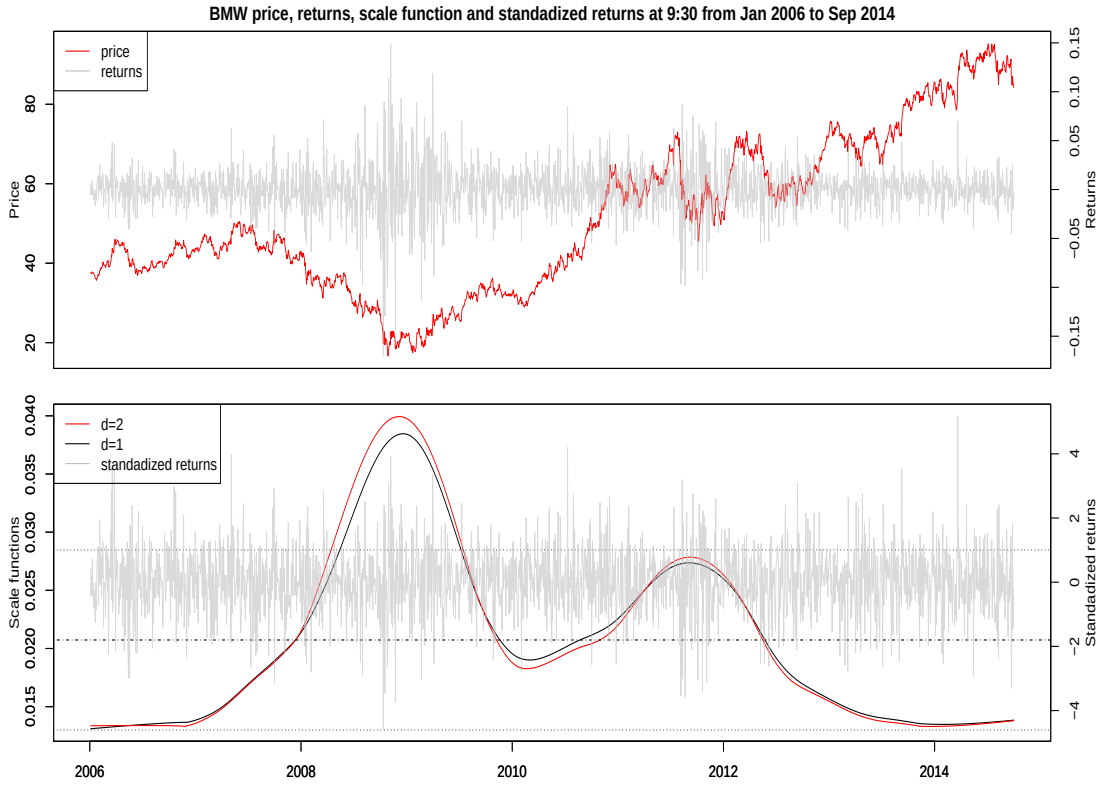

Figure 5.11: The smoothing results of BMW at 09:30

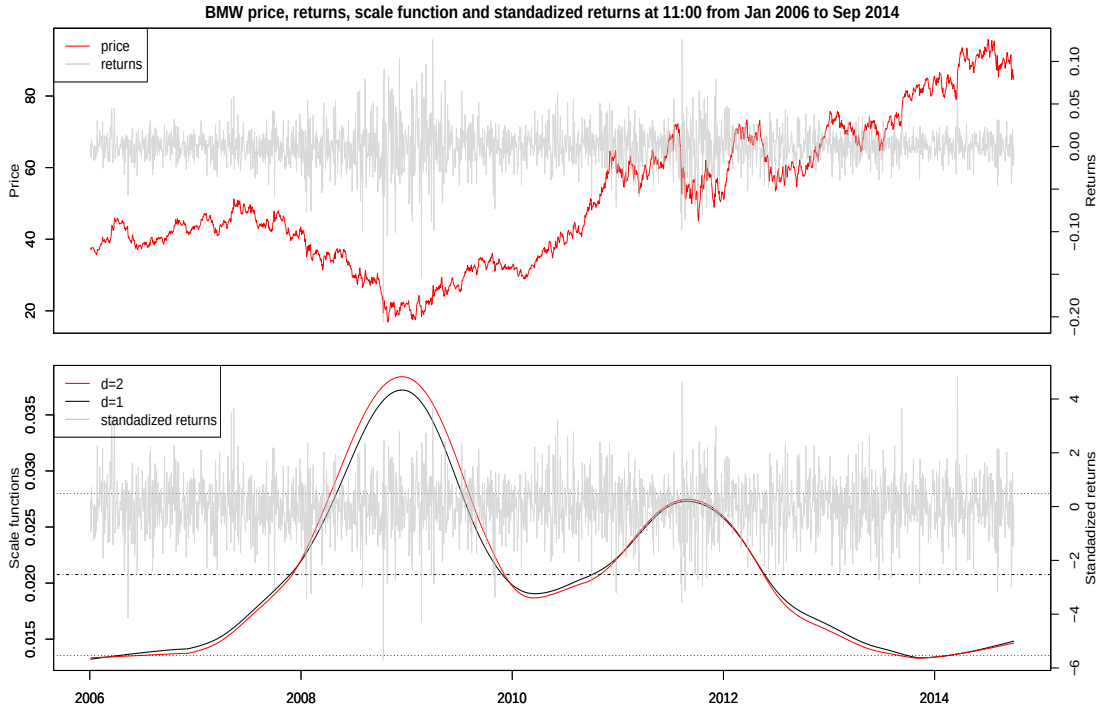

Figure 5.12: The smoothing results of BMW at 11:00

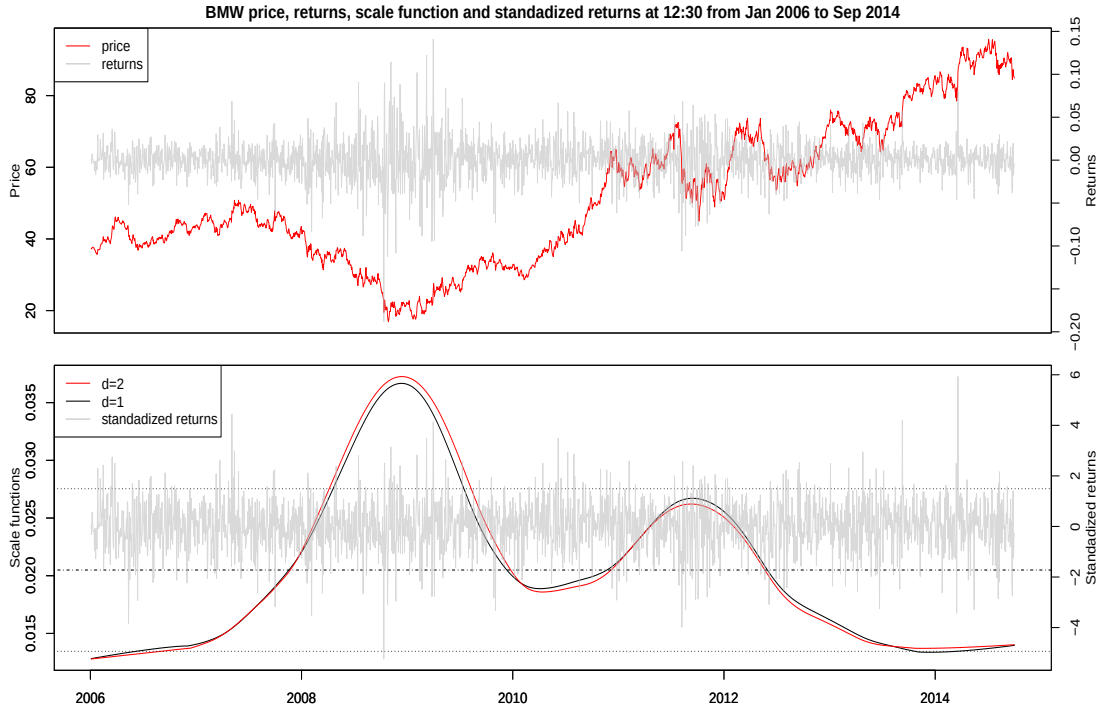

Figure 5.13: The smoothing results of BMW at 12:30

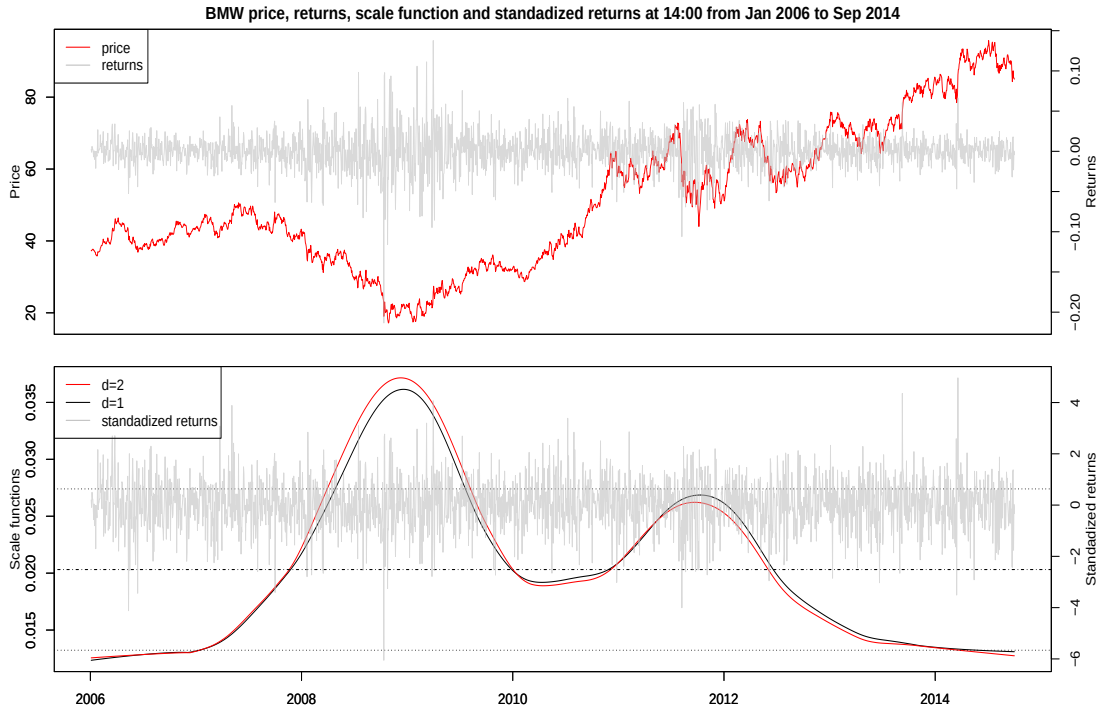

Figure 5.14: The smoothing results of BMW at 14:00

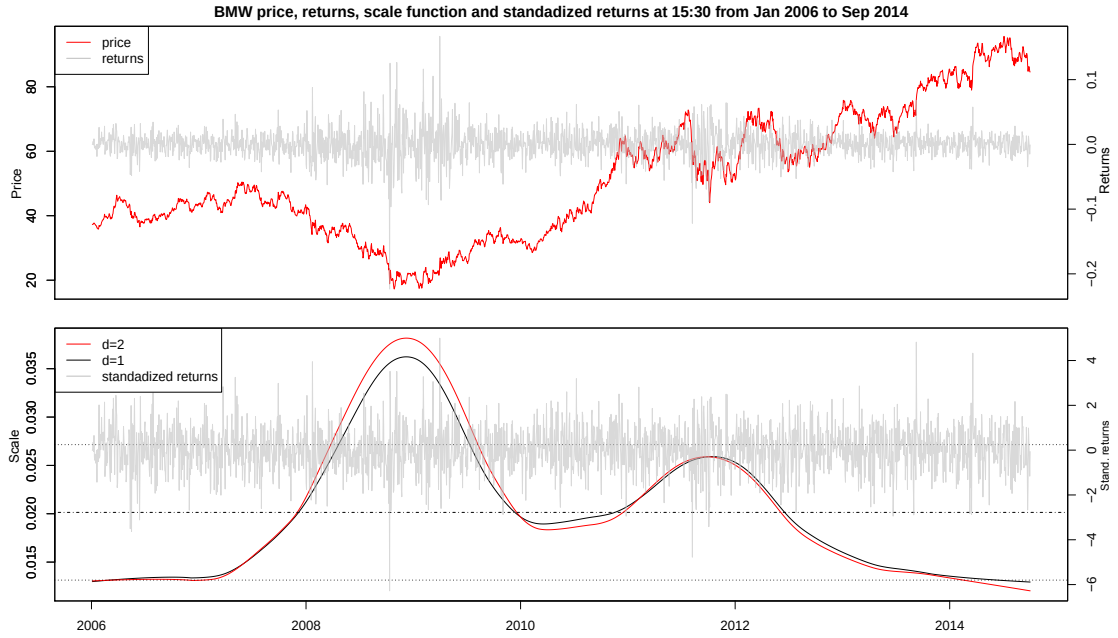

Figure 5.15: The smoothing results of BMW at 15:30

off the peak, However, the rest have still intersections. In the Euro debt crisis, the scale function with $d = 1$ at 09:30 is below the one with $d = 2$ and then overlaps. At 12:30, 14:00 and 15:30 the overlap of the two lines remains in short. Then, the scale function with $d = 2$ is below the one with $d = 1$, however above at the beginning of the second financial crisis.

Table 5.9: Selected bandwidths at fixed time points of BMW

T	09:30	11:00	12:30	14:00	15:30
$d = 1$	0.1047	0.1043	0.1047	0.1050	0.1054
$d = 2$	0.1020	0.1061	0.1075	0.1071	0.1067

Comparing the selected bandwidth with $d = 1$ and $d = 2$ in Table 5.9, smaller bandwidth is always obtained with $d = 1$. Therefore, the standardized returns are estimated by the scale function with power one. The fitted APARCH, EGARCH and CGARCH models of orders (1,1), (1,2), (2,1) and (2,2) are fitted to the standardized returns. The BIC values are listed in Table 5.10. In this example, the best order of all cases is also order (1,1).

Table 5.10: BIC of all selected models of BMW

T	09:30	11:00	12:30	14:00	15:30
APARCH-t(1,1)	2.7862	2.7904	2.7923	2.8003	2.7906
APARCH-t(1,2)	2.7897	2.7933	2.7958	2.8039	2.7941
APARCH-t(2,1)	2.7931	2.7973	2.7993	2.8073	2.7964
APARCH-t(2,2)	2.7965	2.8002	2.8016	2.8108	2.7996
EGARCH-t(1,1)	2.7842	2.7881	2.7905	2.7988	2.7892
EGARCH-t(1,2)	2.7878	2.7909	2.7941	2.8024	2.7927
EGARCH-t(2,1)	2.7912	2.7938	2.7969	2.8058	2.7933
EGARCH-t(2,2)	2.7945	2.7968	2.7995	2.8049	2.7968
CGARCH-t(1,1)	2.7959	2.7976	2.8004	2.8055	2.8026
CGARCH-t(1,2)	2.7995	2.8000	2.8037	2.8092	2.8064
CGARCH-t(2,1)	2.7996	2.8013	2.8041	2.8092	2.8041
CGARCH-t(2,2)	2.8027	2.8035	2.8071	2.8119	2.8076

In the following tables, the shape values at 09:30, 11:00, 12:30 and 14:00 are between 5 and 8. The distribution is nearly heavy tails and the eighth moment of the models does not exist but the fourth moment exists. At 15:30 the shape values of the APARCH, EGARCH and CGARCH models are all larger than 8, meaning that the distribution is also nearly heavy tails and the eighth moment exists. Also, it is found in the tables that the estimated leverage parameters γ of the APARCH(1,1) model at 09:30, 11:00, 12:30 and 15:30 are 0.72, 0.8, 0.81 and 0.797, indicating a strong leverage effect, while at 14:00, the value of γ is just 0.37, showing that the leverage effect is weak. The estimated sign effect α of the EGARCH(1,1) model at these five trading time points are -0.072, -0.069, -0.062, -0.056 and -0.09, which are all smaller than 0. The leverage effect exists in the EGARCH model and it reach its peak at 15:30 and its low at 14:00.

In the CGARCH(1,1) model the estimated α are larger than 0.99 and φ are equal to 0 at the trading time points. The sum of α and β are around 0.9 but

Table 5.11: Estimated coefficients of the selected models of BMW at 09:30

	APARCH(1,1)		EGARCH(1,1)		CGARCH(1,1)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
μ	0.0260	0.0191	0.0261	0.0185	0.0370	0.0193
ω	0.0597	0.0185	-0.0012	0.0046	0.0070	0.0005
α_1	0.0471	0.0152	-0.0726	0.0230	0.0602	0.0199
β_1	0.9009	0.0253	0.9392	0.0555	0.8522	0.0654
γ_1	0.7239	0.3196	0.0941	0.0539	-	-
δ	1.2406	0.3924	-	-	-	-
η_{11}	-	-	-	-	0.9932	0.0000
η_{21}	-	-	-	-	0.0000	0.0000
shape	6.5687	0.9092	6.5137	0.9122	6.3533	0.8636

Table 5.12: Estimated coefficients of the selected models of BMW at 11:00

	APARCH(1,1)		EGARCH(1,1)		CGARCH(1,1)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
μ	0.0278	0.0194	0.0281	0.0193	0.0372	0.0193
ω	0.0752	0.0269	-0.0006	0.0045	0.0060	0.0004
α_1	0.0426	0.0168	-0.0692	0.0183	0.0632	0.0187
β_1	0.8898	0.0354	0.9276	0.0260	0.7930	0.0713
γ_1	0.8004	0.4057	0.0843	0.0292	-	-
δ	1.1979	0.3336	-	-	-	-
η_{11}	-	-	-	-	0.9942	0.0000
η_{21}	-	-	-	-	0.0000	0.0000
shape	6.0425	0.7978	5.9940	0.7878	6.0006	0.7820

Table 5.13: Estimated coefficients of the selected models of BMW at 12:30

	APARCH(1,1)		EGARCH(1,1)		CGARCH(1,1)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
μ	0.0363	0.0158	0.0340	0.0194	0.0386	0.0195
ω	0.0661	0.0219	-0.0018	0.0045	0.0050	0.0003
α_1	0.0453	0.0137	-0.0617	0.0175	0.0644	0.0186
β_1	0.9019	0.0302	0.9276	0.0254	0.8041	0.0671
γ_1	0.8082	0.2651	0.0956	0.0291	-	-
δ	0.7496	0.3598	-	-	-	-
η_{11}	-	-	-	-	0.9951	0.0000
η_{21}	-	-	-	-	0.0000	0.0000
shape	6.7253	0.9246	6.6818	0.9171	6.6061	0.8890

Table 5.14: Estimated coefficients of the selected models of BMW at 14:00

	APARCH(1,1)		EGARCH(1,1)		CGARCH(1,1)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
μ	0.0317	0.0198	0.0317	0.0191	0.0386	0.0196
ω	0.0945	0.0340	-0.0021	0.0050	0.0028	0.0002
α_1	0.0502	0.0181	-0.0557	0.0179	0.0633	0.0170
β_1	0.8547	0.0445	0.9140	0.0287	0.8192	0.0598
γ_1	0.3677	0.2215	0.1083	0.0293	-	-
δ	1.7718	0.7278	-	-	-	-
η_{11}	-	-	-	-	0.9973	0.0000
η_{21}	-	-	-	-	0.0000	0.0000
shape	7.6362	1.1848	7.6910	1.2046	7.4619	1.1223

Table 5.15: Estimated coefficients of the selected models of BMW at 15:30

	APARCH(1,1)		EGARCH(1,1)		CGARCH(1,1)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
μ	0.0203	0.0197	0.0172	0.0197	0.0337	0.0196
ω	0.1006	0.0276	-0.0034	0.0056	0.0008	0.0000
α_1	0.0493	0.0217	-0.0898	0.0199	0.0644	0.0157
β_1	0.8462	0.0362	0.9018	0.0259	0.8229	0.0472
γ_1	0.7967	0.4200	0.1152	0.0275	-	-
δ	1.5174	0.4334	-	-	-	-
η_{11}	-	-	-	-	0.9992	0.0000
η_{21}	-	-	-	-	0.0000	0.0000
shape	8.6502	1.4846	8.6297	1.4785	8.0625	1.2700

smaller than 0.99. The correlation of these parameters is obvious $0 < (\alpha + \beta) < \rho < 1$. Therefore, volatility can reflect the shock immediately. The volatility impact on the short-run component diminish soon but last long on the long-run component.

In the figures, if there is positive news, the APARCH and EGARCH models have lower volatility than the CGARCH model as the marked area by square and have higher volatility for negative news as the marked area by circle. Moreover, the negative news causes a larger change than positive news. However, at 14:00 the APARCH model expresses a weak leverage effect. From Fig. 5.19, the leverage effect also cannot be detected by comparing with the EGARCH and CGARCH models. For the CGARCH models, the first few estimations can also be neglected.

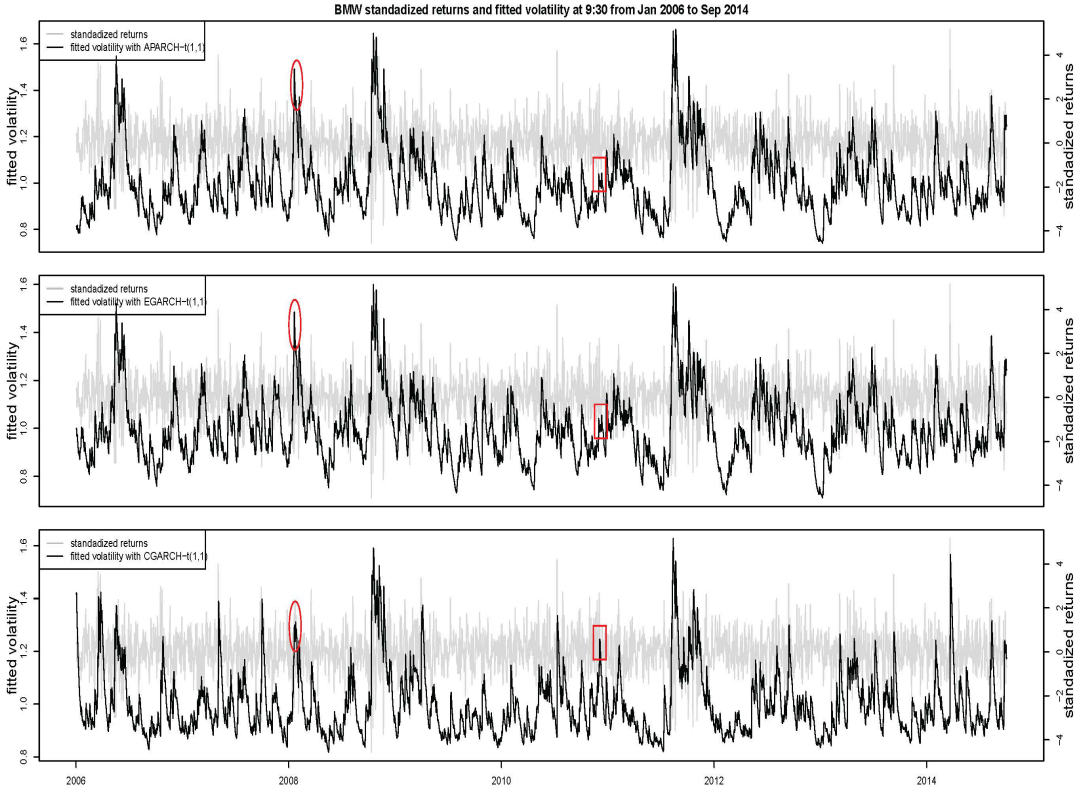

Figure 5.16: The volatility series of different models of BMW at 09:30

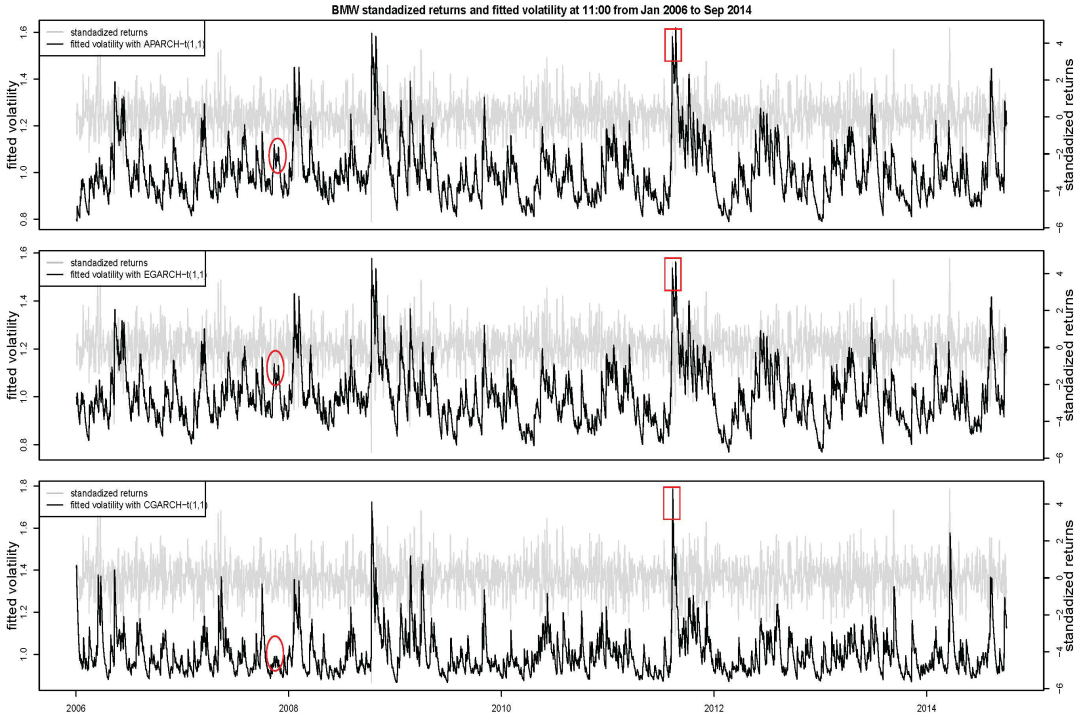

Figure 5.17: The volatility series of different models of BMW at 11:00

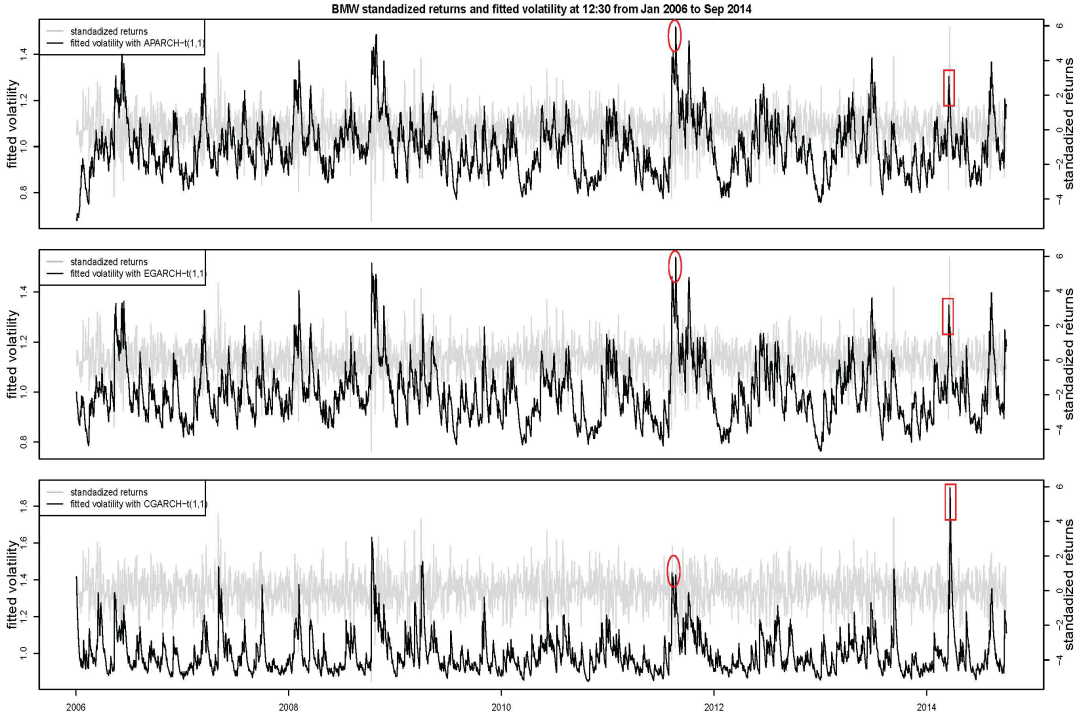

Figure 5.18: The volatility series of different models of BMW at 12:30

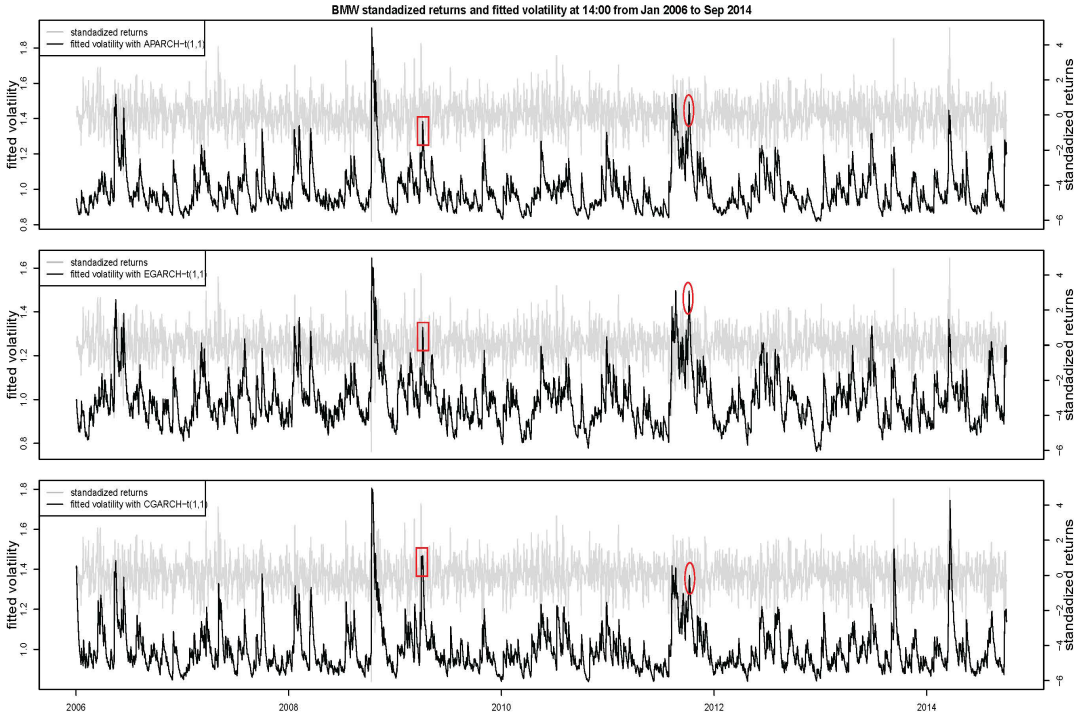

Figure 5.19: The volatility series of different models of BMW at 14:00

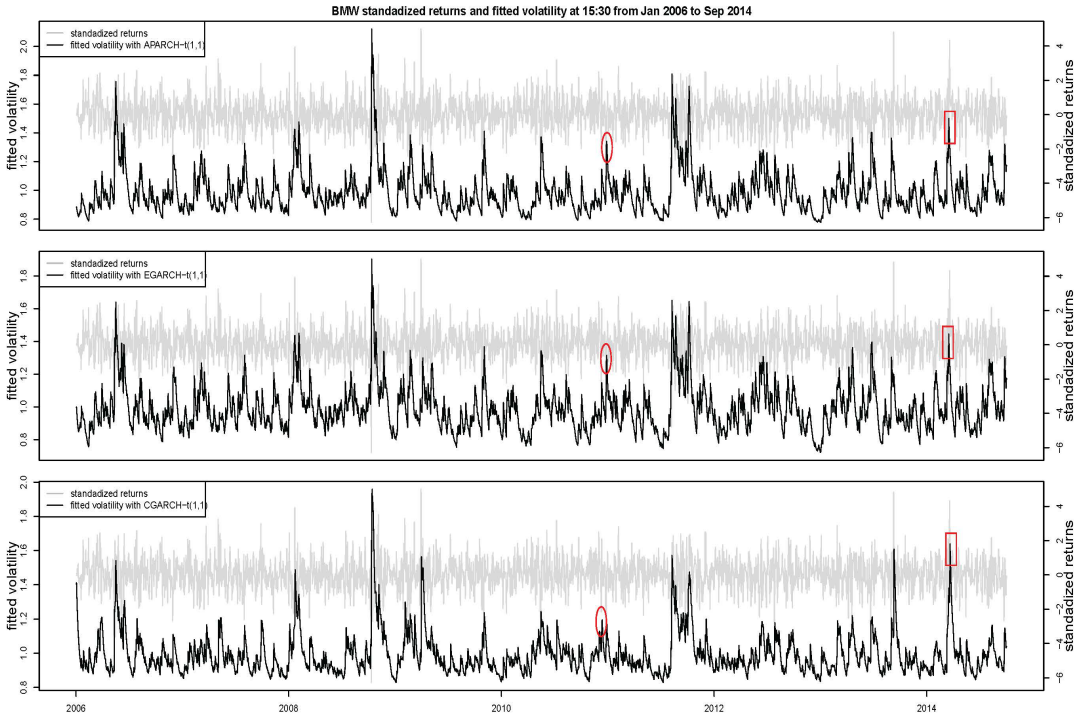

Figure 5.20: The volatility series of different models of BMW at 15:30

5.4 Final remarks

In this chapter, the semiparametric volatility models are applied to the high-frequency returns at fixed trading time points by introducing a smooth scale function into the standard GARCH model. Therefore, the conditional heteroskedasticity and scale change in a financial time series can be modeled at the same time. In the empirical research, the semiparametric model works well with the data at fixed trading time points. It can express the trend of the returns and the leverage effect at the different trading time points. The SemiAPARCH and SemiEGARCH model show up the leverage effect and in the SemiAPARCH model, the leverage effect is obvious. In the SemiCGARCH model, the immediate shock impact on the short-run component is detected and the persistence in the long-run is also strong.

Chapter 6

A Box-Cox semiparametric multiplicative error model[1]

A general class of SemiMEM (semiparametric multiplicative error) models is proposed by introducing a scale function into a MEM (multiplicative error) class model to analyze the non-negative observations. The estimation of the scale function is not limited by any parametric models specification and the moments condition is also reduced via the Box-Cox transformation. For the purpose, an equivalent scale function is applied in a local linear approach and converted to the scale function under weak moment conditions. The equivalent scale function estimation and the bandwidth, the constant factor in the asymptotic variance and the power transformation parameters estimation are proposed based on the iterative plug-in (IPI) algorithms. In the power transformation estimation, the maximum likelihood estimation (MLE), the normality test and the quantile-quantile regression (QQr) are employed and simulation algorithms for the confidence interval of estimated power transformation parameter are also developed by the block bootstrap method. The algorithms fit the selected real data well.

6.1 Introduction

The MEM model was built up by Engle (2002) to model the common non-negative financial data in practice, such as the mean duration (MD), the absolute returns

[1]Chapter 6 is based on the working paper: *A Box-Cox semiparametric multiplicative error model* (Zhang, 2019a), CIE, 2019–05.

(AR), the trading volume (VO) and the trading number (TR). The MEM models were first developed as autoregressive conditional duration (ACD, Engle and Russell, 1998) model for featuring the stochastic process of the time between events. Manganelli (2005) applied the MEM model to the non-negative expected trading volume series and proposed the autoregressive conditional volume (ACV) model. Besides, the MEM framework was extended to a more general functional form by Box-Cox transformation. The Box-Cox ACD model (Dufour and Engle, 2000; Hautsch, 2002; Fernandes and Grammig, 2006) were proposed as a flexible model to analyze the process of the conditional mean to recent durations based on Box-Cox transformation. If the Box-Cox parameter reduces to zero, the duration can be modeled by the Log-ACD model (Bauwens and Giot, 2000) or the EACD model (Karanasos, 2008) and the log-data follows an ARMA process. Recently, the MEM model is also applied widely. Taylor and Xu (2017) proposed a log-vMEM model, discussing the cross-dependent error terms and the non-negative conditional mean without any restriction. The multiplicative error model with volatility jumps (MEM-J, Caporin et al., 2017) was developed to investigate the probability and density of the extreme values in the daily volatility. The MEM model is also feasible under the semiparametric framework, such as the SemiGARCH model (Feng, 2004), the Spline-GARCH model (Engle and Rangel, 2008), the general Box-Cox SemiGARCH model (Zhang et al., 2017), etc.

In this chapter, a general class of semiparametric MEM model, applying a time-varying scale function with the Box-Cox transformation into the MEM model to analyze the non-negative financial time series, is developed. Nonparametric estimation of the scale function is studied in detail. It is shown that the scale function can be estimated using nonparametric regression based on the Box-Cox transformation of the data and it is closely related to its equivalent scale function, which is obtained in the local linear regression. Note that the difference between the scale function and the equivalent scale function is only a constant parameter, depending on the power transformation of the data under a weak moment condition, while, the parametric model can be also estimated under the weak moment condition after removing the long term trend component. An iterative plug-in

(IPI, Gasser et al. 1991) algorithm is developed for the bandwidth, the power parameter and the lag-window estimator c_f selection. Following Beran and Feng (2002), the initial bandwidth is selected by exponential inflation method (EIM) as $b_0 = T^{-5/7}$ instead of that with the power minus one of the sample size (Gasser et al., 1991) to minimize the rate of mean square error (MSE). In the c_f estimation, the spectral density at the origin is of great interest and the data-driven algorithm (Bühlmann, 1996; Feng and Gries, 2017; Feng et al., 2019) are employed according to the Bartlett-window, leading to an optimal c_f choice rather than manual input. Meanwhile, the power transformation parameter is also estimated in the data-driven algorithm with various criteria, such as the maximum likelihood estimation (MLE, Box and Cox, 1964), the normality test (Jarque-Bera test, JB; Shapiro-Wilk test, SW) and the quantile-quantile regression (QQr). A block bootstrap method is put forward to the power transformation parameter confidence interval (CI) estimation without sample distribution assumptions due to its dependence. The simulation results show that the absolute values of the power parameter are far smaller than one, satisfying the requirement of the weak moments conditions in real financial markets and the properties of the power parameter depend on the considered sample classes, e.g. for trading volume, the power parameter is about zero, indicating a possible logarithmic data transformation. Finally, the selection of the above parameters are implemented in the same IPI procedure and all the parameters can achieve a convergence value in a few IPI steps. The applications to the non-negative data samples show that the algorithm is feasible.

The chapter is organized as follows. In Section 6.2, the model is interpreted. Section 6.3 proposes the semiparametric estimation and the properties of the estimators. The data-driven algorithms are raise in Section 6.4. Data examples in Section 6.5 show the proposal works well in real financial markets. Section 6.6 concludes the chapter. The proof of some results is provided in the appendix.

6.2 The model

A general SemiMEM is proposed to analyze the nonnegative financial data. To reduce the moment condition, the Box-Cox transformation is also introduced.

6.2.1 The SemiMEM model

Let $x_1, ..., x_n$, denote the observed non-negative observations. The MEM model was introduced by Engle (2002) for modeling positive time series. The MEM class is a wide class of models including the ARCH and GARCH family as a special case. Numerous parameterizations for the expected variables are proposed and studied in the literature. In this section, the MEM will be generalized to a semiparametric class by introducing a smooth mean function into the parametric MEM model so that slowly changing dynamics caused by the economic environment can be modeled. This leads to the conditional distribution defined by

$$X_t = m(\tau_t)\psi_t\eta_t, \tag{6.1}$$

where $\tau_t = t/n$ is the rescaled time, $m(\cdot) > 0$ is a smooth trend, which is the localized unconditional mean function or the scale function in X_t, $\psi_t \geq 0$ denote the conditional mean, $\eta_t \geq 0$ are i.i.d. innovations with unit mean and

$$x_t = \psi_t\eta_t \tag{6.2}$$

is the descaled stationary process. Throughout this chapter, the notation of the condition, i.e. the past information set, $\mathcal{F}_{t-1}$ will be omitted for simplicity. The use of re-scaled time $\tau_t \in [0, 1]$ is a standard technique in nonparametric regression with time series errors. Due to $E(x_t|\mathcal{F}_{t-1}) = \psi_t E(\eta_t|\mathcal{F}_{t-1}) = \psi_t$, it is obvious that ψ_t is the conditional mean. To ensure that model (6.1) is uniquely defined, it is assumed that $E(\psi_t) = 1$. However, it is not necessary and the quantities $E(\eta_t)$ and $E(\psi_t)$ may also be determined by the situation under consideration or by the estimation procedure used. This model will be called a varying scale MEM model (VSMEM), which is also indeed the SemiMEM model (Feng (2014),

Feng and Zhou(2015)). Equation (6.1) defines indeed a sequence of models. The process $\{X_t\}$ is nonstationary unless $m(\cdot)$ is constant. But it can be shown that the SemiMEM model is locally stationary following Dahlhaus (1997). Our purpose is to develop consistent data-driven estimators of $m(\cdot)$ under certain moment condition $E(X_t^k) < \infty$, where $k > 0$ is a real number, e.g. $k = 2$. Then an approximation of the descaled stationary process x_t can be achieved which can be further studied following some known proposals in the literature. We will see that the use of some suitable power transformation X_t^λ works well, where with $|\lambda| \leq k/4$ is an exponent. Note that X_t^λ belongs to the SemiMEM class.

A closely related class of models is the general semiparametric GARCH framework proposed by Feng (2004). The definition of such a model in the second order sense is given by

$$r_t = s(\tau_t)\sqrt{h_t}\varepsilon_t, \tag{6.3}$$

where $s^2(\cdot) > 0$ is the variance of r_t, h_t is the conditional variance and ε_t are i.i.d. $N(0,1)$ innovations. The descaled stationary process $\xi_t = \sqrt{h_t}\varepsilon_t$ stands for an ARCH type model. It is assumed that $E(h_t) = 1$ to ensure that the model is well defined. Here $s^2(\cdot)$ is the scale function in r_t^2 and $s(\cdot)$ is the scale function in r_t. Due to the assumptions $E(\varepsilon_t^2) = 1$ and $E(h_t) = 1$, we have $E(r_t^2) = s^2(\tau_t)$. But $s(\tau_t)$ is not the mean of $|r_t|$. We will see that the difference between $s(\tau_t)$ and $E(|r_t|)$ is a constant factor depending on the distribution of ξ_t. Model (6.3) will be called a varying scale GARCH model (VSGARCH, also SemiGARCH). A nonparametric trend in the mean or a parametric regression for the mean based on some exogenous variables can also be included in (6.3). Assume again that $E(r_t^k) < \infty$ with e.g. $k = 4$. The scale function $s(\cdot)$, up to a constant factor, can be estimated consistently from $|r_t|^\lambda$ by some data-driven nonparametric regression algorithm, provided $0 < \lambda \leq k/4$ is used. Again, $|r_t|^\lambda$ also belongs to the SemiMEM class.

We see, $m(\cdot)$ in (6.1) and $s(\cdot)$ in (6.3) can be estimated in the same way. And the conditional mean ψ_t in (6.1) can also be modeled following the idea of ARCH and GARCH models for the conditional variance h_t in (6.3) (Engle, 2002). But the SemiGARCH and SemiMEM classes do not coincide with each other. On

the one hand, the signs of the observations of a SemiGARCH process may play an important role theoretically and in practice. Hence an original SemiGARCH process is not a member of the SemiMEM class. On the other hand, if the original process is non-negative, e.g. a duration process, it is of course not a member of the SemiGARCH class.

The descaled x_t in (6.2) process can be modeled following any suitable parameterization. Different parameterizations will lead to different semiparametric models. For instance, if X_t is a duration process, ψ_t can then be modeled by the well known ACD model introduced by Engle and Russell (1998) with

$$\psi_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{q} \beta_j \psi_{t-j}, \tag{6.4}$$

where p and q are the orders, $\alpha_0 > 0$, $\alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q \geq 0$ are unknown parameters such that $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. Equations (6.1) and (6.4) together define a semiparametric varying scale ACD model (VSACD, also SemiACD).

The overall mean in the SemiACD model can be thought of as a weighted sum of the unconditional local mean $m(\tau_t)$, the last q conditional means and the last p observations, which reflect long run, middle term and short term dynamics in the mean of X_t. Their weights are $1 - \sum \alpha_i - \sum \beta_j$; $\beta_1, ..., \beta_q$; and $\alpha_1, ..., \alpha_p$, respectively. Due to the restriction of the conditional mean, the scale parameter, α_0, is no more a free parameter, because it holds $\alpha_0 = 1 - \sum \alpha_i - \sum \beta_j$. That is α_0 in the conditional mean of the rescaled process is itself the weight for the unconditional local mean. The difference between the ACD and the SemiACD is just that the unconditional mean in the former is a constant but it is a smooth nonparametric function in the latter. Moreover, let $m_t^* = m(\tau_t)\psi_t$, we have

$$m_t^* = \alpha_0(\tau_t) + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j m_{t-j}^* + O(n^{-1}), \tag{6.5}$$

where $\alpha_0(\tau_t) = \alpha_0 m(\tau_t)$ is a time-varying parameter. We see, the SemiACD model is approximately an ACD model with a time-varying scale parameter.

In the remaining part of the chapter, only the estimation of $m(\cdot)$ in (6.1)

and the estimation of the unknown parameters in (6.4) based on the descaled data will be discussed in detail and the estimation of $s(\cdot)$ in (6.3) and of the unknown parametric parameters can be done following the same procedure. The scale function $m(\cdot)$ will be first estimated without using any further parametric assumption. It allows further parameter estimation under different, not only the ACD, specifications.

6.2.2 The Box-Cox SemiMEM model

The discussion of the scale function $m(\cdot)$ and the conditional mean ψ_t in the SemiMEM model are always considered. In the section, we study the estimation of the slow scale change based on the Box-Cox power transformation x_t^λ for some $|\lambda| < 1$. Here a nonparametric estimator of the scale function in x_t will be estimated first. Then the back-transformed estimator will be used. We will see that the resulting estimator is usually not a consistent estimator of $m(\cdot)$ but of some equivalent scale function.

In this section, any function of the form $\tilde{m}(\tau_t) = C \cdot m(\tau_t)$ with $C > 0$ will be called an equivalent scale function. Because $\tilde{x}_t = X_t/\tilde{m}(\tau_t)$, then $\tilde{x}_t = C^{-1}x_t$ is a stationary process having the same properties as x_t but with a different scale parameter. Hence the resulting estimator based on x_t can be used to remove the effect of the slowly changing scale in X_t. There are different further transformations, which can be used for estimating an equivalent scale function. The power transformations (or equivalently the Box-Cox transformations with $|\lambda| < 1$) are just the most simple examples. For a more general description on this point, we refer the reader to Eagleson and Müller (1997).

We see, the development of a consistent data-driven estimator is always possible based some power transformation. If it is assumed that $E(X_t) < \infty$, the maximal allowed λ is $1/4$. For conducting maximum likelihood estimators of the unknown parameters, the condition $E(X_t^2) < \infty$ was indicated by Lee and Hansen (1994) and Engle and Russell (1998). Now, $|\lambda| \leq 1/2$ can be used. Assume that $E(X_t^k) < \infty$. Let $\delta = \min(k/4, 1)$ for some $|\lambda| \leq \delta$. We have the Box-Cox SemiMEM model

as

$$X_t = m_\lambda(\tau_t)x_{t,\lambda}, \tag{6.6}$$

where $x_{t,\lambda}$ is stationary with $E(x_{t,\lambda}^\lambda) = 1$. If $x_{t,\lambda}$ satisfies a specific parameterization, e.g. an ACD process, model (6.6) becomes a Box-Cox SemiACD (or a Box-Cox VSACD) process. Obviously, the stationary process $x_{t,\lambda} = c_\lambda^{-1/\lambda}x_t$ and the equivalent scale function $\tilde{m}(\tau_t) = m_\lambda(\tau_t) = C_\lambda^{1/\lambda} \cdot m(\tau_t)$. The value of C_λ is determined by λ and the marginal distribution of x_t. Under the assumption $E(x_t) = 1$ as used in (6.2) we have $C_1 \equiv 1$. This shows again why commonly proposed estimators of $m(\cdot)$ are based on x_t. However, model (6.1) can also be defined according to the λ_0-th moments of the process for some $|\lambda_0| \leq k/4$. That is, we can assume that $E(x_t^{\lambda_0}) = 1$ and $E(\psi_t^{\lambda_0}) = 1$. This implies that $E(X_t^{\lambda_0}) =: m_{\lambda_0}(\tau_t)$ is the scale function in $X_t^{\lambda_0}$. Under this definition we have $C_{\lambda_0} \equiv 1$. The Box-Cox SemiACD model in the section has a close relation with the parametric Box-Cox ACD model, i.e. if the power parameter is nonzero, a varying scale Box-Cox ACD is considered and if the power parameter is zero, then the model reduces to a logarithmic form.

6.3 The model estimation and properties

The Box-Cox SemiMEM models can be estimated using a semiparametric procedure. The scale function can be estimated and removed under very weak moment condition based on suitable power transformation of the data. The conditional variance can be analyzed using any parametric model.

6.3.1 Estimation of $m_\lambda(\tau_t)$

In Efromovich (1999), the scale function can be estimated by a general nonparametric regression process. Following the proposal, Equation (6.6) can be written as a nonparametric regression model

$$X_t^\lambda = g_\lambda(\tau_t) + g_\lambda(\tau_t)\zeta_{t,\lambda}, \tag{6.7}$$

where $g_\lambda(\tau_t) = c_\lambda m^\lambda(\tau_t) = [m_\lambda(\tau_t)]^\lambda$ and $\zeta_{t,\lambda} = x_{t,\lambda}^\lambda - 1$ with $E(\zeta_{t,\lambda}) = 0$ and $\text{var}\,(\zeta_{t,\lambda}) = 1$. Hence $g_\lambda(\cdot)$ can be estimated for instance by kernel (Feng and Heiler 1998 and Feng, 2004) or local liner regression (Fan and Yao, 1998).

Since by definition the observations x_t are non-negative, kernel estimates of $g_\lambda(\tau_t)$ based on non-negative kernels are always positive. However, at boundary points (i.e. $\tau_t < b$ or $\tau_t > 1 - b$), the rate of convergence of kernel estimators is lower than in the interior unless boundary kernels are used. Alternatively, one may use local linear estimates which have the same rate of convergence for all $\tau_t \in [0, 1]$. However, boundary kernels or local linear regression lead to estimates that may be negative at boundary points. We therefore propose to use the estimator $\hat{g}_\lambda(\tau_t) = |\tilde{g}_\lambda(\tau_t)|$ where $\tilde{g}_\lambda(\tau_t) = \hat{a}_0(x)$ is a local linear estimate obtained by minimizing

$$Q(a_0, a_1) = \sum_{t=1}^{T} \left\{ X_t^\lambda - a_0(\tau) - a_1(\tau)(\tau_t - \tau) \right\}^2 K\left(\frac{\tau_t - \tau}{b}\right) \tag{6.8}$$

with K being a symmetric non-negative kernel function and $0 < b < \frac{1}{2}$ the bandwidth. Using $\hat{g}_\lambda(\tau_t)$ instead of $\tilde{g}_\lambda(\tau_t)$ can be justified as follows. First, note that $\hat{g}_\lambda(\tau_t) = \tilde{g}_\lambda(\tau_t)$, if $\tilde{g}_\lambda(\tau_t) \geq 0$. If $\tilde{g}_\lambda(\tau_t) < 0$, we have $|\tilde{g}_\lambda(\tau_t) - g_\lambda(\tau_t)| = g_\lambda(\tau_t) + |\tilde{g}_\lambda(\tau_t)|$ and $|\hat{g}_\lambda(\tau_t) - g_\lambda(\tau_t)| = ||\tilde{g}_\lambda(\tau_t)| - g_\lambda(\tau_t)| < g_\lambda(\tau_t) + |\tilde{g}_\lambda(\tau_t)|$. Hence $E[(\hat{g}_\lambda(\tau_t) - g_\lambda(\tau_t))^2] \leq E[(\tilde{g}_\lambda(\tau_t) - g_\lambda(\tau_t))^2]$. That is the MSE of $\hat{g}_\lambda(\tau_t)$ is no larger than that of $\tilde{g}_\lambda(\tau_t)$. Moreover, the possibility of negative values is a finite sample problem and can only occur at boundary points, because in the interior the weights of a local linear estimate are exactly equal to the kernel weights and are hence non-negative. Finally, even at a boundary point, $\hat{g}_\lambda(\tau_t)$ and $\tilde{g}_\lambda(\tau_t)$ coincide with asymptotic probability one, as shown in Zhang et al. (2017).

Let that $\hat{m}_\lambda(\cdot)$ be a consistent estimator of $m_\lambda(\cdot)$, we can obtain $\hat{x}_{t,\lambda}^* = x_{t,\lambda}/[\hat{m}_\lambda(\tau_t)]^{1/\lambda}$ is an approximation of the stationary process $x_{t,\lambda}^*$. Define $Z_{t,\lambda} = X_t^\lambda$, then $\hat{z}_{t,\lambda}^* = Z_{t,\lambda}/\hat{m}_\lambda(\tau_t)$ is its approximate estimation of $x_{t,\lambda}$ with Box-Cox transformation. A suitable parametric model can be fitted to either $\hat{x}_{t,\lambda}^*$ or $\hat{z}_{t,\lambda}^*$.

Assume that the moments condition $E(X_t^k) < \infty$ and $E(x_t^k) = 1$ hold, for $k \neq 0$. The scale function $m(\tau_t)$ in the Box-Cox SemiMEM model is estimated by its equivalent scale $g_\lambda(\tau_t)$ as $\hat{m}(\tau_t) = \hat{C}_\lambda^{-1/\lambda} \cdot \hat{g}_\lambda(\tau_t)$, where $\hat{C}_\lambda^{-1/\lambda} = [\frac{1}{n}\sum_{t=1}^{n} \hat{x}_{t,\lambda}^k]^{1/k}$.

Zhang et al. (2017) discussed the estimation under a Box-Cox SemiGARCH framework and the constant parameter of the equivalent scale function is regarded related to the mean of the stationary second order stationary process. Following the definition of Box-Cox SemiMEM model, for $\lambda \neq 0$, we can conclude that the scale function is estimated by its equivalent scale function as $\hat{m}_\lambda(\tau_t) = c_\lambda^{1/\lambda}\hat{m}(\tau_t)$. Then,

$$\begin{aligned} X_t &= m_\lambda(\tau_t) \cdot x_{t,\lambda} \\ &= c_\lambda^{1/\lambda} \cdot m(\tau_t) \cdot x_{t,\lambda} \\ &= m(\tau_t) \cdot x_t, \end{aligned} \tag{6.9}$$

it is obvious that $x_{t,\lambda} = c_\lambda^{-1/\lambda} x_t$. Under the assumptions of Proposition, if $E(x_t^k) = 1$ and $k \neq 0$ hold, the power k-th expectation of the transformed non-negative variable is

$$\begin{aligned} E(x_{t,\lambda}^k) &= E[(c_\lambda^{-1/\lambda} \cdot x_t)^k] \\ &= c_\lambda^{-k/\lambda} \cdot E(x_t^k) \\ &= c_\lambda^{-k/\lambda}. \end{aligned} \tag{6.10}$$

Also, the constant for the equivalent scale function is

$$c_\lambda^{-1/\lambda} = [E(x_{t,\lambda}^k)]^{1/k}. \tag{6.11}$$

Obviously, if the stationary process follows a GARCH process and a second order transformation, the constant of the equivalent scale function is

$$\begin{aligned} c_\lambda^{-1/\lambda} &= \sqrt{E(\hat{x}_{t,\lambda}^2)} \\ &= \sqrt{\frac{1}{n}\sum_{t=1}^{n} \hat{x}_{t,\lambda}^2}, \end{aligned} \tag{6.12}$$

and for the ACD model, under the assumption $E(x_t) = 1$, the constant reads as

$$\begin{aligned} c_\lambda^{-1/\lambda} &= E(\hat{x}_{t,\lambda}) \\ &= \frac{1}{n}\sum_{t=1}^{n} \hat{x}_{t,\lambda}, \end{aligned} \tag{6.13}$$

which is indeed the mean of the stationary process.

6.3.2 Properties of $\hat{g}_\lambda(\tau_t)$

The asymptotic properties of $\hat{g}_\lambda(\tau_t)$ are similar to those of a local linear estimator of the mean function in the presence of heteroskedastic time series errors and are closely related to known results on nonparametric regression with dependent errors (see e.g. Altman, 1990, Hart, 1991 and Beran and Feng, 2002). The results given in this section do not depend on any parametric specification of $x_{t,\lambda}$.

Note that a local linear estimator at point x generates an equivalent kernel $K_\tau(u)$ which is the same as $K(u)$ for $b \leq \tau_t \leq 1-b$ and equal to a boundary kernel at boundary points. To reduce the variance, we will use a varying bandwidth b_τ at boundary points, such that the length of the window is always $2b$. For kernel functions K and K_τ, define $R(K) = \int K^2(u)du$, $I(K) = \int u^2K(u)du$, $R(K) = \int K^2(u)du$ and $I(K) = \int u^2K(u)du$. Furthermore, let $\gamma(k) = cov(x_{t,\lambda}, x_{t+k,\lambda})$. Assume that $\gamma(k)$ are absolutely summable and c_f denotes the value of the spectral density of $x_{t,\lambda}$ at the origin $\lambda = 0$. Under the assumptions A1 through A5 stated in the appendix the following holds, as $n \to \infty$, $b \to 0$ and $nb \to \infty$:

The bias and variance of $\hat{g}_\lambda(\tau_t)$ are

$$B[\hat{g}_\lambda(\tau_t)] = E(\hat{g}_\lambda(\tau_t)) - g_\lambda(\tau_t) = \frac{1}{2}b_\tau^2 g''_\lambda(\tau_t)I(K_\tau) + o(b_\tau^2), \tag{6.14}$$

and

$$var(\hat{g}_\lambda(\tau_t)) = \frac{2\pi c_f g_\lambda^2(\tau_t)R(K)}{nb_\tau} + o\left(\frac{1}{nb_\tau}\right) = \frac{V_\tau}{nb_\tau} + o\left(\frac{1}{nb_\tau}\right), \tag{6.15}$$

where $V_\tau = 2\pi c_f g_\lambda^2(\tau_t)R(K)/(nb_\tau)$.

Based on the bias and variance of the estimated equivalent scale function, the mean integrated squared error of $\hat{g}_\lambda$, $\text{MISE}(\hat{g}_\lambda)=\int_0^1 \{\hat{g}_\lambda(\tau) - g_\lambda(\tau)\}^2 d\tau$, is

$$\text{MISE}(\hat{g}_\lambda) = b^4 \frac{\int [g''_\lambda(\tau)]^2 d\tau I^2(K)}{4} + \frac{2\pi c_f \int g^2_\lambda(\tau) d\tau R(K)}{nb} + \max\left\{ o(b^4), o\left(\frac{1}{nb}\right)\right\}. \tag{6.16}$$

Then, the asymptotically optimal bandwidth, which minimizes the dominating part of the MISE is given by

$$b_A = \left(2\pi c_f \frac{R(K)}{I^2(K)} \frac{\int g^2_\lambda(\tau) d\tau}{\int [g''_\lambda(\tau)]^2 d\tau}\right)^{1/5} T^{-1/5} = \left(2\pi c_f \frac{R(K)}{I^2(K)} \frac{I(g^2_\lambda)}{I[(g''_\lambda)^2]}\right)^{1/5} T^{-1/5}, \tag{6.17}$$

where $I(g^2_\lambda) = \int g^2_\lambda(\tau) d\tau$ and $I([g''_\lambda]^2) = \int [g''_\lambda(\tau)]^2 d\tau$.

Further, if a bandwidth $b = o(b_A)$ is used, the bias is asymptotically negligible and

$$\sqrt{nb}[\hat{g}_\lambda - g_\lambda] \xrightarrow{\mathcal{D}} N(0, V), \tag{6.18}$$

where V is as $V = 2\pi c_f g^2_\lambda R(K)/(nb)$.

Note that both the kernel $K(u)$ and the bandwidth b_τ in the bias and variance of $\hat{g}(\tau_t)$ depend on τ_t. But the effect of boundary points on the MISE of $\hat{g_\lambda}$ is asymptotically negligible. Thus, the asymptotically optimal global bandwidth can be calculated using the MISE over the whole interval $\tau_t \in [0, 1]$. Eq. (6.17) provides the basis for developing a plug-in bandwidth selector. The difference between the formula of the optimal bandwidth b_A here, compared to nonparametric regression with i.i.d. errors is that the two unknown constants c_f and $I(g_\lambda)$ are different. Here, the factor c_f measures the effect of the stationary time series errors on b_A. The constant $I(g_\lambda)$ is determined by the (deterministic) heteroskedasticity characterized by the scale function in (6.7). The result of Eq. (6.18) shows that $\hat{g}_\lambda$ is asymptotically unbiased and asymptotically normal if a bandwidth of a smaller order than b_A is used.

6.4 The data-driven algorithms

In this section, data-driven estimation algorithms of the bandwidth and the c_f selection are developed. Besides, a bootstrap method is proposed to simulate the confidence interval of the power parameter.

6.4.1 The bandwidth selection

In this section, we develop a data-adaptive bandwidth selector for the SemiACD model following the iterative plug-in method (IPI, see Gasser et al. 1991). For simplicity, global bandwidth selection will be considered. The plug-in bandwidth selector is based on an iteration algorithm where estimates of c_f, $I(g_\lambda^2)$ and $I([g_\lambda'']^2)$ are plugged into (6.17). The key point here is the estimation of $I([g_\lambda'']^2)$, because the estimation of c_f and $I(g_\lambda^2)$ are relatively easy. The IPI method has been extended successfully to nonparametric regression with time series errors (see e.g. Herrmann et al. 1992, Brockmann et al. 1993, Beran and Feng 2002 and Ghosh and Draghicescu 2002). We will, therefore, use this approach here.

Let b_{j-1} denote the bandwidth in the $(j-1)$th iteration. The IPI algorithm estimates $I([g_\lambda'']^2)$ in the jth iteration by $\hat{I}_j([g_\lambda'']^2) = n^{-1}\sum_{t=1}^{n}[\hat{g}_{\lambda_j}''(\tau_t)]^2$, where $\hat{g}_{\lambda_j}''(\tau_t)$ is estimated using a bandwidth b_{2j} obtained from b_{j-1} by a so-called inflation method. The word 'inflation' comes from the fact that b_{2j} is much larger than that the bandwidth for estimating m itself. Once the estimate $\hat{I}_j([g_\lambda'']^2)$ is calculated, it is inserted into (6.17) to calculate b_j in the jth iteration. This procedure is repeated until the selected bandwidth converges.

Two choices have to be made to define the algorithm. One needs to fix an initial bandwidth b_0, and a concrete inflation method has to be specified. Gasser et al. (1991) propose to set $b_0 = n^{-1}$. However, this bandwidth cannot be used for estimating $I(m)$ in the first iteration. Beran and Feng (2002) therefore suggest $b_0 = n^{-5/7}$ which is a very small bandwidth but satisfies the assumptions in Theorem 6.1. As it turns out, the final bandwidth is not sensitive to the choice of b_0. However, the number of required iterations depends heavily on b_0. A data-driven

possibility that reduces the number of iterations is $b_0 = \hat{b}_{\mathrm{CV}}$ where $\hat{b}_{\mathrm{CV}}$ is the bandwidth selected by the cross-validation criterion (Wahba and Wold 1975).

With respect to the inflation method, the so-called exponential inflation method proposed (EIM) in Beran and Feng (2002) with $b_{2j} = (b_{j-1})^{\alpha}$ (where $0 < \alpha < 1$) is well suited in the current context. We propose to choose $\alpha = 5/7$ which means that the rate of the MSE of $\hat{I}([g''_{\lambda}]^2)$ is minimized. Note that, in comparison, the original multiplicative inflation method (MIM) of Gasser et al. (1991) defined by $b_{2j} = b_{j-1}n^{1/10}$ does not work well in the current context. The main reason is that the rate of convergence of the estimated bandwidth is $O(n^{-1/5})$ which is clearly lower than the rate of convergence $O(n^{-2/7})$ achieved by the EIM method (see the theorem below). An additional problem with the MIM is that for large sample sizes T the inflation factor is too strong which often leads to poor final bandwidth. In each iteration, $I(g_{\lambda}^2)$ can be estimated from $\hat{g}_{\lambda}$ which is based on the previous bandwidth b_{j-1}. Since $\hat{b}_{\mathrm{CV}}$ is already a consistent estimator of b_A, using $b_0 = \hat{b}_{\mathrm{CV}}$ is likely to provide a good initial estimate of $I(g_{\lambda}^2)$.

The asymptotic performance of the selected $\hat{b}_A$ by the IPI algorithms in the following subsection is given by the following theorem. For this purpose, the following assumptions are required.

A1. The scale function $g_{\lambda}(\tau_t)$ is strictly positive, bounded, and at least twice continuously differentiable on [0, 1].

A2. The kernel $K(u)$ is a symmetric density with compact support [-1, 1].

A3. $\{x_t\}$ is a stationary ACD process defined by (6.2).

Theorem 6.1 *Under Assumptions A1 to A3 and the additional assumption that* $E(x_t^4) < \infty$*, we have*

$$\hat{b}_A = b_A[1 + O_p(n^{-2/7}) + O(n^{-1/3})]. \tag{6.19}$$

A sketched proof of Theorem 6.1 is given in the appendix. The $O_p(n^{-2/7})$ term in (6.19) is caused by the error in $\hat{I}([g''_{\lambda}]^2)$ and whereas the $O(n^{-1/3})$ term is due to the error in $\hat{c}_f$. If the parametric specification in (6.2) and (6.4) is used,

then the (relative) effect of the error in $\hat{c}_f$ will be of the order $O(n^{-2/5})$ (see e.g. Feng, 2004), which is smaller than $O(n^{-1/3})$. However, the $O(n^{-1/3})$ term is still asymptotically negligible compared to $O_p(n^{-2/7})$. Hence the asymptotic properties and in particular the rate of convergence of $\hat{b}_A$ are determined by those of $\hat{I}([g''_\lambda]^2)$. Note in particular that, although the bandwidth selection problem under model (6.7) is more complex, the rate of convergence of the selected bandwidth is the same as for the DPI (direct plug-in) bandwidth selector proposed by Ruppert et al. (1997) in the context of nonparametric regression with i.i.d. errors. Furthermore, b_A is not well defined, if $g_\lambda(\tau_t) \equiv g_0$, because $I([g''_\lambda]^2) = 0$. Nevertheless, the SemiACD model and the proposed algorithm are still applicable in this case. For instance, if y_t follows an MEM model, it can be shown that b_j converges to a nonzero constant as $j \to \infty$. Besides, it can be shown that $\hat{\theta}$ (obtained from $\hat{x}_t$) has the same asymptotic properties as $\tilde{\theta}$ (obtained from the MEM observations x_t), because $\hat{b} >> O_p(n^{-1/2})$. Moreover, suppose that no maximal number of iterations is fixed. Then $(nb_j)^{-1}$ is asymptotically of the order n^{-1} though $b_j < 1$. Therefore $\hat{g}_\lambda(\tau_t)$ is $\sqrt{n}$-consistent, with some loss in efficiency compared to a parametric estimator.

6.4.2 The c_f estimation

The remaining unknown quantity c_f can be estimated using any nonparametric estimator of the spectral density. The spectral density of $x_{t,\lambda}$ is

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_j(k) e^{ik\lambda}, -\pi \leq \lambda \leq \pi. \tag{6.20}$$

If $\lambda = 0$, we can obtain that $c_f = f(0)$. In the following, we will use the lag-window estimator with the Bartlett-window (see e.g. Priestley, 1981)

$$\hat{c}_f^j = \frac{1}{2\pi} \sum_{k=-K}^{K} w_k \hat{\gamma}_j(k) \tag{6.21}$$

to estimate the c_f. In the formula, $\hat{\gamma}_j(k)$ denotes the sample autocovariance at lag k calculated from the residuals in the $(j-1)$th iteration, $w_k = 1 - k/(K+1)$ and

$K << T$ is the bandwidth of the lag-window, included in $\hat{c}_f$.

The optimal choice of K is of the order $O(n^{1/3})$. To focus on the main results, we do not address the issue of data-driven algorithms for selecting K. In the applications we will use $K = [c_f n^{1/3}]$ with c_f selected in the same IPI algorithm. The proposed IPI data-driven algorithm, including the bandwidth selection with power transformation and the c_f estimation, reads as follows:

1. Select b_0 using the CV criterion ignoring correlations and changes in the scale. Set $j = 1$.

2. Select bandwidth correlations and changes in the scale, using J_1 IPI-iterations.

3. In the j-th iteration for $j \geq J_1$ carry out the following calculations:

   a) Estimate $\hat{g}_\lambda(\tau_t)$ with b_{j-1} and let $\hat{x}_{t,\lambda} = X_t/[\hat{g}_\lambda(\tau_t)]^{1/\lambda}$.

   b) Estimate $\hat{c}_f$ with the Bartlett-lag-window estimator.

      i) Set the starting Bartlett-window as $M_0 = [n/2]$, where $[\cdot]$ is the integer part.

      ii) Global estimation. Following Bühlmann (1996), estimate the integration of the first derivative $\int f^{(1)}(\lambda)d\lambda$ with the Bartlett-window width $K'_j = K_{j-1}/n^{2/21}$. Insert the estimates into the optimal global window width equation (Bühlmann, 1996, Eq. 5), then calculate K_j. Increase j by one and repeat the above procedures until the selected K converges or reach the maximal 20 iterations. Denote the selected optimal global window width as K_{Gl}.

      iii) Local estimation. Calculate $\int f^{(1)}(\lambda)d\lambda$ with the optimal local window width $K_{Lo} = K_{Gl}/n^{2/21}$ and insert the estimates into optimal local window width equation (Bühlmann, 1996) with $\lambda = 0$

$$K_{opt} = n^{-1/3}\left(\frac{\int_{-\pi}^{\pi}(f(\lambda)^2)d\lambda}{3\int_{-\pi}^{\pi} f^{(1)}(\lambda)^2 d\lambda}\right)^{1/3}. \tag{6.22}$$

      The finally selected Bartlett-window width is denoted as $\hat{K} = K_{opt}$.

c) Denote by $\hat{I}(g_\lambda)$ the estimate of $I(g_\lambda)$ obtained using b_{j-1} and by $\hat{I}(g''_\lambda)$ the estimate of $I(g''_\lambda)$ based on bandwidth $b_{2j} = b_{j-1}^{5/7}$. Consider the power transformation parameter in the range $[-1, 1]$ and let the power transformation parameter increase interval be 0.01. In the k-th iteration with $k > 1$, estimate the trend $\hat{g}_{\lambda_k}(\tau_t)$ using $\hat{\lambda}_{k-1}$. Remove the trend and obtain the optimal transformation parameter λ_k by the MLE, JB, SW and QQr criteria. Increase k by one and repeat the previous steps until reach the convergence or the maximal number of iterations.

d) Improve b_{j-1} by

$$b_j = \left(2\pi \hat{c}_f \frac{R(K)}{I^2(K)} \frac{\hat{I}(g_\lambda)}{\hat{I}(g''_\lambda)} \right)^{1/5} n^{-1/5}. \tag{6.23}$$

e) Increase j by one and repeatedly carry out a) to c) until convergence or a given maximal number of iterations has been reached.

The finally selected bandwidth $\hat{b}_A$ is obtained in the last iteration of Step 3. In the IPI procedures of the bandwidth, c_f and the power parameter estimation, we find that if the selected statistics converge, the values usually are not affected by the initial inputs. Hence, the proposed IPI algorithm can be carried out without starting restrictions. In the section for simplicity, the Bartlett-window is applied in the c_f selection algorithm. Besides, Bühlmann (1996) also discussed a C^2-window for the optimal window width. Finally, Feng and Gries (2017) introduced that the estimation quality of c_f can be improved if the optimal bandwidth for calculating the $x_{t,\lambda}$ is considered and vice versa.

6.4.3 The confidence interval simulation of λ

In the previous section, we obtain the stationary time series $x_{1,\lambda}, \ldots, x_{t,\lambda}$ by removing the time-varying scale function. For estimating the power transformation parameter λ, we propose to construct an estimator $\hat{\lambda}$ based on the descaled time series via the MLE, JB, SW and QQr criteria. The confidence interval of λ at a given confidence level with MLE has been well discussed. However, the confi-

dence intervals depending on the other criteria, are unknown in practice and often very complicated. Bootstrap methods provide a general solution for estimating the confidence interval of λ without the consideration of the time series model assumptions.

The bootstrap was published by Efron (1979) for i.i.d. random variables as a simulation method using the resampling technique to estimate the variables distributions and the statistic inference, such as the bias, the variance, the confidence intervals, the reject probability in a hypothesis test, etc. The advantage of the bootstrap method is that the simulation algorithm requires no distribution or parametric assumption of the under analysis data set or process. The bootstrap provides information about the whole sampling distribution and performs computational efficient rather than the other resampling technique– jackknife (Tukey, 1958), recognized as the "leave-one-out" method.

For the time series, however, the dependent and correlated data are not suitable to directly apply the bootstrap algorithm. As discussed by Künsch (1989), Hall (1992), Liu and Singh (1992) and Politis and Romano (1993), for the dependent time series data, the block bootstrap method is the potential solution to estimate the unknown distribution by dividing the data into several blocks, to hold the original time series dependence structure within blocks. Because the asymptotic properties of the estimator may be affected by the block selection, i.e. the dependence of the resampling time series is always regarding to the randomly selected blocks. Thus, a modification bootstrap– moving block bootstrap (MBB) was proposed by selecting the optimal block length, also recognized as the overlapping block bootstrap, which preserves the data structure of the original series in each formed block. In the MBB, the length of the block is

$$l = n_o/N_{bl}, \tag{6.24}$$

where n_o is the length of the original time series and N_{bl} is the number of the resampling blocks and according to the MBB, the independence of the l subsampling is for sure. If dependent data is considered, then an unnecessary requirement

is

$$l \to \infty \text{ and } N_{bl}^{-1} l \to 0 \text{ as } N_{bl} \to \infty. \tag{6.25}$$

Obviously, the selection of the block length is the most concerned problem in the block bootstrap (Efron and Tibshirani, 1993). As indicated by Bühlmann and Künsch (1999) and Bühlmann (2002), they proposed the bootstrap variance is equivalent to a lag weight estimator of the spectral density of the bootstrapped variable's influence function (IF, Hampel et al., 1986) at the interesting origin and the block length is obvious obtained as inverse of the bandwidth. Furthermore, they also developed a data-driven algorithm, suggesting a two-step procedure selecting the optimal block length in the blockwise bootstrap and in this algorithm, the IF is estimated first, then optimal block length is obtained as the estimated value of the lag weight spectral density with certain lag window (e.g. Bartlett window) at frequency zero. Following the algorithm, the optimal block length is approximate to the cubic root of the sample size. Politis and White (2004) discussed a data-based block length selection algorithm and found that the optimal block length order for the stationary bootstrap is also one third.

Besides, we still consider a model-based bootstrap method (such as autoregressive bootstrap (ARB), Efron, 1979), by removing the nonstationary trend and building up the i.i.d. error terms under the MEM model assumption. Due to the model-based bootstrap depending on not only the parameters in models but also the identified structure with the original data. Obviously, the model-based resampling is greatly affected by both the model and its structure and the asymptotic properties of original data can not be correctly revealed if the model misspecification occurs, leading to the inconsistent between the built-up i.i.d data and the original ones. In the section, the MBB methodology works very well in practice, however, the model-based idea seems not.

There are several procedures to calculate the confidence intervals of the resampling data sets, such as the percentile method, the bootstrap-t method, the bias-corrected (BC) method, the bias-corrected and accelerated (BCa) method, the approximate bootstrap confidence (ABC) method and calibration. For simplicity, we consider only the ordinary percentile method in the chapter.

In the percentile approach, we consider the $\alpha/2$ and $(1-\alpha)/2$ percentiles of the bootstrap distribution of $\hat{\lambda}$. Let the number of the bootstrap replications of $\hat{\lambda}$ be N_B. Then, the order statistic of λ

$$\hat{\lambda}_1^* \leq \cdots \leq \hat{\lambda}_{k_L}^* \leq \cdots \leq \hat{\lambda}_{k_U}^* \leq \cdots \leq \hat{\lambda}_{N_B}^*, \tag{6.26}$$

and the bootstrap confidence interval (BCI) at the $100(1-\alpha)\%$ is

$$\lambda \in [\hat{\lambda}_{k_L}^*, \hat{\lambda}_{k_U}^*], \tag{6.27}$$

where $k_L = [\frac{\alpha}{2}(N_B+1)]$ and $k_U = [(N_B+1) - k_L]$

The percentile method gets rid of the assumption limitation of the bootstrap variable in theory, however, if the resampling replication number is too small, the simulation may not perform well, requiring the percentiles corrections, new resampling histogram assumptions, bias-correction factors and acceleration parameters.

6.5 The empirical examples

The selection of the $\hat{\lambda}$ is an important issue regarding to the model selection in the risk management. In the section, we apply the proposal to the non-negative transaction data (MD, AR, VO and TR) of Siemens (SIE) and Deutsche Bank (DBK) from 2000 to 2013. The MD, VO and TR data are organized as equidistant daily data from the intraday high-frequency transaction records, meanwhile, the AR data is calculated as the absolute returns of the daily stock close price.

From Fig. 6.1 to Fig. 6.8, the histograms of the selected $\hat{\lambda}$ with the MLE, JB, SW and QQr are displayed respectively. Because the sample sizes of the considered data sets are about 3500, we apply the length of 16 (the approximate cubic root of the size of the observation) as the optimal block length in the bootstrap procedures.

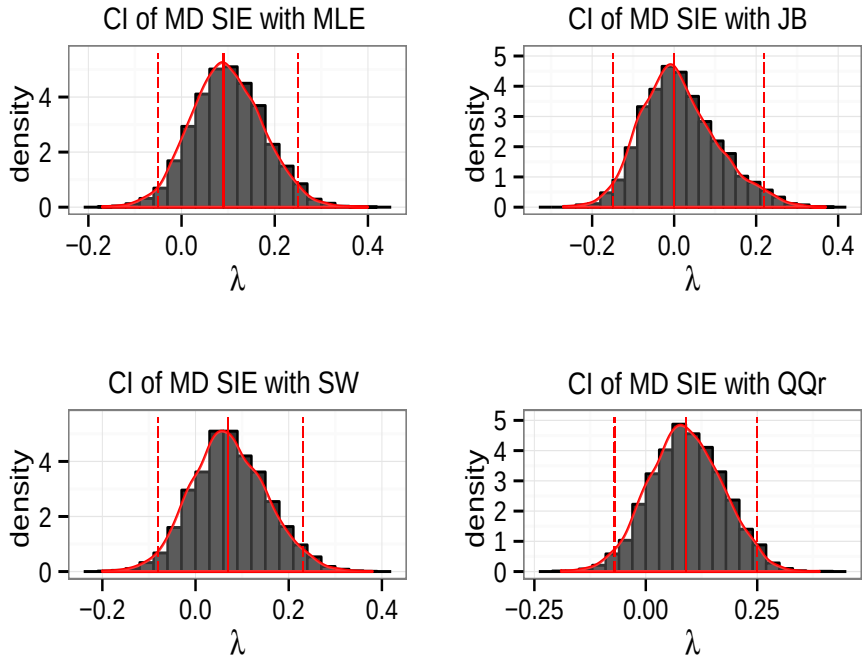

Figure 6.1: Simulated λ confidence interval of the SIE mean duration

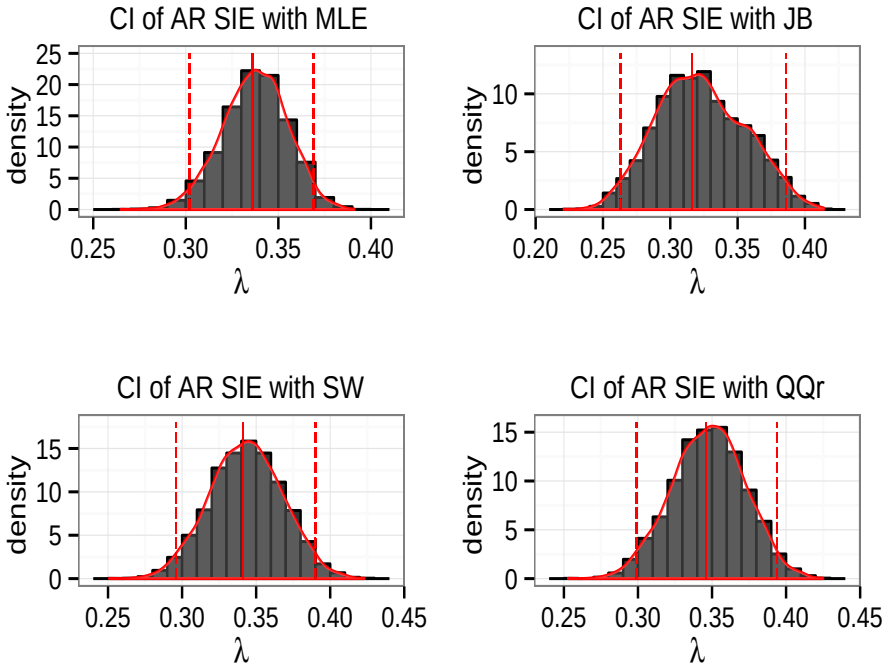

Figure 6.2: Simulated λ confidence interval of the SIE absolute returns

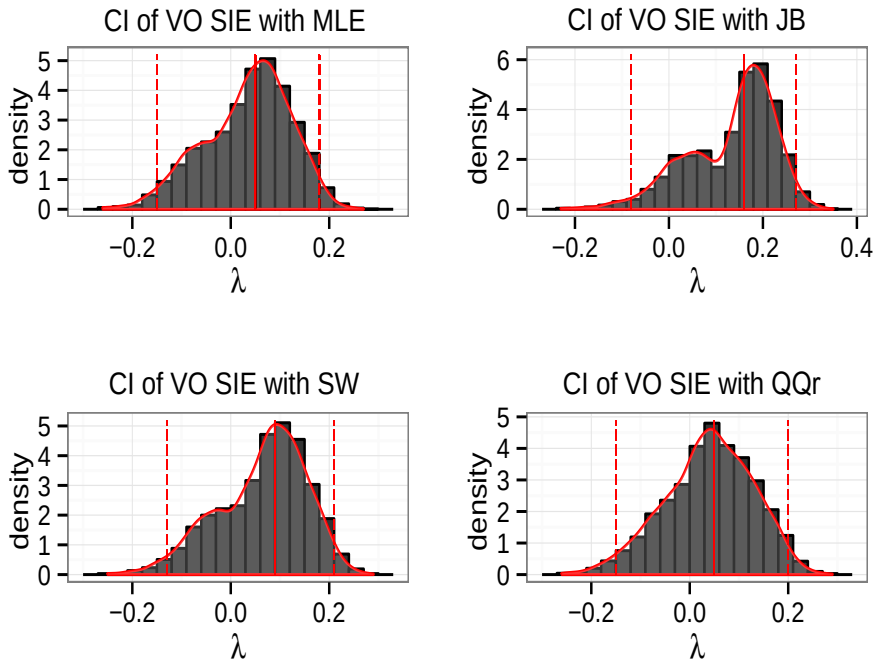

Figure 6.3: Simulated λ confidence interval of the SIE trading volume

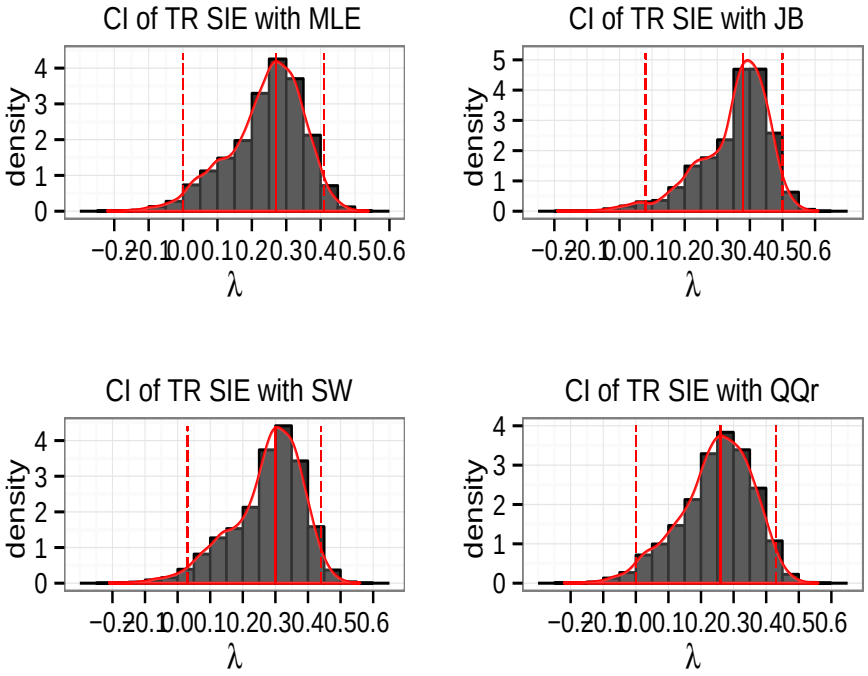

Figure 6.4: Simulated λ confidence interval of the SIE trading numbers

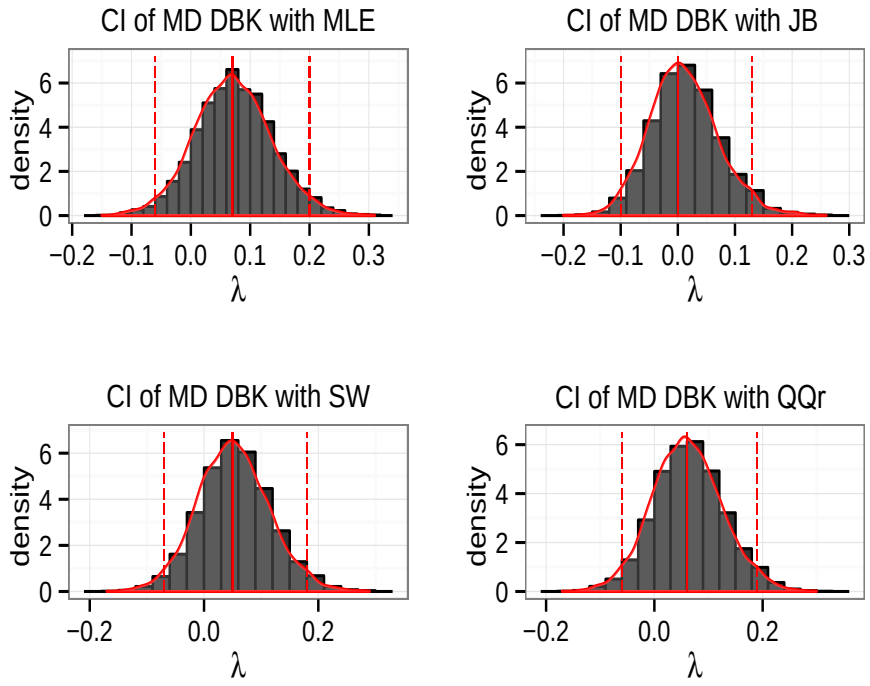

Figure 6.5: Simulated λ confidence interval of the DBK mean duration

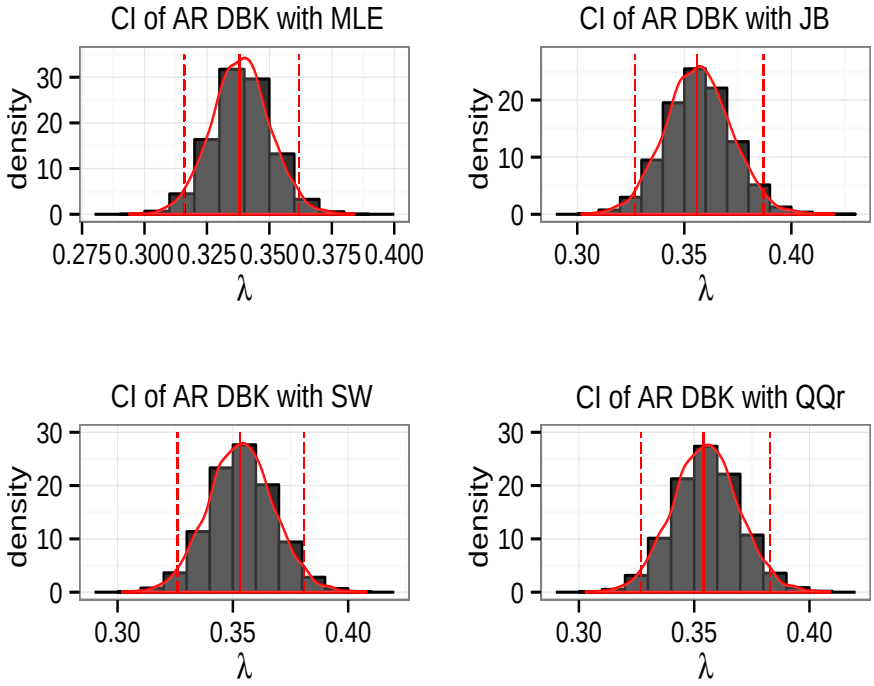

Figure 6.6: Simulated λ confidence interval of the DBK absolute returns

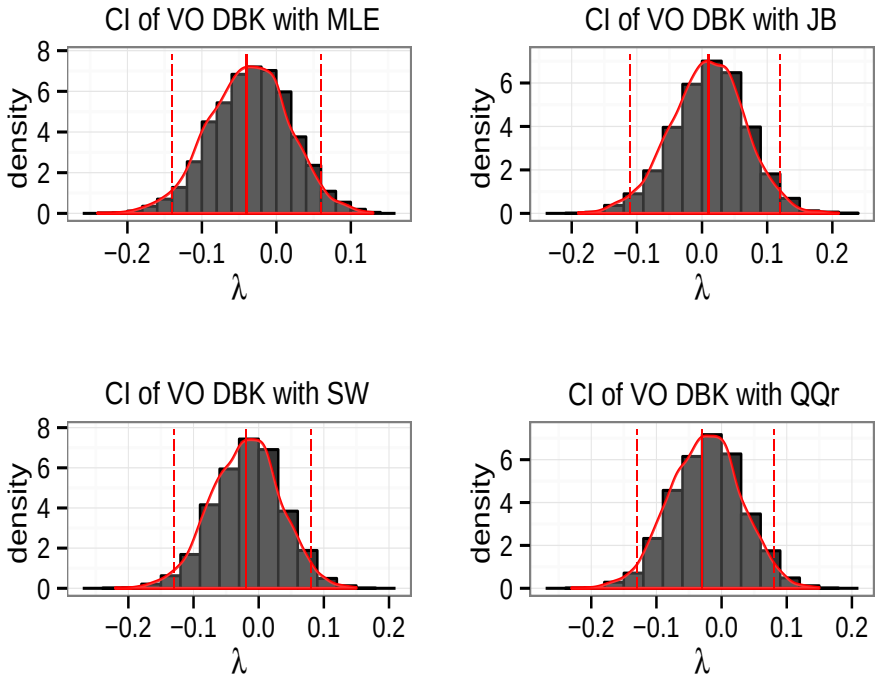

Figure 6.7: Simulated λ confidence interval of the DBK trading volume

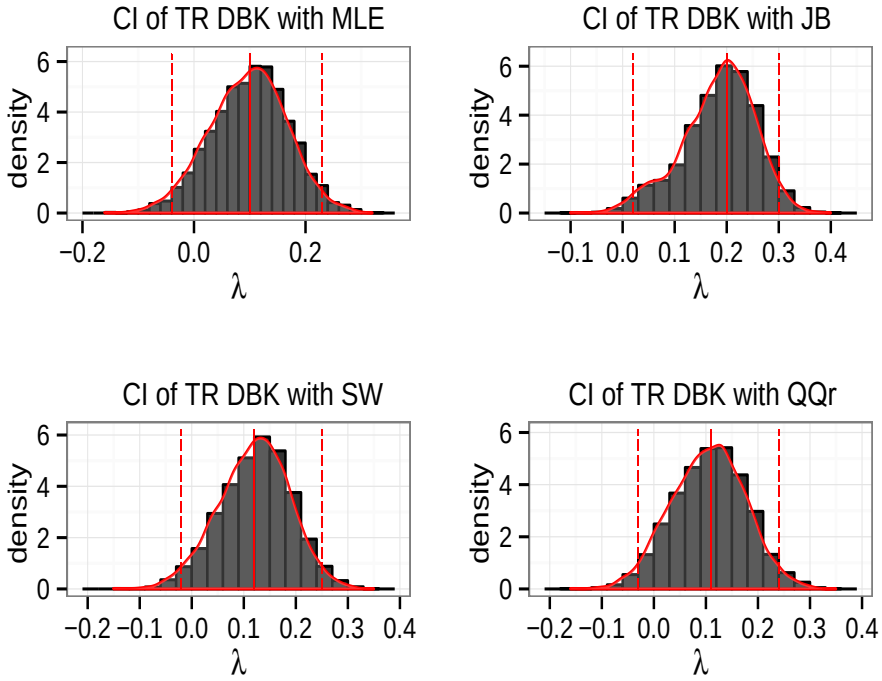

Figure 6.8: Simulated λ confidence interval of the DBK trading numbers

In the figures, we find that the $\hat{\lambda}$ of MD is very close to 0 in all cases and the confidence interval also covers the origin. In Table 6.1, the CIs of SIE's MD are $[-0.05, 0.25]$ with the MLE, $[-0.08, 0.23]$ with the SW, $[-0.15, 0.22]$ with the JB and $[-0.07, 0.25]$ with the QQr and he CIs of DBK's MD are $[-0.06, 0.20]$ with the MLE, $[-0.07, 0.18]$ with the SW, $[-0.10, 0.13]$ with the JB and $[-0.06, 0.19]$ with the QQr. Due to the small value $\hat{\lambda}$ and its CI, it seems that a logarithmic transformation is suitable in the power transformation process. In Fig. 6.9 and 6.10, it is obvious that after the Box-Cox power transformation with $\hat{\lambda}$, both the histogram and the Q-Q plot of SIE and DBK perform better than before. Therefore, in the semiparametric analysis, a Semi-Log-MEM process is preferred in this case.

Similar to the MD, the VO series also possesses an almost zero $\hat{\lambda}$ and a through-origin CI, which also indicates a log-transform consideration. However, the AR set performs a little different. The $\hat{\lambda}$ with all methods are definitely positive and the values are between 0.32 and 0.36. Like the returns in the SemiGARCH model with Box-Cox transformation (Zhang et al., 2017), a third-root power transformation based on a SemiMEM process seems to be applicable to the AR series. Besides, the results of the TR series are between the above two circumstances. The simulated results of SIE are close to the AR-type, however, those of DBK are in the form of the MD-VO-type. Thus, for the TR series, both the SemiMEM process and the Semi-Log-MEM process are the possible power transformation models, depending on the exact $\hat{\lambda}$ selection.

Table 6.1: The selected λ, confidence interval and bandwidth

		MLE			SW			JB			QQr		
		λ	CI	b	λ	CI	b	λ	CI	b	λ	CI	b
SIE	MD	0.09	[-0.05, 0.25]	0.07	0.07	[-0.08, 0.23]	0.07	0.00	[-0.15, 0.22]	0.08	0.09	[-0.07, 0.25]	0.07
	AR	0.34	[0.30, 0.37]	0.12	0.34	[0.30, 0.39]	0.12	0.32	[0.26, 0.39]	0.12	0.35	[0.30, 0.39]	0.12
	VO	0.05	[-0.15, 0.18]	0.07	0.09	[-0.13, 0.21]	0.07	0.16	[-0.08, 0.27]	0.07	0.05	[-0.15, 0.20]	0.07
	TR	0.27	[0.00, 0.41]	0.08	0.30	[0.03, 0.44]	0.08	0.38	[0.08, 0.50]	0.08	0.26	[0.00, 0.43]	0.08
DBK	MD	0.07	[-0.06, 0.20]	0.07	0.05	[-0.07, 0.18]	0.07	0.00	[-0.10, 0.13]	0.07	0.06	[-0.06, 0.19]	0.07
	AR	0.34	[0.32, 0.36]	0.07	0.35	[0.33, 0.38]	0.07	0.36	[0.33, 0.39]	0.07	0.35	[0.33, 0.38]	0.07
	VO	-0.04	[-0.14, 0.06]	0.08	-0.02	[-0.13, 0.08]	0.08	0.01	[-0.11, 0.12]	0.08	-0.03	[-0.13, 0.08]	0.08
	TR	0.10	[-0.04, 0.23]	0.07	0.12	[-0.02, 0.25]	0.07	0.20	[0.02, 0.30]	0.07	0.11	[-0.03, 0.24]	0.07

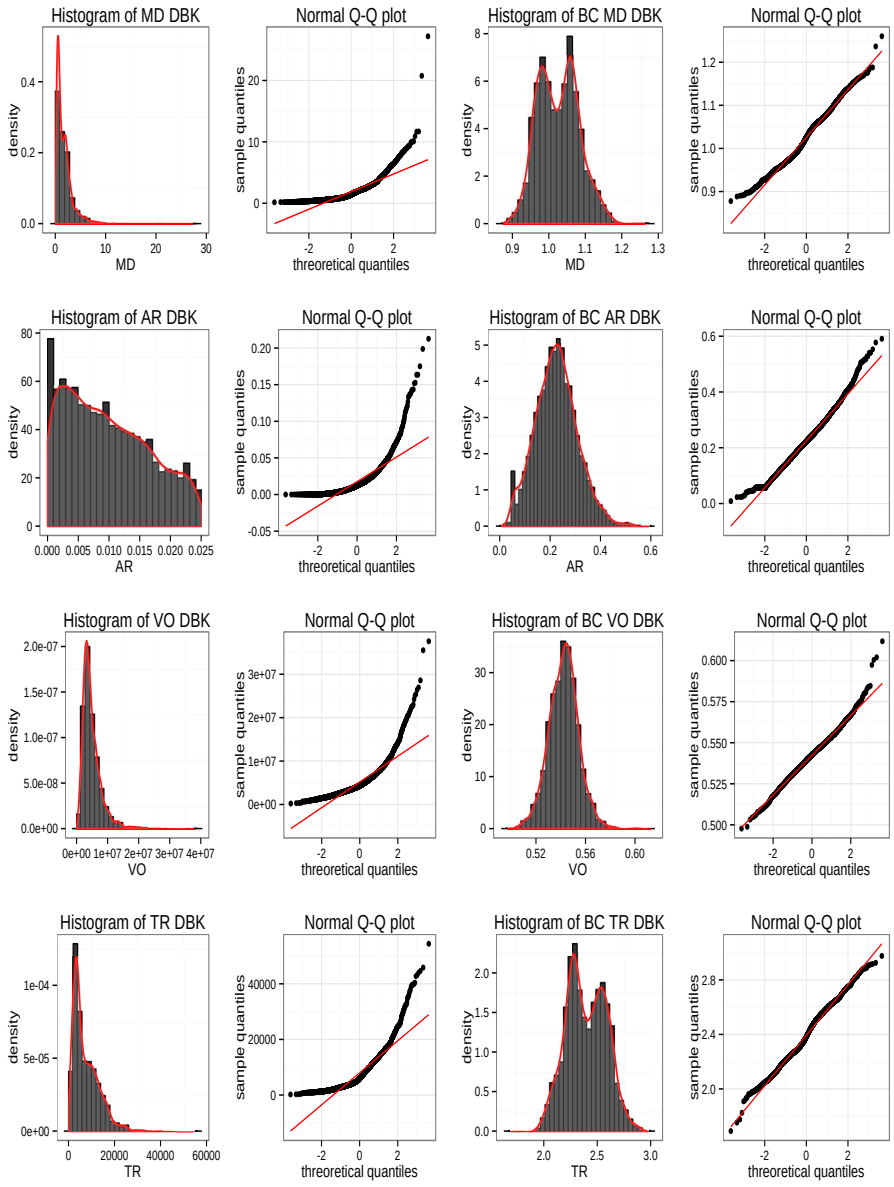

Figure 6.9: The histogram and Q-Q plot of DBK

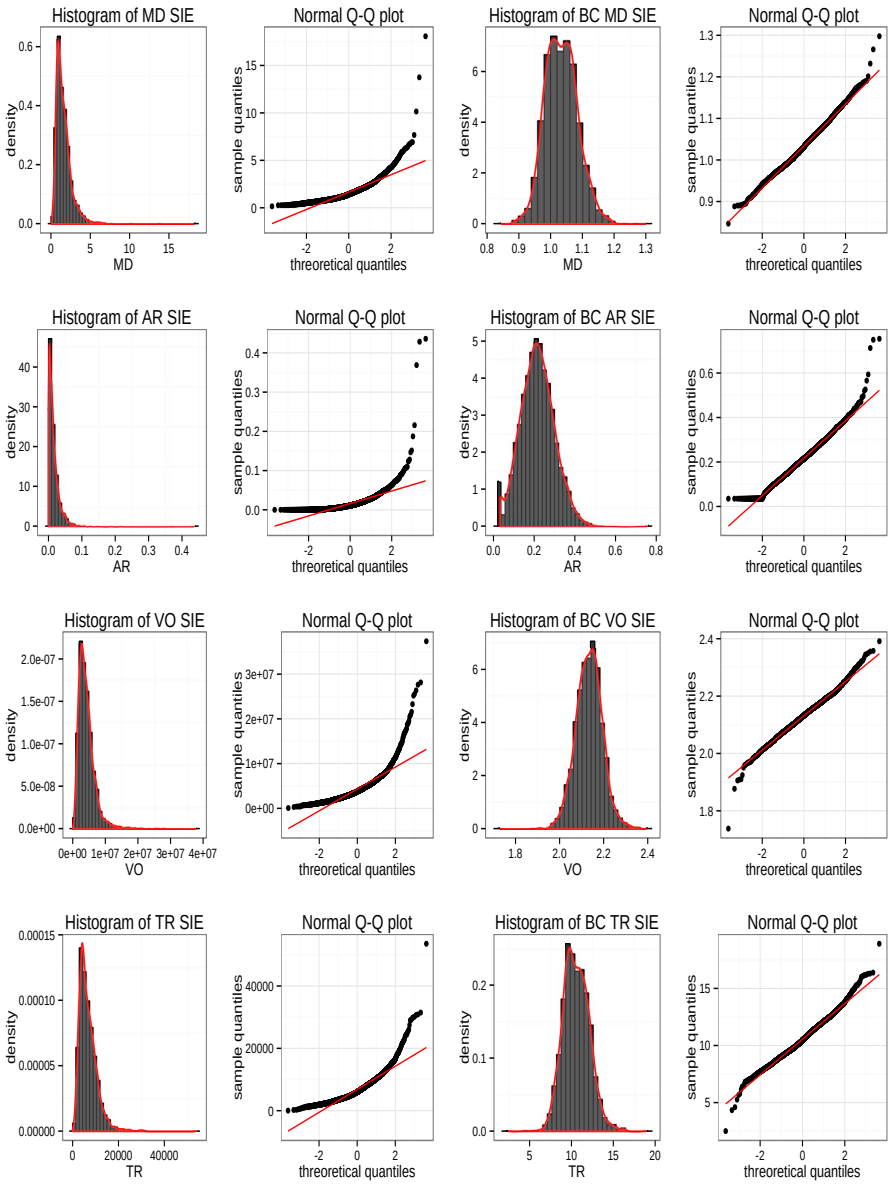

Figure 6.10: The histogram and Q-Q plot of SIE

To describe the simulation quality of the $\hat{\lambda}$'s CI, we have to bring the length (L) and shape (Δ) statistics into consideration. The two statistics are defined as

$$L = \hat{\lambda}^*_{k_U} - \hat{\lambda}^*_{k_L}, \tag{6.28}$$

and

$$\Delta = (\hat{\lambda}^*_{k_U} - \hat{\lambda})/(\hat{\lambda}^*_{k_L} - \hat{\lambda}). \tag{6.29}$$

Table 6.2: The length and shape of simulated CI

		MLE		SW		JB		QQr	
		L	Δ	L	Δ	L	Δ	L	Δ
SIE	MD	0.30	1.14	0.31	1.07	0.37	1.47	0.32	1.00
	AR	0.07	0.97	0.09	1.14	0.13	1.32	0.09	1.02
	VO	0.33	0.65	0.34	0.55	0.35	0.46	0.35	0.75
	TR	0.41	0.52	0.41	0.52	0.42	0.40	0.43	0.65
DBK	MD	0.26	1.00	0.25	1.08	0.23	1.30	0.25	1.08
	AR	0.04	1.09	0.05	1.29	0.06	1.07	0.05	1.07
	VO	0.20	0.92	0.21	0.91	0.23	0.92	0.21	1.10
	TR	0.27	0.91	0.27	0.93	0.28	0.56	0.27	0.93

In the chapter, we employ the percentile CI, which is a first-order accurate procedure. If the coverage probabilities are identical, the two statistics can be applied to detect the CI quality. Therefore, we evaluate the simulation quality of CI through the L and Δ.

In Table 6.2, we can see that in most case, the lengths of the CI with MLE, SW and QQr are always more accurate than those with JB, e.g. the CI length of MD with the MLE, SW and QQr are 0.3, 0.31 and 0.32, however, the CI length of JB reach even 0.37. For another, from the performance of the shape, the shapes of the MLE and QQr are closer to one than the other two's, which means that the $\hat{\lambda}$ lies closer to the center of the CI, so that the CIs selected by the MLE and QQr are more asymmetric. In general, the CIs with the MLE and QQr perform well in practice, while the CI with JB seems to be not as good as the other ones. In addition, the CI with SW is not so stable, especially in the shape performance, e.g. the VO shape of SIE is only 0.55 and the AR shape of DBK is as high as 1.29. So, in the latter descale process, we are going to apply the $\hat{\lambda}$ by MLE (tiny difference from $\hat{\lambda}$ with QQr) as the considered power.

Note that the CIs discussed in this section only refer to the first-order accurate percentile CI. For the purpose of the CI modification, the second-order accurate methods can be introduced, such as the bootstrap-t and the BCa methods (Efron, 2003). In addition to the discussed data, we have also tried the realized volatility (RV), however the $\hat{\lambda}$ of RV is always negative, due to the extreme values. Hence, the RV is no longer considered here.

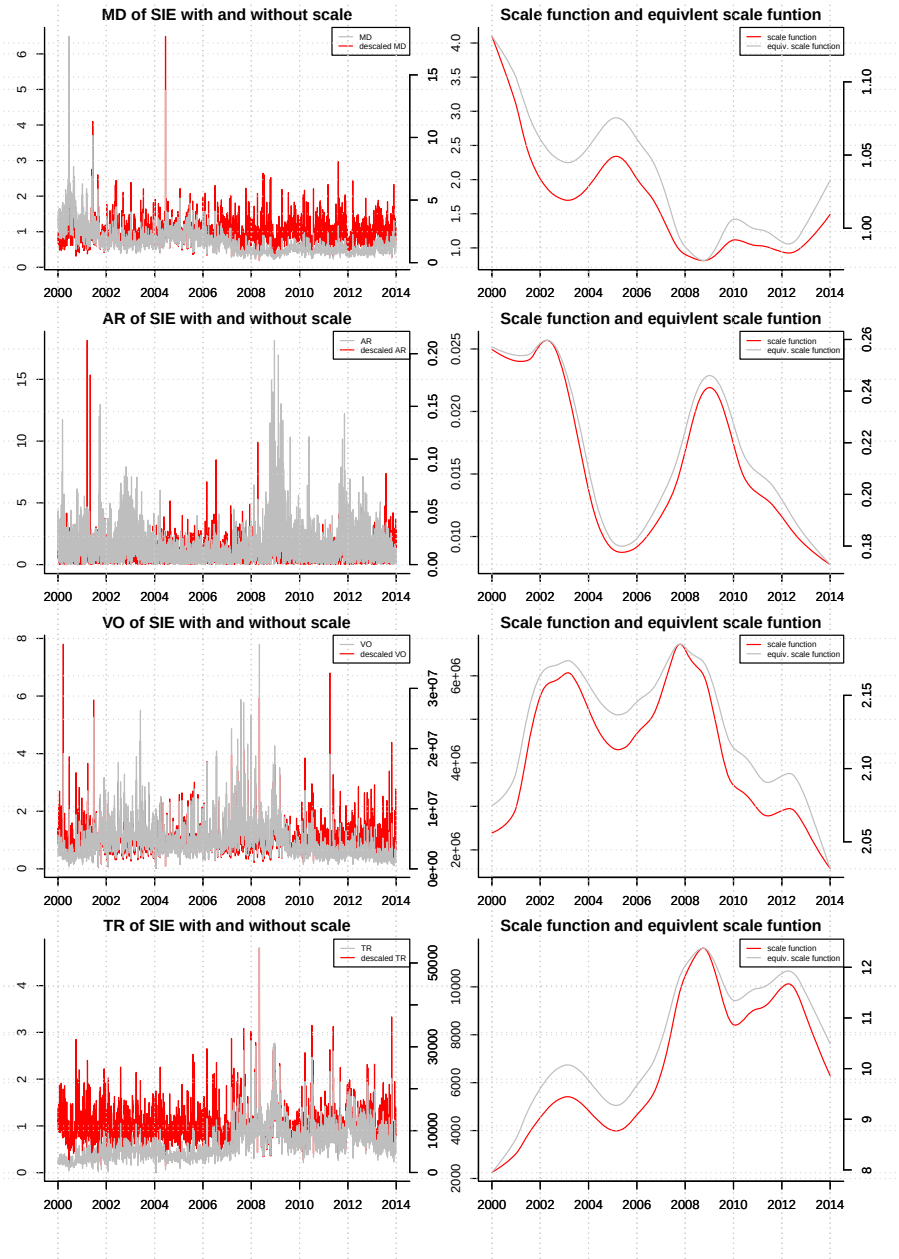

Figure 6.11: Estimation results of SIE from Jan 2000 to Dec 2013

Finally, in Fig. 6.11 and Fig. 6.12, the original and descaled data are displayed at the left side, while the estimation of the scale function and the equivalent scale function are at the right side.

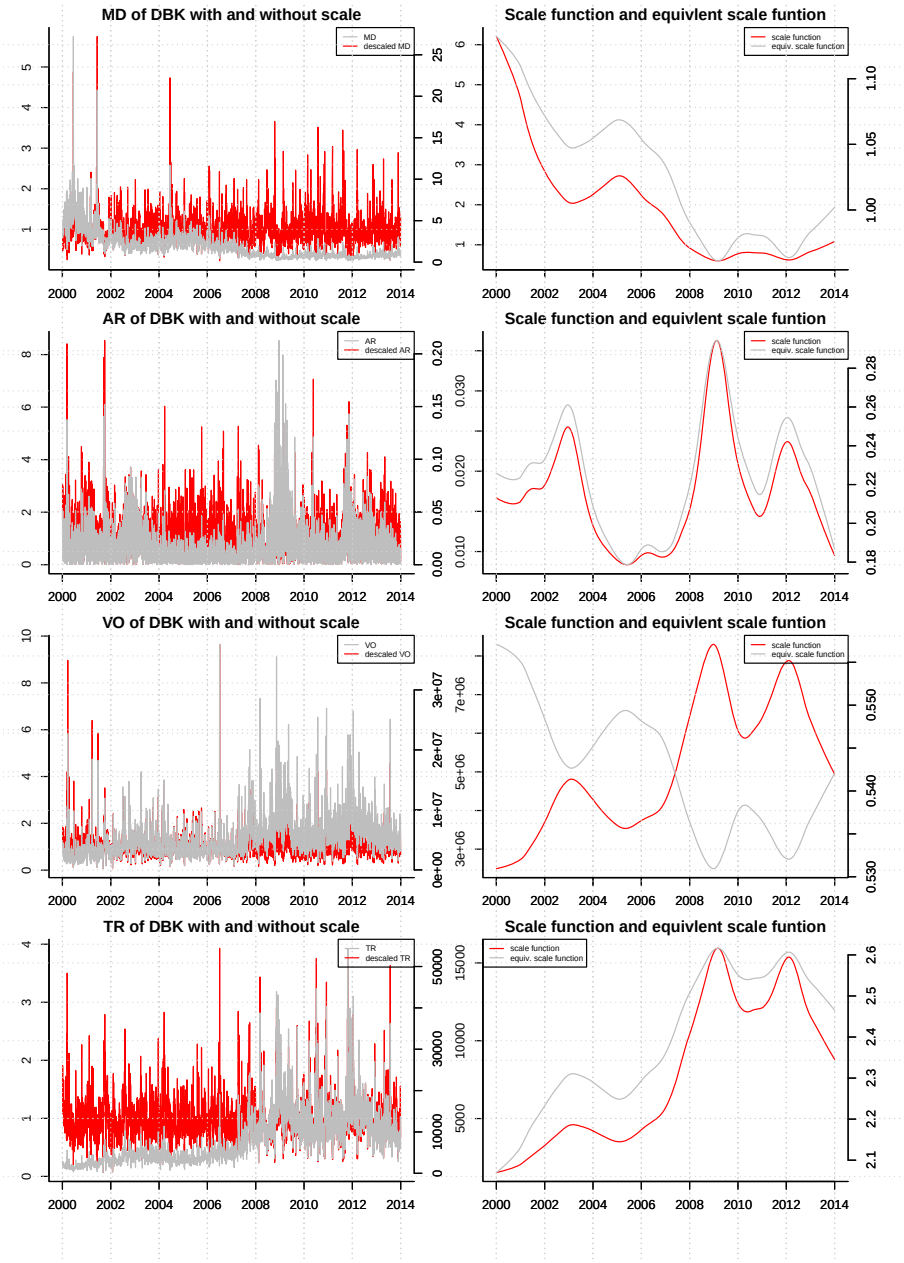

Figure 6.12: Estimation results of DBK from Jan 2000 to Dec 2013

Comparing the data, we can see obviously that the descaled data are more stationary than the original ones, bringing the smaller volatility after removing

the trend function from the nonstationary process. Further, we have to pay attention to the change of the scale function. During the global financial crisis, the transaction risk reaches its peak, combining with the high transaction volume and numbers and low transaction duration. Therefore, we can see that the MD is at its minimum value, however, the AR, VO and TR all reach the peak values. Besides, the data are selected from DAX 30, where the German 'neuer Markt'[2] (2002-2003) can also be observed by reaching the changing points in the sub-period.

6.6 Final remarks

We discussed a general SemiMEM model with Box-Cox power transformation to analyze the non-negative data, such as MD, AR, VO and TR by selecting the power parameter with the MLE, SW, JB and QQr criteria. In the IPI procedures, we found that, for the MD and VO, the Semi-Log-MEM model is optimal, due to the approaching zero $\hat{\lambda}$, while for the AR, the $\hat{\lambda}$ is significantly positive, leading to a SemiMEM process. Both the Semi-Log-MEM and SemiMEM are suitable for the TR, depending on the $\hat{\lambda}$ selection in cases. Furthermore, we have discovered the simulated CI of $\hat{\lambda}$ without model or distribution assumption by introducing the block bootstrap method for dependent variables. The simulation CI quality is also considered. With the accurate coverage probabilities, we introduced the length and shape parameters to detect the simulated CI quality with the four mentioned criteria and it is found that the CI qualities with the MLE and QQr methods are better. Besides, IPI algorithms are also developed, referring to the c_f selection. The c_f is selected using the algorithm in Bühlmann (1996), estimating a Bartlett-window estimator and the optimal window width is proved to be an $O(T^{-1/3})$ term. Hence, in the chapter, we selected the cubic root of the sample size as the optimal block length for simplicity.

Finally, some contributions will be possibly carried out in the future. First, the long memory parameter can be introduced into the framework. The long mem-

[2]Listed companies of Nemax 50 declined sharply in share prices within three years since the dotcom crash in 2000.

ory models, such as the Semi-FI-MEM and Semi-FI-Log-MEM models, can be discussed with the suitable power transformation under weak moment conditions, developing the algorithm of the synchronous selection of the differencing parameter d and the power parameter λ. Then, the power transformation technique can be updated. In the chapter, we applied the Box-Cox transformation and with this method, the transformation is only close to the normal distribution, leading to the application of the other distribution, such as sin-arcsin distribution, sinh-normal distribution, Birnbaum-Saunders distribution, and so on. Further, the intraday high-frequency financial data can be introduced. Set up the spatial general SemiMEM models by modeling the high-frequency data in the daily and intraday dimension.

Chapter 7

Further topics

7.1 Introduction

Financial markets have grown rapidly within the last decades and the corresponding financial instruments have become difficult to handle. Issues of particular importance are highly volatile markets that increase the level of risks for investment decisions. The need for risk indicators has led to a rapid growth in research on the price volatility, the trading volume and the trading duration, which are the non-negative variables in financial markets. Most of the studies have regarded the above non-negative variables as crucial risk factors in the financial market. Therefore, the variables are related to uncertainty, since it is crucial in risk management, portfolio and investment decisions. As a result, the variables are approximated by using statistical computing methods to carry out perceptible values, such as the daily changes and intraday changes.

The most common approaches to calculate the discussed non-negative variables is characterized in parametric, such as the GARCH model and its extensions and nonparametric models. The restrictiveness of the GARCH models and their corresponding weaknesses led to a simpler and more flexible nonparametric approach to consider the risk in the market. Besides, the bias increases with the sampling frequency. The literature has investigated, which level of intervals is optimal and moreover, which sampling scheme is superior. Generally, the calendar time sampling, the business time sampling and the tick time sampling are discussed. In this chapter, a new sampling method proposed by Feng, the so-called k-method, is introduced, where each k-th observation is selected in terms of the data densi-

ty and obviously, the first and last values of the selected observations are fixed. Moreover, a general class of semiparametric MEM models is suggested with the semiparametric ACD model as a special case. A Box-Cox transformation is applied in the general semiparametric MEM process. The purpose of the chapter is to analyze the high-frequency non-negative variables with a Box-Cox semiparametric MEM model. A data-driven iterative plug-in algorithm introduced is used to carry out the estimation of the scale function and the bandwidth and power parameter λ selection. The selection of the power parameter λ is achieved by the MLE and Jarque-Bera (JB) statistics. Furthermore, the selection and estimation of appropriate parametric models are also given. The parametric models are a general MEM class, which makes the semiparametric process as parametric model free.

The chapter is organized as follows. Section 7.2 discusses the sampling scheme methods of the ultra high-frequency financial data. In Section 7.3, the SemiMEM model is interpreted and the data-driven algorithms are also provided. The empirical analysis of different trading days is given in section 7.4. Finally, the chapter ends with a brief conclusion.

7.2 The sampling schemes

In this section, the sampling schemes of high-frequency data are introduced and a new k-method is also proposed as the sampling schemes in the following empirical part.

7.2.1 The CTS, TTS and BTS

One major problem, which commonly arises in high-frequency data analysis, is that not all available transactions can be implemented at one time. The number of daily-recorded data can be overwhelmingly high, which makes the handling of the data very difficult, i.e. a sub-grid approach is obligatory. Consequently, the first step to select the sub-grid is to choose the sampling scheme and the second step is

to select the target average sampling frequency (Dong and Tse, 2014). According to Oomen (2006), the time sampling schemes were primarily differentiated as regulators of data capturing. The calendar time sampling (CTS) scheme selects the data or transactions on equidistant calendar time, for instance every 10 minutes. Following the transaction time sampling (TTS) scheme, the data in event times are without predefined and it is not equidistant. Obviously, in the TTS, each individual transaction is recorded and the most frequently available information is provided. Another sampling scheme is the tick-time sampling (TickTS), similar to the TTS but all zero returns are removed. Moreover, transactions are commonly chosen based on regularly spaced numbers of ticks, e.g. every 5 or 10 ticks to develop the observed price process. Besides, Dacorogna et al. (1993) proposed the business time sampling (BTS) to analyze the diurnal and weekly seasonality in the volatility. Practically, the CTS and TTS schemes are more widely used than the BTS.

7.2.2 The k-method

In this section, we discuss a new sampling method, called the k-method. The idea is to take every k-th observation of any given time series, obtaining a new series with M intervals and $M+1$ observations. M is the desired number of intervals in the analysis and usually the same for all transaction days and its value varies, depending only on the analysis. For the non-negative financial data set, starting from the first observation, every k-th observation will be selected in terms of the data density. Obviously, k can be determined by

$$k = \lfloor \frac{N}{M} \rfloor, \tag{7.1}$$

where N denotes the number of the non-negative observations per day and $n_0 = 1, n_1, \ldots, n_{(M-1)}, n_M = N$ should be chosen on the equal terms if possible. As already noted, the first value $n_0 = 1$ and the last value $n_M = N$ are fixed. According to the analysis of the non-negative variables in this chapter, M is defined equal to 510 so as to make sure that the interval of the selected series is 510 and the new

data set contains 511 observations, including the first and last observations from the original series. Subsequently, every k-th observation will be chosen from the first observation to the end of the series.

The k-method is similar to the BTS, however not exactly the same. Following the BTS, the size of the selected data set from each observed transaction day may be different, which therefore leads to the incomparability among transaction days. The k-method, however, allows the comparison between various selected transaction days, since all selected data sets contain the same $M+1$ observations. Consequently, in this chapter, the k-method is carried out as the sampling scheme method.

7.3 The SemiMEM models to high-frequency data

In the MEM approach, the non-negative data can be specified as

$$X_t = m(\tau_t)\psi_t\eta_t, \tag{7.2}$$

where $\tau_t = t/n$ denotes the rescaled time and $m(\cdot) > 0$ defines a smooth trend, which presents either the localized unconditional mean function or the scale function in X_t. Moreover, the conditional mean is here given by $\psi_t \geq 0$ and $\eta_t \geq 0$ are i.i.d. innovations with a unit mean,

$$x_t = \psi_t\eta_t, \tag{7.3}$$

signifies the descaled stationary process. A common technique in nonparametric regression with time series errors is the application of a rescaled time, written as $\tau_t \in [0, 1]$. Since $E(x_t|\mathcal{F}_{t-1}) = \psi_t$, it becomes evident that ψ_t is the conditional mean. In order to guarantee a unique definition of model (7.2), it is required that $E(\psi_t) = 1$. As a result, this model is called a SemiMEM model. The equation of model (7.2) certainly illustrates a sequence of models. In the case, if $m(\cdot)$ is not constant, the process $\{X_t\}$ is nonstationary.

Besides, it is also studied the estimation of the slowly scale change that bases

on the Box-Cox power transformation x_t^λ for any $-1 < \lambda < 1$. The the SemiMEM model with power transformation is defined as

$$X_t = m_\lambda(\tau_t) x_{t,\lambda}, \tag{7.4}$$

where $x_{t,\lambda}$ is noted as a stationary process with $E(x_{t,\lambda}^\lambda) = 1$. The Equation (7.4) can be rewritten with nonparametric regression as

$$X_t = g_\lambda(\tau_t) + g_\lambda(\tau_t)\zeta_{t,\lambda}, \tag{7.5}$$

where $m_\lambda(\tau_t) = c_\lambda^{1/\lambda} m(\tau_t) = [g_\lambda(\tau_t)]^{1/\lambda}$ and $\zeta_{t,\lambda}$ is defined as $\zeta_{t,\lambda} = x_{t,\lambda}^\lambda - 1$ with zero mean and unit variance. As indicated by Zhang et al. (2017) and Zhang (2019a), it is suggested to use the estimator $\hat{g}_\lambda(\tau_t) = |\tilde{g}_\lambda(\tau_t)|$, where $\tilde{g}_\lambda(\tau_t) = \hat{a}_0(x)$ is a local linear estimate obtained by minimizing the following equation

$$Q(a_0, a_1) = \sum_{t=1}^{T} \{x_t - a_0(\tau) - a_1(\tau)(\tau_t - \tau)\}^2 K\left(\frac{\tau_t - \tau}{b}\right), \tag{7.6}$$

where K is defined as a symmetric non-negative kernel function and $0 < b < 1/2$ is the bandwidth.

The $\hat{\lambda}$ selection algorithm in Zhang et al. (2017) will be carried out in the chapter. A starting bandwidth b_0 and power parameter λ_0 is essential to be defined. For λ_0, different values will be considered in empirical research. In this chapter, λ_0 will be set equal to the values, 1, -1, 0,5 and -0,5. The final values of $\hat{\lambda}$ and $\hat{b}_A$ are obtained in a six-step iteration process. Finally, it is found that the final selected estimated value is independent of the initial inputs. Different methods can be used to specify λ, such as the MLE, the normality (JB and SW) test and the quantile-quantile regression (QQr). In the following, the MLE, JB, SW statistics and the QQr will be applied in the empirical work of the λ selection.

7.4 The empirical analysis

In this section, the analysis of the high-frequency realized volatility (RV), trading volume (VO) and trading duration (TR) with Box-Cox SemiMEM models will be carried out. In the semiparametric process, the power transformation parameter is selected with the MLE, JB, SW and QQr criteria. Besides, parametric ACD models with the Burr (BACD), the exponential (EACD), the Weibull (WACD) and the Gamma (GACD) distributions are considered. In the empirical study, all sample days are considered, however, only one day is selected to analyze.

7.4.1 The data

The data, Allianz (ALV), Siemens (SIE), BMW and Deutsche Bank (DBK) from DAX 30 are discussed. Each data covers four weeks from 12.09.2011 to 07.10.2011, including 20 trading days per company, to analyze the daily pattern in high-frequency financial data. The source of data is the database Thomson Reuters. The table below gives an data overview.

Table 7.1: The UHF observations

Company	Obs. period		Average obs.	Total obs.
	Start	End		
Allianz	09 Sep, 2011	07 Oct, 2011	15559	311187
BMW	09 Sep, 2011	07 Oct, 2011	15394	307888
Deutsche Bank	09 Sep, 2011	07 Oct, 2011	25630	512600
Siemens	09 Sep, 2011	07 Oct, 2011	16560	331215

The above data sets are applied to the designed algorithm. In the first step, the data is sampled by the k-method. The selected data has 511 observations each day, due to the fixed open and close time points of the financial market (from 9:00 to 17:30). The non-negative financial data sets are applied to an IPI algorithm

to obtain suitable power transformation parameters λ with MLE, JB, SW and QQr. The $\hat{\lambda}$ values will also be estimated in the algorithm. Parametric models will be fitted to the descaled standardized non-negative data. For simplicity, only the empirical results of ALV are discussed.

7.4.2 The analysis of the intraday trading volume

In the analysis of TR, we consider the data of Allianz on September 27, 2011, as an example. In Fig. 7.1, it shows the final selected power parameter values with the IPI processes by JB, MLE, SW and QQr, respectively. Obviously, all λ values are identical from the second IPI procedure and the identical λ is the selected convergence power parameter, which is used as the power of the Box-Cox transformation. However, as displayed in Table 7.2, we get unlike but very close λ with different criteria, for example, in this case, the stable λ of ALV are $-0.64, -0.68, -0.68, -0.73$ of JB, MLE, SW and QQr, respectively. Besides, the λ value is irrelevant with the initial inputs for each criteria, defined as $-1, -0.5, 0, 0.5, 1$, reaching the same stable value. It is also proved in Fig. 7.1. The horizontal axis is defined as the six IPI steps and the vertical axis is the initial starting λ values. It displays the IPI processes of Allianz with the four criteria and obviously, the λ reaches its convergence no matter how large the initial inputs are. After the 6-step process, the power parameter λ will be finally selected. As shown in the figure, it is not important which inputs from $[-1, 1]$ are set, because the finally selected λ is always obtained the same value. It is also apparent that the data set will reach its convergence values after the second IPI procedure, in particular in the λ selection with MLE, which is common, however only a few need a third step, for instance, the λ selected via JB. Figure 7.1 also shows the average trading volume after the Box-Cox transformation with MLE. Obviously, from the histogram, it is possible to obtain a transformed distribution that approximates to normal, if optimal selected λ is applied. The scale function is also displayed in Fig. 7.1. In the plots, it is found that the scale functions the example has a U-curve pattern. The fitted parametric model and semiparametric ACD models are listed in Table 7.3. Semiparametric models perform better than parametric models form BIC and

MSE. The semiparametric models obtain smaller BIC values with the same data and especially, the MSE values are much smaller than those of parametric models.

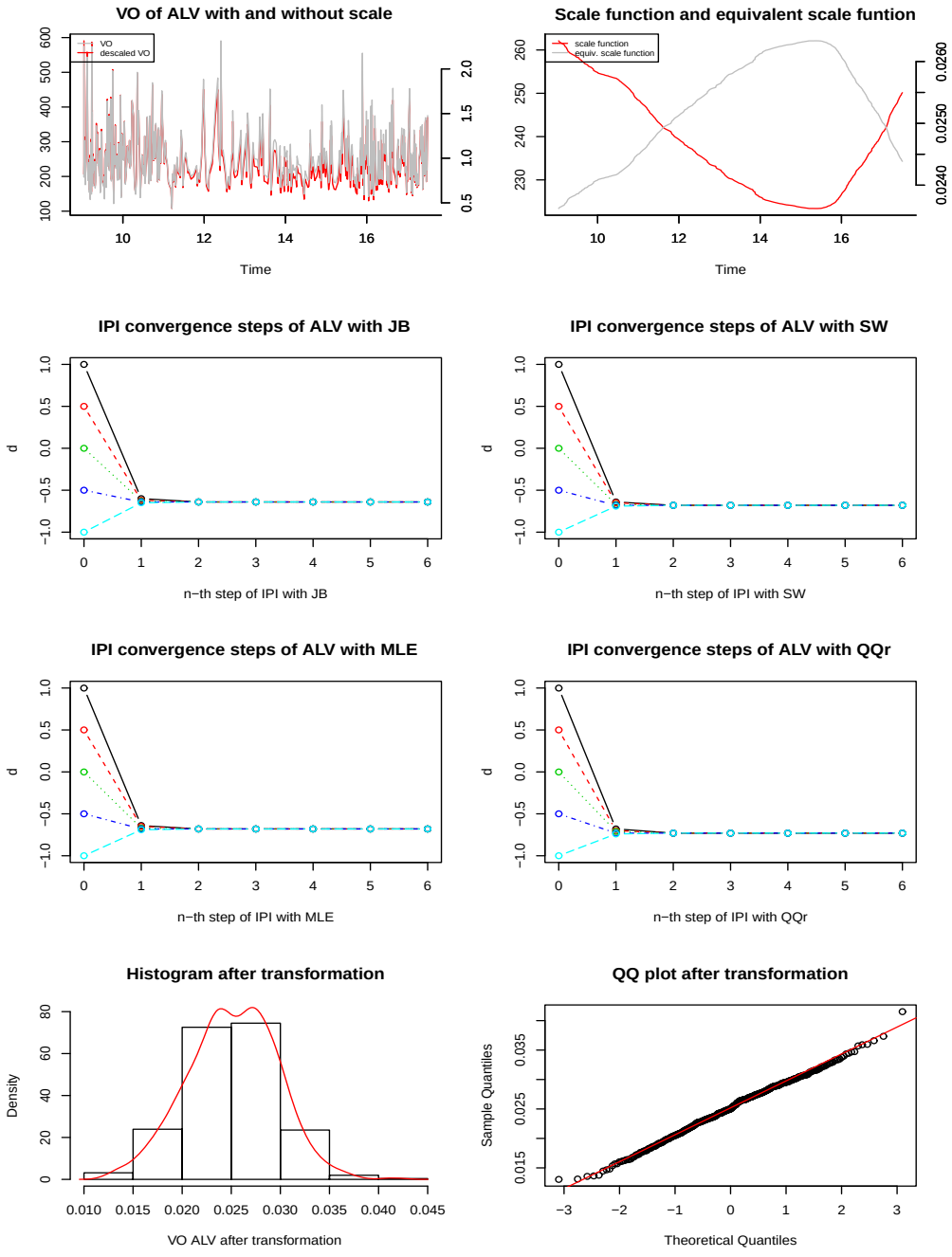

Figure 7.1: The power transformation results of ALV VO on 27 Sep, 2011

Table 7.2: The IPI selected $\hat{\lambda}$ of ALV VO with MLE, JB, SW and QQr

Day	λ_0	MLE	JB	SW	QQr	Day	λ_0	MLE	JB	SW	QQr	Day	λ_0	MLE	JB	SW	QQr	Day	λ_0	MLE	JB	SW	QQr
	1.00	0.34	0.42	0.38	0.30		1.00	-0.37	-0.35	-0.37	-0.39		1.00	-0.22	-0.21	-0.21	-0.22		1.00	-0.79	-0.80	-0.81	-0.82
	0.50	0.34	0.42	0.38	0.30		0.50	-0.37	-0.35	-0.37	-0.39		0.50	-0.22	-0.21	-0.21	-0.22		0.50	-0.79	-0.80	-0.81	-0.82
1	0.00	0.34	0.42	0.38	0.30	6	0.00	-0.37	-0.35	-0.37	-0.39	11	0.00	-0.22	-0.21	-0.21	-0.22	16	0.00	-0.79	-0.80	-0.81	-0.82
	-0.50	0.34	0.42	0.38	0.30		-0.50	-0.37	-0.35	-0.37	-0.39		-0.50	-0.22	-0.21	-0.21	-0.22		-0.50	-0.79	-0.80	-0.81	-0.82
	-1.00	0.34	0.42	0.38	0.30		-1.00	-0.37	-0.35	-0.37	-0.39		-1.00	-0.22	-0.21	-0.21	-0.22		-1.00	-0.79	-0.80	-0.81	-0.82
	1.00	-0.14	-0.07	-0.11	-0.15		1.00	-0.42	-0.42	-0.42	-0.42		1.00	-0.68	-0.64	-0.68	-0.73		1.00	-0.54	-0.52	-0.53	-0.54
	0.50	-0.14	-0.07	-0.11	-0.15		0.50	-0.42	-0.42	-0.42	-0.42		0.50	-0.68	-0.64	-0.68	-0.73		0.50	-0.54	-0.52	-0.53	-0.54
2	0.00	-0.14	-0.07	-0.11	-0.15	7	0.00	-0.42	-0.42	-0.42	-0.42	12	0.00	-0.68	-0.64	-0.68	-0.73	17	0.00	-0.54	-0.52	-0.53	-0.54
	-0.50	-0.14	-0.07	-0.11	-0.15		-0.50	-0.42	-0.42	-0.42	-0.42		-0.50	-0.68	-0.64	-0.68	-0.73		-0.50	-0.54	-0.52	-0.53	-0.54
	-1.00	-0.14	-0.07	-0.11	-0.15		-1.00	-0.42	-0.42	-0.42	-0.42		-1.00	-0.68	-0.64	-0.68	-0.73		-1.00	-0.54	-0.52	-0.53	-0.54
	1.00	-0.39	-0.39	-0.39	-0.40		1.00	0.24	0.42	0.32	0.17		1.00	-0.23	-0.17	-0.21	-0.25		1.00	-0.30	-0.26	-0.29	-0.31
	0.50	-0.39	-0.39	-0.39	-0.40		0.50	0.24	0.42	0.32	0.17		0.50	-0.23	-0.17	-0.21	-0.25		0.50	-0.30	-0.26	-0.29	-0.31
3	0.00	-0.39	-0.39	-0.39	-0.40	8	0.00	0.24	0.42	0.32	0.17	13	0.00	-0.23	-0.17	-0.21	-0.25	18	0.00	-0.30	-0.26	-0.29	-0.31
	-0.50	-0.39	-0.39	-0.39	-0.40		-0.50	0.24	0.42	0.32	0.17		-0.50	-0.23	-0.17	-0.21	-0.25		-0.50	-0.30	-0.26	-0.29	-0.31
	-1.00	-0.39	-0.39	-0.39	-0.40		-1.00	0.24	0.42	0.32	0.17		-1.00	-0.23	-0.17	-0.21	-0.25		-1.00	-0.30	-0.26	-0.29	-0.31
	1.00	-0.46	-0.54	-0.48	-0.37		1.00	-0.60	-0.65	-0.62	-0.53		1.00	-0.73	-0.73	-0.72	-0.67		1.00	-0.42	-0.44	-0.42	-0.32
	0.50	-0.46	-0.54	-0.48	-0.37		0.50	-0.60	-0.65	-0.62	-0.53		0.50	-0.73	-0.73	-0.72	-0.67		0.50	-0.42	-0.44	-0.42	-0.32
4	0.00	-0.46	-0.54	-0.48	-0.37	9	0.00	-0.60	-0.65	-0.62	-0.53	14	0.00	-0.73	-0.73	-0.72	-0.67	19	0.00	-0.42	-0.44	-0.42	-0.32
	-0.50	-0.46	-0.54	-0.48	-0.37		-0.50	-0.60	-0.65	-0.62	-0.53		-0.50	-0.73	-0.73	-0.72	-0.67		-0.50	-0.42	-0.44	-0.42	-0.32
	-1.00	-0.46	-0.54	-0.48	-0.37		-1.00	-0.60	-0.65	-0.62	-0.53		-1.00	-0.73	-0.73	-0.72	-0.67		-1.00	-0.42	-0.44	-0.42	-0.32
	1.00	-0.68	-0.73	-0.71	-0.61		1.00	0.10	0.17	0.13	0.11		1.00	-0.52	-0.52	-0.52	-0.49		1.00	-0.44	-0.39	-0.42	-0.46
	0.50	-0.68	-0.73	-0.71	-0.61		0.50	0.10	0.17	0.13	0.11		0.50	-0.52	-0.52	-0.52	-0.49		0.50	-0.44	-0.39	-0.42	-0.46
5	0.00	-0.68	-0.73	-0.71	-0.61	10	0.00	0.10	0.17	0.13	0.11	15	0.00	-0.52	-0.52	-0.52	-0.49	20	0.00	-0.44	-0.39	-0.42	-0.46
	-0.50	-0.68	-0.73	-0.71	-0.61		-0.50	0.10	0.17	0.13	0.11		-0.50	-0.52	-0.52	-0.52	-0.49		-0.50	-0.44	-0.39	-0.42	-0.46
	-1.00	-0.68	-0.73	-0.71	-0.61		-1.00	0.10	0.17	0.13	0.11		-1.00	-0.52	-0.52	-0.52	-0.49		-1.00	-0.44	-0.39	-0.42	-0.46

Table 7.3: The parametric and SemiBC fitting results of VO with the order $(1, 1)$

Data	Model	Para-								SemiBC-							
		ω	α_1	β_1	γ	κ	σ^2	BIC	MSE	ω	α_1	β_1	γ	κ	σ^2	BIC	MSE
ALV	BACD	17.05	0.05	0.88	-	9.54	2.19	5681.59	5389.63	0.10	0.13	0.76	-	1.63	0.01	621.05	0.09
	EACD	125.50	0.23	0.25	-	-	-	6629.85	5257.34	0.71	0.20	0.09	-	-	-	1036.95	0.09
	GACD	137.29	0.22	0.21	0.37	99.46	-	5703.93	5255.77	0.74	0.18	0.08	0.02	23470.00	-	85.70	0.09
	WACD	78.29	0.24	0.42	3.29	-	-	5852.30	5306.35	0.60	0.23	0.17	3.38	-	-	237.46	0.09
BMW	BACD	38.75	0.20	0.64	-	5.54	1.54	5951.55	11459.16	0.27	0.17	0.55	-	5.47	1.47	404.36	0.20
	EACD	25.59	0.17	0.71	-	-	-	6569.25	11473.04	0.20	0.16	0.64	-	-	-	1032.45	0.21
	GACD	36.24	0.18	0.66	0.08	983.88	-	5978.80	11457.10	1.93	0.03	-0.99	0.87	7.54	-	496.03	0.22
	WACD	14.64	0.16	0.78	2.25	-	-	6131.98	11559.52	0.13	0.14	0.73	2.26	-	-	588.14	0.21
DBK	BACD	110.89	0.15	0.63	-	6.28	1.25	6562.66	27569.58	0.50	0.12	0.38	-	6.52	1.28	200.03	0.10
	EACD	96.01	0.15	0.66	-	-	-	7376.96	27568.53	0.47	0.13	0.41	-	-	-	1037.55	0.10
	GACD	39.46	0.10	0.82	0.26	156.45	-	6568.81	27767.33	0.54	0.12	0.34	0.04	8732.49	-	202.69	0.10
	WACD	75.75	0.16	0.68	3.05	-	-	6679.71	27618.81	0.33	0.15	0.52	3.06	-	-	330.24	0.10
SIE	BACD	38.40	0.26	0.56	-	5.80	1.16	5722.27	5874.17	0.67	0.24	0.09	-	6.02	1.14	238.78	0.11
	EACD	53.05	0.31	0.44	-	-	-	6483.37	5895.81	0.75	0.25	0.00	-	-	-	1035.48	0.11
	GACD	13.64	0.21	0.73	0.62	22.29	-	5770.09	5917.64	0.71	0.26	0.03	0.43	52.36	-	272.65	0.11
	WACD	69.90	0.32	0.34	2.74	-	-	5875.40	5921.77	0.58	0.14	0.28	2.93	-	-	377.78	0.11

7.4.3 The analysis of the intraday trading duration

In the following the Allianz SE data will be analyzed in detail for the given first trading day, which is on September 12, 2011. In Fig. 7.3, the six steps IPI procedures for the lambda selection with JB, MLE, SW and QQr, respectively. The initial lambda values as before defined as 1, 0.5, 0, -0.5 and -1. When $\lambda = 0$, a logarithmic transformation should be considered. The optimal power transformation parameter λ of the Box-Cox transformation is found in the figures when the curve reaches their convergence values. In this case, the selected λ as showed in Table 7.4, are $0.51, 0.54, 0.54, 0.56$ for the above criteria. Besides, it is known that the duration series, which is the time between trades, is shorter on average at the beginning and at the end of the trading day, while about at noon durations reach their highest values. Meaning that the trading is active at the open and close time and reduced during midday. Here, in this figure, the diurnal pattern can be obviously identified. The scale function has lower values also at the beginning and the end of the day. With increasing observation values the scale function rises and likewise, the trend decreases with the reduction of durations. Thus the estimated trend displays the inverse U-shape that is expected for intraday durations. Further, the fitted results of the ACD models with different distributions are in Table 7.5. It is found that the mean and variance existence condition of ACD models, which are $\alpha + \beta < 1$ and $\beta^2 + 2\alpha\beta + 2\alpha^2 < 1$, are satisfied by most models. Meanwhile, from the BIC and MSE values, it also supports that the semiparametric models perform better than the parametric models. Finally, as a difference between the companies, it can be outlined that the data sets give some information about the firms' performance on the stock market, hence also about the size of the companies. The highest trading activity and shortest average durations are recognized for large scale firms. The high market capitalization shows that high market shares and high investments are made, which approves the assumption on the firm scale.

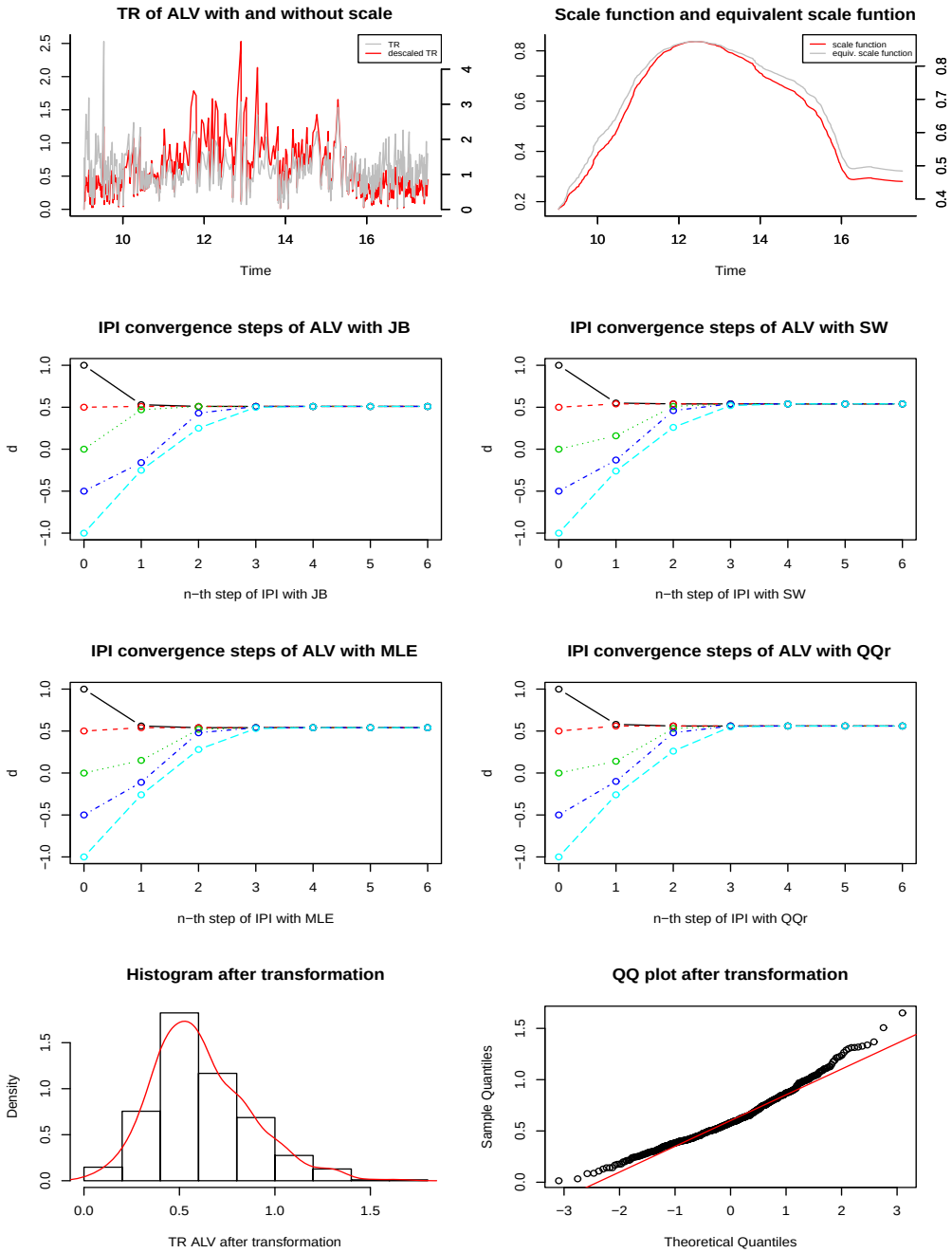

Figure 7.2: The power transformation results of ALV TR on 12 Sep, 2011

Table 7.4: The IPI selected $\hat{\lambda}$ of ALV TR with MLE, JB, SW and QQr

Day	λ_0	MLE	JB	SW	QQr	Day	λ_0	MLE	JB	SW	QQr	Day	λ_0	MLE	JB	SW	QQr	Day	λ_0	MLE	JB	SW	QQr
	1.00	0.54	0.51	0.54	0.56		1.00	0.55	0.57	0.58	0.58		1.00	0.39	0.39	0.38	0.38		1.00	0.47	0.48	0.48	0.48
	0.50	0.54	0.51	0.54	0.56		0.50	0.55	0.57	0.58	0.58		0.50	0.39	0.38	0.38	0.38		0.50	0.47	0.48	0.48	0.48
1	0.00	0.54	0.51	0.54	0.56	6	0.00	0.55	0.57	0.58	0.58	11	0.00	0.39	0.39	0.38	0.38	16	0.00	0.47	0.48	0.48	0.49
	-0.50	0.54	0.51	0.54	0.56		-0.50	0.55	0.57	0.58	0.58		-0.50	0.39	0.39	0.38	0.38		-0.50	0.47	0.48	0.48	0.48
	-1.00	0.54	0.51	0.54	0.56		-1.00	0.55	0.57	0.58	0.58		-1.00	0.39	0.37	0.37	0.37		-1.00	0.47	0.48	0.48	0.48
	1.00	0.44	0.43	0.44	0.44		1.00	0.44	0.45	0.45	0.46		1.00	0.62	0.63	0.63	0.63		1.00	0.40	0.41	0.41	0.41
	0.50	0.44	0.43	0.44	0.44		0.50	0.44	0.45	0.45	0.46		0.50	0.62	0.63	0.63	0.63		0.50	0.40	0.41	0.41	0.41
2	0.00	0.44	0.43	0.44	0.44	7	0.00	0.43	0.45	0.44	0.45	12	0.00	0.62	0.63	0.63	0.63	17	0.00	0.40	0.41	0.41	0.41
	-0.50	0.44	0.43	0.44	0.44		-0.50	0.44	0.45	0.45	0.46		-0.50	0.62	0.63	0.63	0.63		-0.50	0.40	0.41	0.41	0.41
	-1.00	0.44	0.43	0.44	0.44		-1.00	0.44	0.45	0.45	0.46		-1.00	0.62	0.62	0.62	0.62		-1.00	0.40	0.41	0.41	0.41
	1.00	0.34	0.34	0.34	0.34		1.00	0.39	0.40	0.40	0.41		1.00	0.52	0.55	0.56	0.58		1.00	0.47	0.46	0.47	0.46
	0.50	0.34	0.34	0.34	0.34		0.50	0.39	0.40	0.40	0.41		0.50	0.52	0.55	0.56	0.58		0.50	0.47	0.46	0.47	0.46
3	0.00	0.34	0.34	0.34	0.34	8	0.00	0.39	0.40	0.40	0.40	13	0.00	0.52	0.54	0.55	0.57	18	0.00	0.47	0.46	0.47	0.46
	-0.50	0.34	0.34	0.34	0.34		-0.50	0.39	0.40	0.40	0.40		-0.50	0.52	0.54	0.55	0.57		-0.50	0.47	0.46	0.47	0.46
	-1.00	0.34	0.34	0.34	0.34		-1.00	0.39	0.40	0.40	0.40		-1.00	0.52	0.54	0.54	0.57		-1.00	0.47	0.47	0.47	0.47
	1.00	0.24	0.19	0.22	0.25		1.00	0.59	0.61	0.61	0.61		1.00	0.42	0.42	0.42	0.42		1.00	0.41	0.41	0.41	0.41
	0.50	0.24	0.19	0.22	0.25		0.50	0.58	0.61	0.61	0.61		0.50	0.42	0.42	0.42	0.42		0.50	0.41	0.41	0.41	0.41
4	0.00	0.24	0.19	0.22	0.25	9	0.00	0.58	0.61	0.61	0.61	14	0.00	0.41	0.42	0.42	0.42	19	0.00	0.41	0.41	0.41	0.41
	-0.50	0.24	0.19	0.22	0.25		-0.50	0.58	0.61	0.61	0.61		-0.50	0.41	0.42	0.42	0.42		-0.50	0.41	0.41	0.41	0.41
	-1.00	0.24	0.19	0.22	0.25		-1.00	0.58	0.61	0.61	0.61		-1.00	0.41	0.42	0.42	0.42		-1.00	0.41	0.41	0.41	0.41
	1.00	0.29	0.29	0.29	0.27		1.00	0.32	0.31	0.32	0.31		1.00	0.50	0.51	0.51	0.51		1.00	0.44	0.44	0.44	0.43
	0.50	0.29	0.29	0.29	0.27		0.50	0.32	0.31	0.32	0.31		0.50	0.50	0.51	0.51	0.51		0.50	0.44	0.44	0.44	0.43
5	0.00	0.29	0.29	0.29	0.27	10	0.00	0.32	0.31	0.32	0.31	15	0.00	0.50	0.51	0.51	0.51	20	0.00	0.44	0.44	0.44	0.43
	-0.50	0.29	0.29	0.29	0.27		-0.50	0.32	0.31	0.32	0.31		-0.50	0.50	0.51	0.51	0.51		-0.50	0.44	0.44	0.44	0.43
	-1.00	0.29	0.29	0.29	0.27		-1.00	0.32	0.31	0.32	0.31		-1.00	0.50	0.51	0.51	0.51		-1.00	0.44	0.44	0.44	0.43

Table 7.5: The parametric and SemiBC fitting results of TR with the order $(1, 1)$

Data	Model	Para-								SemiBC-							
		ω	α_1	β_1	γ	κ	σ^2	BIC	MSE	ω	α_1	β_1	γ	κ	σ^2	BIC	MSE
ALV	BACD	0.01	0.23	0.75	-	2.07	0.11	-127.32	0.07	0.46	0.33	0.21	-	2.19	0.15	743.74	0.27
	EACD	0.01	0.25	0.72	-	-	-	125.83	0.08	0.45	0.35	0.21	-	-	-	1020.92	0.28
	GACD	0.01	0.20	0.78	1.95	0.97	-	-123.05	0.07	0.42	0.30	0.28	1.80	1.15	-	755.45	0.27
	WACD	0.01	0.20	0.78	1.91	-	-	-129.25	0.07	0.41	0.29	0.30	1.94	-	-	749.94	0.27
BMW	BACD	0.17	0.26	0.57	-	1.63	0.24	867.83	0.50	0.31	0.27	0.43	-	1.60	0.18	930.27	0.50
	EACD	0.17	0.25	0.57	-	-	-	948.98	0.50	0.32	0.26	0.42	-	-	-	1015.46	0.50
	GACD	0.17	0.25	0.58	1.05	1.65	-	874.06	0.50	0.34	0.25	0.42	1.18	1.39	-	934.23	0.50
	WACD	0.17	0.23	0.60	1.39		-	873.28	0.49	0.36	0.24	0.41	1.43		-	930.31	0.50
DBK	BACD	0.02	0.33	0.59	-	2.74	0.29	-950.07	0.01	0.37	0.31	0.33	-	2.85	0.29	598.62	0.20
	EACD	0.02	0.34	0.57	-	-	-	-553.08	0.01	0.40	0.33	0.27	-	-	-	1027.57	0.20
	GACD	0.02	0.30	0.62	1.48	2.08	-	-935.17	0.01	0.36	0.29	0.36	1.55	2.05	-	615.41	0.20
	WACD	0.01	0.25	0.68	2.20	-	-	-928.89	0.01	0.30	0.23	0.47	2.28		-	622.13	0.20
SIE	BACD	0.02	0.20	0.77	-	1.89	0.03	60.02	0.10	0.53	0.24	0.23	-	1.99	0.04	771.31	0.29
	EACD	0.02	0.20	0.76	-	-	-	284.19	0.10	0.59	0.25	0.17	-	-	-	1029.69	0.29
	GACD	0.02	0.20	0.77	1.88	0.98	-	60.19	0.10	0.50	0.23	0.26	2.02	0.94	-	771.51	0.29
	WACD	0.02	0.20	0.77	1.85	-	-	53.96	0.10	0.51	0.23	0.26	1.95	-	-	765.37	0.29

7.4.4 The analysis of the intraday realized volatility

Take the transaction day on September 26th, 2011 of Allianz as an example. The smoothing results are given in below Fig. 7.5. In the figure, it displays the average realized volatility of the Allianz. The x-axis defines the time of the selected trading day, whereas the y-axis shows the value of the realized volatility. It is interesting that the given time series possesses a very high value at the very beginning of the trading day. It is, however, a typical pattern for the RV, which starts very high, reduces slightly during the day, and increases again at the end of the trading day. The typical pattern of the RV is also known as the 'volatility-smile' and is in a U-curve shape. Therefore, Fig. 7.5 shows a common RV process, with very high value at around 9:00 when the market begins and reduces quickly to almost zero right after. During the day, the RV is constant with some peaks, such as at around 9:25, 9:50, 13:45 and 15:00. At 16:00, when is very close to the end of the trading day, the RV increases again. The figure reveals well the time-varying characteristics of the volatility. Besides, the figure displays the average RV together with scale function, which fits perfectly with the course of the time series. The transformed realized volatilities with MLE is also given. The level of the RV values has fallen dramatically, however, the average RV values seem to be more stable. Due to the changing values of the RV, a volatility cluster-effect can be recognized and the scale function fits perfectly with the transformed RV. Furthermore, the figure displays the standardized realized volatility after removing the trend in red. Here, the RV looks very similar to the daily average RV, only with smaller values, but the outliers are exactly the same. The values of the RV range from 1 to 8, but are almost constant at around 0.7. The standardized RV also seems to be stationary with clear clusters, except for some outliers.

An IPI algorithm is carried out in the selection of the power transformation parameter λ. The estimated λ is then used as the Box-Cox power transformation parameter in the scale function of the SemiMEM model. According to Zhang et al. (2017), a fixed λ is crucial in the financial market. In Fig. 7.5, the IPI processes with JB, MLE, SW and QQr are displayed. In the selection, different initial λ values, such as -1, -0.5, 0, 0.5, and 1 are set and moreover, a six-step IPI

process is applied. The used data set of Allianz on September 26th, 2011 converges to its final λ value with MLE after the 3rd IPI procedure and the rest methods are after the 2nd procedure to reach the convergence. The selected λ values of MLE starting from different initial inputs reach the same final stable value, hence $\lambda = 0.159$. The convergence value is also found through the other three methods and the convergence value is very close to that selected by MLE. The discussed data set is then transformed by the Box-Cox transformation. With the selected λ, the power transformed RV is very close to the normal distribution from the histogram plot and Q-Q plot.

The scale function of RV on all transaction days are displayed. It is shown that the U-curve still exists for the RV data. The results of IPI procedures are shown in Table 7.6, which indicate that the stable power parameter λ reaches quickly in the very first few procedures. The fitting results with different ACD $(1, 1)$ models are listed in Table 7.7. In the table, we can see all the fitted models are stationary with the sum of all the coefficients are smaller than one, e.g. the sum of the WACD model of SIE is $0.26 + 0.48 = 0.74 < 1$, the sum of the SemiEACD model $0.28 + 0.29 = 0.57 < 1$ is also smaller than one. Further, it is also discovered that, in most cases, the semiparametric ACD models have much smaller BIC and MSE values than the parametric models, which has proved the advantage of the semiparametric process. For example, the BIC and MSE of the two models are 3167.72, 74.58 and 867.25, 0.62, respectively.

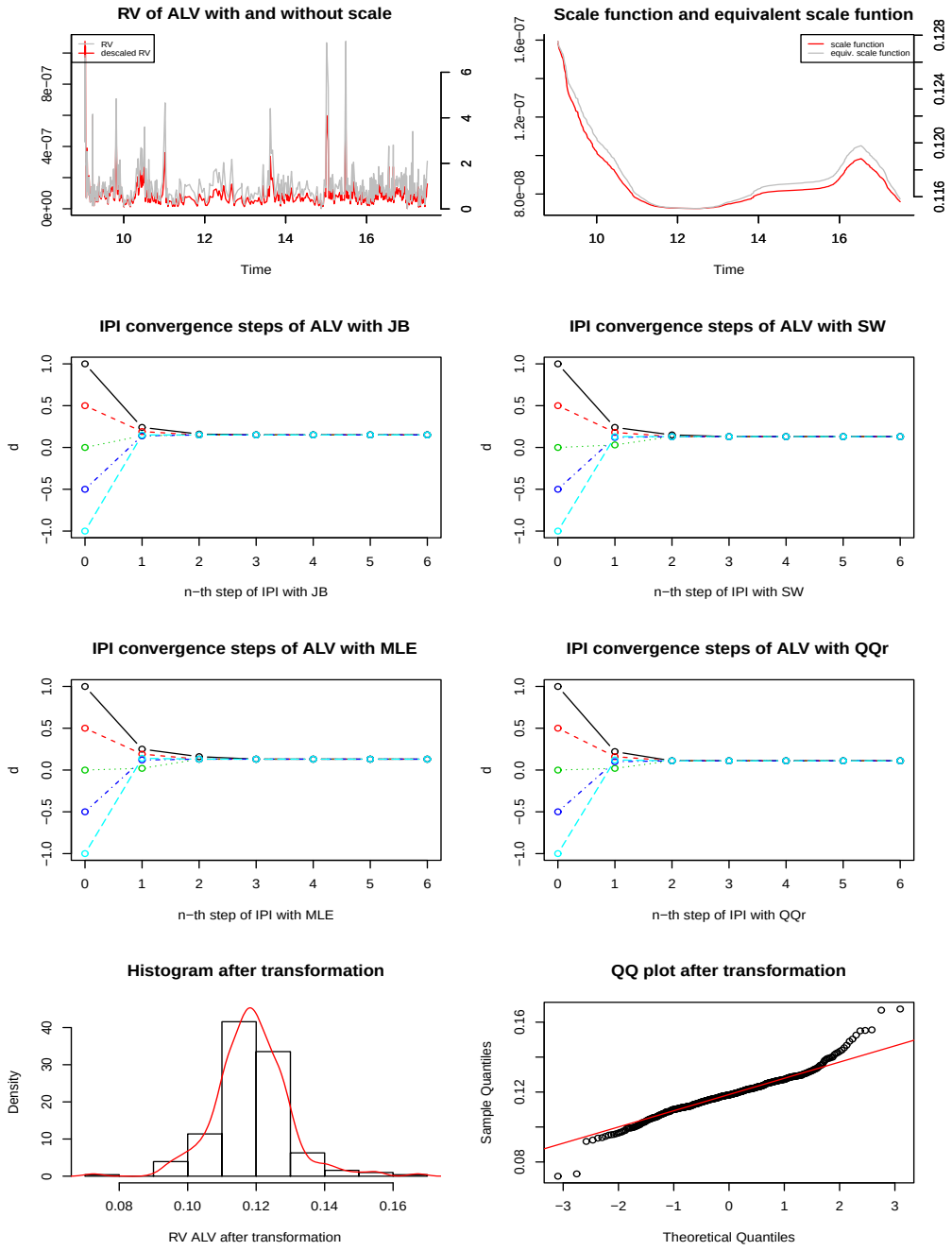

Figure 7.3: The power transformation results of ALV RV on 26 Sep, 2011

Table 7.6: The IPI selected $\hat{\lambda}$ of ALV RV with MLE, JB, SW and QQr

Day	λ_0	MLE	JB	SW	QQr	Day	λ_0	MLE	JB	SW	QQr	Day	λ_0	MLE	JB	SW	QQr	Day	λ_0	MLE	JB	SW	QQr
	1.00	0.26	0.14	0.21	0.26		1.00	0.24	0.09	0.18	0.27		1.00	0.13	0.15	0.13	0.11		1.00	0.15	0.09	0.12	0.14
	0.50	0.26	0.14	0.21	0.26		0.50	0.24	0.09	0.18	0.27		0.50	0.13	0.15	0.13	0.11		0.50	0.15	0.09	0.12	0.14
1	0.00	0.26	0.14	0.21	0.26	6	0.00	0.23	0.08	0.17	0.26	11	0.00	0.13	0.15	0.13	0.11	16	0.00	0.15	0.09	0.12	0.14
	-0.50	0.26	0.14	0.21	0.26		-0.50	0.23	0.08	0.17	0.26		-0.50	0.13	0.15	0.13	0.11		-0.50	0.15	0.09	0.12	0.14
	-1.00	0.26	0.14	0.21	0.26		-1.00	0.23	0.08	0.17	0.26		-1.00	0.13	0.15	0.13	0.11		-1.00	0.15	0.09	0.12	0.14
	1.00	0.22	0.11	0.18	0.27		1.00	0.43	0.43	0.43	0.43		1.00	0.07	-0.01	0.03	0.06		1.00	0.24	0.19	0.23	0.25
	0.50	0.22	0.11	0.18	0.27		0.50	0.43	0.43	0.43	0.43		0.50	0.07	-0.01	0.03	0.06		0.50	0.24	0.19	0.23	0.25
2	0.00	0.22	0.11	0.18	0.27	7	0.00	0.43	0.43	0.43	0.43	12	0.00	0.07	0.00	0.03	0.06	17	0.00	0.24	0.19	0.22	0.25
	-0.50	0.22	0.11	0.18	0.27		-0.50	0.43	0.43	0.43	0.43		-0.50	0.07	0.00	0.03	0.06		-0.50	0.24	0.19	0.22	0.25
	-1.00	0.22	0.11	0.18	0.27		-1.00	0.43	0.43	0.43	0.43		-1.00	0.07	0.00	0.03	0.06		-1.00	0.24	0.19	0.22	0.25
	1.00	0.14	0.03	0.09	0.14		1.00	0.44	0.44	0.44	0.44		1.00	-0.01	-0.07	-0.03	0.00		1.00	0.03	-0.01	0.01	0.02
	0.50	0.14	0.03	0.09	0.14		0.50	0.44	0.44	0.44	0.44		0.50	-0.01	-0.07	-0.03	0.00		0.50	0.03	-0.01	0.01	0.02
3	0.00	0.14	0.02	0.09	0.14	8	0.00	0.44	0.44	0.44	0.44	13	0.00	-0.01	-0.07	-0.03	0.00	18	0.00	0.03	-0.01	0.01	0.02
	-0.50	0.14	0.02	0.09	0.14		-0.50	0.44	0.44	0.44	0.44		-0.50	-0.01	-0.07	-0.03	0.00		-0.50	0.03	-0.01	0.01	0.02
	-1.00	0.14	0.02	0.09	0.14		-1.00	0.44	0.44	0.44	0.44		-1.00	-0.01	-0.07	-0.03	0.00		-1.00	0.03	-0.01	0.01	0.02
	1.00	0.01	-0.05	-0.01	0.02		1.00	0.24	0.12	0.20	0.26		1.00	0.29	0.24	0.27	0.45		1.00	-0.13	-0.19	-0.15	-0.10
	0.50	0.01	-0.05	-0.01	0.02		0.50	0.24	0.12	0.20	0.26		0.50	0.29	0.24	0.27	0.44		0.50	-0.13	-0.19	-0.15	-0.10
4	0.00	0.01	-0.05	-0.01	0.02	9	0.00	0.24	0.12	0.19	0.26	14	0.00	0.29	0.24	0.27	0.43	19	0.00	-0.14	-0.19	-0.15	-0.11
	-0.50	0.01	-0.05	-0.01	0.02		-0.50	0.24	0.12	0.19	0.26		-0.50	0.29	0.24	0.27	0.43		-0.50	-0.14	-0.19	-0.15	-0.11
	-1.00	0.01	-0.05	-0.01	0.02		-1.00	0.24	0.12	0.19	0.26		-1.00	0.29	0.24	0.27	0.43		-1.00	-0.14	-0.19	-0.15	-0.11
	1.00	0.08	0.01	0.05	0.09		1.00	0.15	0.10	0.13	0.13		1.00	0.38	0.37	0.38	0.39		1.00	-0.08	-0.09	-0.09	-0.10
	0.50	0.08	0.01	0.05	0.09		0.50	0.15	0.10	0.13	0.13		0.50	0.38	0.37	0.38	0.39		0.50	-0.08	-0.09	-0.09	-0.10
5	0.00	0.08	0.01	0.05	0.09	10	0.00	0.15	0.10	0.12	0.13	15	0.00	0.38	0.37	0.38	0.39	20	0.00	-0.08	-0.09	-0.09	-0.10
	-0.50	0.08	0.01	0.05	0.09		-0.50	0.15	0.10	0.12	0.13		-0.50	0.38	0.37	0.38	0.39		-0.50	-0.08	-0.09	-0.09	-0.10
	-1.00	0.08	0.01	0.05	0.09		-1.00	0.15	0.10	0.12	0.13		-1.00	0.38	0.37	0.38	0.39		-1.00	-0.08	-0.09	-0.09	-0.10

Table 7.7: The parametric and SemiBC fitting results of RV with the order $(1, 1)$

Data	Model	Para-								SemiBC-							
		ω	α_1	β_1	γ	κ	σ^2	BIC	MSE	ω	α_1	β_1	γ	κ	σ^2	BIC	MSE
ALV	BACD	1.92	0.28	0.51	-	2.69	0.71	2998.44	74.51	0.30	0.27	0.42	-	2.69	0.68	731.76	0.62
	EACD	2.13	0.28	0.49	-	-	-	3278.84	74.47	0.36	0.28	0.35	-	-	-	1010.94	0.62
	GACD	1.90	0.28	0.51	0.31	25.71	-	3046.71	74.55	0.31	0.28	0.41	0.41	15.52	-	773.49	0.62
	WACD	2.31	0.26	0.48	1.41	-	-	3167.72	74.58	0.42	0.28	0.29	1.51	-	-	867.25	0.62
BMW	BACD	0.40	0.24	0.72	-	2.10	0.45	3074.18	122.56	0.24	0.22	0.53	-	2.12	0.40	805.81	1.19
	EACD	0.54	0.34	0.61	-	-	-	3251.42	125.94	0.29	0.33	0.38	-	-	-	996.40	1.22
	GACD	0.52	0.30	0.65	0.55	6.85	-	3103.06	124.20	0.29	0.30	0.40	0.69	4.73	-	829.89	1.21
	WACD	0.57	0.39	0.56	1.39	-	-	3159.49	128.46	0.28	0.40	0.32	1.51	-	-	866.96	1.26
DBK	BACD	2.34	0.34	0.46	-	3.22	0.90	3185.06	136.46	0.25	0.34	0.40	-	3.09	0.78	643.83	0.57
	EACD	1.91	0.38	0.46	-	-	-	3535.86	137.59	0.23	0.37	0.39	-	-	-	992.88	0.57
	GACD	2.11	0.34	0.48	0.07	617.03	-	3206.72	137.05	0.24	0.34	0.42	0.22	69.75	-	660.09	0.57
	WACD	1.71	0.39	0.46	1.59	-	-	3347.55	137.82	0.23	0.39	0.37	1.73	-	-	757.87	0.57
SIE	BACD	1.25	0.24	0.58	-	3.02	0.69	2664.26	34.19	0.37	0.22	0.40	-	3.07	0.66	642.14	0.58
	EACD	0.88	0.23	0.65	-	-	-	3027.49	34.33	0.20	0.19	0.61	-	-	-	1024.61	0.58
	GACD	1.13	0.24	0.60	0.19	95.90	-	2691.33	34.23	0.31	0.22	0.47	0.22	75.21	-	673.35	0.58
	WACD	0.60	0.21	0.70	1.59	-	-	2837.92	34.58	0.10	0.17	0.73	1.61	-	-	824.25	0.59

7.5 Final remarks

In this section, we introduce first a new data process scheme k-method to make sure the high-frequency non-negative financial data of different firms shares the same length as 511 per day. Then, the selected data sets are applied under the Box-Cox SemiMEM framework with ACD models following different distributions. Furthermore, by implementing a data-driven algorithm in the transformation parameter selection, a great improvement in the scale function estimation of Semi-ACD models becomes evident. The power parameter λ, which is applied in the scale function estimation, is obtained after the six steps IPI searching process via JB, MLE, SW and QQr criteria, so as to make the considered data set are close to the normal distribution, combing with the reduction to the moment requirements. In the IPI procedure, the selected λ tends to a convergence value and it is treated as the power in the Box-Cox SemiMEM models. Further, some daily patterns of the scale function are also discovered, such as the U-curve of VO and RV, the inverse U-curve of TR. From the fitting results, we conclude that the semiparametric models have smaller BIC and MSE values than the parametric ones, proving the good performance of the Box-Cox SemiMEM models in practice.

Chapter 8

Concluding remarks

Following the Basel III and its forthcoming finalization, it is nowadays concerned to analyze and predict the individual and market financial performance, so as to reduce the chances of the unnecessary loss, increase the value of firms and minimize the risk in competitive markets. The dissertation provides a comprehensive overview of the financial time series theory, revealing the hidden laws from the market data and supporting the decision making under the Basel framework. In the study, we have found that the semiparametric models perform well in practice and can be used as a supplement of parametric models in risk management.

In Chapter 3, the framework of general SemiGARCH models is set up by introducing a time-varying trend to present the short-term and long-term market performance by daily transaction data. The scale function reveals the long-term risk component, while the classical parametric GARCH models express the short-term market risk. After removing the scale function, the restriction on the parametric GARCH models do not exist anymore and the general SemiGARCH framework requires no assumption on the parametric part, implying parametric model free. Besides, a power transformation is put forward to reduce the moments requirement of the GARCH models. An IPI algorithm is carried out to estimate the power parameter, reaching a convergence value.

Following Basel III and its coming finalization, we examine the VaR and ES prediction and backtesting with the parametric models and the semiparametric models in Chapter 4. The important innovation is that a traffic light test of ES is carried out by introducing breach indicators. In the empirical research, the ES backtesting works well and it indeed provides a simple and direct method for the

ES backtesting evaluation. Besides, some practical cases are found to support the semiparametric models and they can reveal the market risk reasonably, satisfying both the regulators and firms, which proves that the semiparametric models are necessary risk management tools as a supplement of the parametric models.

The general Box-Cox SemiGARCH framework is applied to the high-frequency data of BMW and Allianz in Chapter 5. The data selected in this Chapter is the high-frequency data happens at the same fixed time point on each considered transaction day. Some GARCH models extensions are applied with a time-varying trend, setting up such as the SemiAPARCH model, the SemiEGARCH model and the SemiCGARCH model. In the empirical part, it is found that the selected fixed time point data share similar performance as the daily data.

A general Box-Cox SemiMEM model is provided to analyze the non-negative data, such as MD, AR, VO and TR in Chapter 6. The general SemiMEM models nest the general SemiGARCH models if the squared returns are considered. In the IPI process, we found that different types of data have different features of the power parameter. Furthermore, a simulated confidence interval of the estimated power parameter is calculated via the block bootstrap method without any model or distribution assumption. Then, IPI algorithms of the correlation factor selection are also developed. The general SemiMEM framework greatly expands the scope to some non-negative financial data.

Finally, some open questions and further research topics are still under consideration. First, the long memory parameter can be introduced into the general SemiGARCH and the general SemiMEM framework to build up the long memory models, such as the Semi-FI-MEM and Semi-FI-Log-MEM models. Then, the power transformation technique should be improved, due to the restriction of Box-Cox transformation on obtaining a complete normal distribution. Instead, the other distribution, such as sin-arcsin, sinh-normal and Birnbaum-Saunders distributions should be considered. Further, spatial models (like in Peitz, 2015) can be considered under the general semiparametric framework. Finally, new methodologies of VaR and ES backtesting should be put forward, such as simulating new test statistics by the Monte Carlo simulation.

References

Abad, P., S. Benito and C. Lopez (2015). The role of the loss functions in value at risk comparison. *The Journal of Risk Model Valuation*, 9, 1–19.

Acerbi, C. and B. Szekely (2014). Backtesting expected shortfall. *Risk*, 32, 1404–1415.

Altman, N. (1990). Kernel smoothing with correlated errors. *Journal of the American Statistical Association*, 85, 749–759.

Altman, N. and C. Leger (1995). Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, 46, 195–214.

Amado, C. and T. Teräsvirta (2017). Specification and testing of multiplicative time-varying GARCH models with applications. *Econometric Reviews*, 36, 421–446.

Angelidis, T. and S. Degiannakis (2007). Backtesting VaR models: a two-stage procedure. *The Journal of Risk Model Validation*, 1, 27–48.

Artzner, P., F. Delbaen, J. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance*, 9, 203–228.

Asar, Ö., I. Ozlem and D. Osman (2017) Estimating Box-Cox power transformation parameter via goodness-of-fit tests. *Communications in Statistics - Simulation and Computation*, 4, 91–105.

Baillie, R., T. Bollerslev and H. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74, 3–30.

Basel Committee on Banking Supervision (1996). Supervisory framework for the use of back-testing in conjunction with the internal models approach to market risk capital requirements.

Basel Committee on Banking Supervision (2012). Fundamental review of the trading book: A revised market risk framework.

Basel Committee on Banking Supervision (2016). Minimum capital requirements for market risk.

Bauwens, L., F. Galli and P. Giot (2008). The moments of Log-ACD models. *Quantitative and Qualitative Analysis in Social Sciences*, 2, 1–28.

Bauwens, L. and P. Giot (2000). The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. *Annales d'Économie et de Statistique*, 60, 117–149.

Beran, J. and Y. Feng (2002). Iterative plug-in algorithms for SEMIFAR models - definition, convergence and asymptotic properties. *Journal of Computational and Graphical Statistics*, 11, 690–713.

Beran, J. and Y. Feng (2002a). Local polynomial fitting with long memory, short memory and antipersistent errors. *Annals of the Institute of Statisitcal Mathematics*, 54, 291–311.

Beran, J. and Y. Feng (2002b). SEMIFAR models – a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis*, 40, 393–419.

Beran, J., Y. Feng and S. Ghosh (2015). Modelling long-range dependence and trends in duration series: an approach based on EFARIMA and ESEMIFAR models. *Statistical Papers*, 56, 431–451.

Beran, J. and D. Ocker (2001). Volatility of stock-market indexes–An analysis based on SEMIFAR models. *Journal of Business & Economic Statistics*, 19, 103–116.

Black, F. (1976). Studies of Stock Price Volatility Changes. In *Proceedings of the 1976 Meetings of the Business and Economics Statistics Section*. American Statistical Association, 177–181.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327.

Box, G. and D. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, Series B, 26, 211–252.

Brockmann, M., T. Gasser and E. Herrmann (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association*, 88, 1302–1309.

Bühlmann, P. (1996). Locally adaptive lag-window spectral estimation. *Journal Time Series Analysis*, 17, 247–270.

Bühlmann, P. (2002). Bootstrap for time series. *Statistical Science*, 17, 52–72.

Bühlmann, P. and H. Künsch (1999). Block length selection in the bootstrap for time series. *Computational Statistics and Data Analysis*, 31, 295–310.

Caporin, M. (2008). Evaluating value-at-risk measures in the presence of long memory conditional volatility. *The Journal of Risk*, 10, 79–110.

Caporin, M., E. Rossi and P. de Magistris (2017). Chasing volatility. *Journal of Econometrics*, 198, 122–145.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39, 841–862.

Christoffersen, P. (2003). *Elements of financial risk management*. Amsterdam: Academic Press.

Conrad, C. (2006). GARCH models with long memory and nonparametric specification. PhD Dissertation, University of Mannheim.

Conrad, C. and M. Karanasos (2006). The impulse response function of the long memory GARCH process. *Economics Letters*, 90, 34–41.

Costanzino, N. and M. Curran (2018). A simple traffic light approach to backtesting expected shortfall. *Risks*, 6, 1–7.

Dacorogna, M., U. Müller, R. Nagler, R. Olsen and O. Pictet (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance*, 12, 413–438.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25, 1–37.

Dahlhaus, R. and S. Rao (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics*, 34, 1075–1114.

Di, J. and A. Gangopadhyay (2011). On the efficiency of a semi-parametric GARCH model. *The Econometrics Journal*, 14, 257–277.

Ding, Z. and C. Granger (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 73, 185–215.

Ding, Z., C. Granger and R. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1, 83–106.

Dong, Y. and Y. Tse (2014). Business time sampling scheme and its application to semi-martingale hypothesis and estimating integrated volatility. *Research Collection School Of Economics*, 12–2014.

Dufour, A and R. Engle (2000). Time and the impact of a trade. *Journal of Finance*, 55, 2467–2489.

Eagleson, G. and H. Müller (1997). Transformations for smooth regression models with multiplicative errors. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59, 173–189.

Efromovich, S. (1999) *Nonparametric curve estimation: methods, theory, and applications*. Springer, New York.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1–26.

Efron, B. (2003). Second thoughts on the bootstrap. *Statistical Science*, 18, 135–140.

Efron, B. and R. Tibshirani (1993). *An introduction to the bootstrap*. Chapman and Hall, New York.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimation of U.K. inflation. *Econometrica*, 50, 987–1008.

Engle, R. (2002). New frontiers for arch models. *Journal of Applied Econometrics*, 17, 425–446.

Engle, R., E. Ghysels and B. Sohn (2008). On the economic sources of stock market volatility. Discussion Paper of NYU and UNC.

Engle, R. and G. Lee (1999). A permanent and transitory component model of stock return volatility. *Cointegration, Causality, and Forecasting: a Festschrift in Honour of Clive W. J. Granger*, 20, 475–497. Oxford: Oxford University Press.

Engle, R. and J. Rangel (2008). The Spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, 21, 1187–1222.

Engle, R. and J. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 66, 1127–1162.

Eubank, R. (1993). Applied nonparametric regression. *Technometrics*, 35, 225–226.

Fama, E. (1965). The behavior of stock-market prices. *Journal of Business*, 38, 34–105.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, 21, 196–216.

Fan, J. and T. Gasser and I. Gijbels (1997). Local polynomial regression: Optimal kernels and asymptotic minimax effciency. *The Annals of Statistics*, 49, 79–99.

Fan, J. and Q. Yao (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85, 645–60.

Feng, Y. (2004). Simultaneously modeling conditional heteroskedasticity and scale change. *Econometric Theory*, 20, 563–596.

Feng, Y. (2013). An iterative plug-in algorithm for decomposing seasonal time series using the Berlin Method. *Journal of Applied Statistics*, 40, 266–281.

Feng, Y. (2014). Data-driven estimation of diurnal patterns of durations between trades on financial markets. *Statistics & Probability Letters*, 92, 109–113.

Feng, Y. (2019). Semiparametric GARCH models for value at risk and expected shortfall–an object-driven procedure. *Preprint April 2019*.

Feng, Y. and T. Gries (2017). Data-driven local polynomial for the trend and its derivatives in economic time series. *CIE WP*, 2017-04, Paderborn University.

Feng, Y., T. Gries and M. Fritz (2019). Data-driven local polynomial for the trend and its derivatives in economic time series. *CIE WP*, 2019-06, Paderborn University.

Feng, Y and S. Heiler (1998). Locally weighted autoregression, in: R. Galata and H. Küchenhoff (Eds.) *Econometrics in theory and practice*, Physica-Verlag, Heidelberg, 101–117.

Feng, Y. and A. McNeil (2008). Modelling of scale change, periodicity and conditional heteroskedasticity in return volatility. *Economic Modelling*, 25, 850–867.

Feng, Y. and L. Sun (2013). A Semi-APARCH approach for comparing long-term and short-term risk in Chinese financial market and in mature financial markets. *CIE WP*, 2013-12, Paderborn University.

Feng, Y. and C. Zhou (2015). Forecasting financial market activity using a semiparametric fractionally integrated Log-ACD. *International Journal of Forecasting*, 31, 349–363.

Fernandes, M. and J. Gramming (2006). A flexible family of autoregressive conditional duration models. *Journal of Econometrics*, 130, 1–23.

Gasser, T., A. Kneip and W. Köhler (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association*, 86, 643–652.

Gasser, T. and H. Müller (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 11, 171–185.

Ghosh, S and D. Draghicescu (2002). An algorithm for optimal bandwidth selection for smooth nonparametric quantile estimation. In Y. Dodge (Ed.), *Statistics for Industry and Technology: Statistical Data Analysis Based on the L1-Norm and Related Methods*, 161–168, Basel: Birkhäuser.

Glosten, L., R. Jagannathan and D. Runkle (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48, 1779–1801.

Gneiting, T. (2012). Making and evaluating point forecasts. *Journal of the American Statistical Association* 106, 746–762.

Gourieroux, C. and A. Monfort (1992). Qualitative threshold ARCH models. *Journal of Econometrics*, 52, 159–199.

Granger, C. and Z. Ding (1995). Some properties of absolute returns: An alternative measure of risk. *Annales d'Economie et de Statistique*, 10, 67–91.

Haas, M. (2001). New Methods in Backtesting. *Financial Engineering Research Center*, Working paper.

Hall, P. (1992). *The bootstrap and Edgeworth expansion.* Springer, New York.

Hall, P. and B. Presnell (1999). Intentionally-biased bootstrap methods. *Journal of the Royal Statistical Society*, 61, 143–158.

Hampel, F., E. Ronchetti, P. Rousseeuw and W. Stahel (1986). *Robust statistics: The approach based on influence function.* Wiley, New York.

Hart, J. (1991). Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society, Series B*, 53, 173–188.

Harvey, A., E. Ruiz and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies*, 61, 247–264.

Hautsch, N. (2002). Modeling intraday trading activity using Box-Cox ACD models. Discussion Paper 02/05, CoFE, University of Konstanz.

Herrmann, E. and T. Gasser (1994). Iterative plug-in algorithm for bandwidth selection in kernel regression estimation. Preprint, Darmstadt Institute of Technology and University of Zürich.

Herrmann, E., T. Gasser and A. Kneip (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika*, 79, 783–795.

Jasiak, J. (1998). Persistence in intertrade durations. *Finance*, 19, 166–195.

John, J. and N. Draper (1980). An alternative family of transformations. *Journal of the Royal Statistical Society*, 29, 190–97.

Karanasos, M. (2008). The statistical properties of exponential ACD models. *Quantitative and Qualitative Analysis in Social Sciences*, 2, 29–49.

Karanasos, M., Z. Psaradakis and M. Sola (2004). On the autocorrelation properties of long-memory GARCH processes. *Journal of Time Series Analysis*, 25, 265–281.

Künsch, H (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17, 1217–1241.

Kupiec, P. (1995). Techniques for verifying the accuary of risk management models. *Journal of Derivatives*, 3, 73–84.

Lee, S. and B. Hansen (1994). Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator. *Econometric Theory*, 10, 29–52.

Liu, R. and K. Singh (1992). Moving blocks jackknife and bootstrap capture weak dependence. In: R. Lepage and L. Billard, Eds., *Exploring the Limits of Bootstrap*, John Wiley, New York.

Lopez, J. (1999). Methods for evaluating value-at-risk models. *Federal Reserve Bank of San Francisco Economic Review*, 2, 3–17.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36, 394–419.

Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets*, 8, 377–399.

Manly, B. (1976). Exponential data transformations. *Statistician*, 25, 37–42.

McNeil, A., R. Frey and P. Embrechts (2015). *Quantitative risk management: concepts, techniques and tools (Revised edition)*, Princeton University Press.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59, 347–370.

Noguchi, K., A. Aue and P. Burman (2016). Exploratory analysis and modeling of stock returns. *Journal of Computational and Graphical Statistics*, 25, 363-381.

Oomen, R. (2006). Properties of realized variance under alternative sampling schemes. *Journal of Business & Economic Statistics*, 24, 219–237.

Peitz, C. (2015). *Die parametrische und semiparametrische Analyse von Finanzzeitreihen: Neue Methoden, Modelle und Anwendungsmöglichkeiten.* Springer Gabler, Wiesbaden.

Politis, D. and J. Romano (1993). The stationary bootstrap. *Journal of the American Statistical Association*, 89, 1301–1313.

Politis, D. and H. White (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23, 53–70.

Priestley, M. (1981). *Spectral analysis and time series* (Volume 1 and 2). Academic Press, London.

Rahman, M. (1999). Estimating the Box-Cox transformation via Shapiro-Wilk W Statistic, *Communications in Statistics - Simulation and Computation*, 28, 223–241.

Rahman, M. and L. Pearson (2008). Anderson-Darling statistic in estimating the Box-Cox transformation parameter. *Journal of Applied Probability and Statistics*, 3, 45–57.

Ruppert, D., M. Wand, U. Holst and O. Hössjer (1997). Local polynomial variance-function estimation. *Technometrics*, 39, 262–273.

Russell, J. and R. Engle (2002). Econometric analysis of discrete-valued, irregularly-spaced financial transactions data. Revised version of Discussion Paper 98-10, University of California, San Diego.

Russell, J. and R. Engle (2010). Analysis of high-frequency data. In *Handbook of financial econometrics: tools and techniques*, 383–426. North-Holland.

Sarma, M., S. Thomas and A. Shah (2003). Selection of value-at-risk models. *Journal of Forecasting*, 22, 337–358.

Schwert, G. (1989). Why does market volatility change over time. *Journal of Finance*, 5, 1115–1153.

Taylor, S. (1986). *Modelling financial time series*, John Wiley & Sons, New York.

Taylor, N. and Y. Xu (2017). The logarithmic vector multiplicative error model: an application to high frequency NYSE stock data. *Quantitative Finance*, 17, 1021–1035.

Tsai, K., J. Lih and J. Ko (2012). The overnight effect on the Taiwan stock market. *Physica A: Statistical Mechanics and its Applications*, 391, 6497–6505.

Tukey, J. (1958). Bias and confidence in not quite large samples (abstract). *The Annals of Mathematical Statistics*, 29, 614.

Van Bellegem, S. and R. von Sachs (2004). Forecasting economic time series with unconditional time-varying variance. *International Journal of Forecasting*, 20, 611–627.

Wahba, G. and S. Wold (1975). A completely automatic french curve: fitting spline functions by cross validation. *Communications in Statistics*, 4, 1–17.

Xu, X. and S. Taylor (1994). The term structure of volatility implied by foreign exchange options. *Journal of Financial and Quantitative Analysis*, 29, 57–74.

Yeo, I. and R. Johnson (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87, 954–59.

Zakoïan, J. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18, 931–955.

Zhang, X. (2019a). A Box-Cox semiparametric multiplicative error model. *CIE WP*, 2019–05, Paderborn University.

Zhang, X. (2019b). Value at risk and expected shortfall under general semiparametric GARCH models. *CIE WP*, 2019–06, Paderborn University.

Zhang, X, Y. Feng and C. Peitz (2017). A general class of SemiGARCH models based on the Box-Cox transformation. *CIE WP*, 2017–05, Paderborn University.

Appendices

Appendix A: Proofs of the Results

Proof of Lemma 3.1

Let $Z_n = n^{\eta_2/2}\{\tilde{g}(\tau_t,\lambda) - B[\tilde{g}(\tau_t,\lambda) - g(\tau_t,\lambda)\}/\tilde{\sigma}$. Under the assumptions of Lemma 3.1, Z_n is asymptotically standard normal. We have

$$\begin{aligned} P(\tilde{g}(\tau_t,\lambda) < 0) &\leq P[|\tilde{g}(\tau_t,\lambda) - g(\tau_t,\lambda)| > g(\tau_t,\lambda)] \\ &= P\{|\tilde{g}(\tau_t,\lambda) - B[\tilde{g}(\tau_t,\lambda)] - g(\tau_t,\lambda) + B[\tilde{g}(\tau_t,\lambda)]| > g(\tau_t,\lambda)\} \\ &\leq P\{|\tilde{g}(\tau_t,\lambda) - B[\tilde{g}(\tau_t,\lambda)] - g(\tau_t,\lambda)| > m(x) - |B[\tilde{g}(\tau_t,\lambda)|\} \\ &\leq P\{|\tilde{g}(\tau_t,\lambda) - B[\tilde{g}(\tau_t,\lambda)] - g(\tau_t,\lambda)| > g(\tau_t,\lambda)/2\} \end{aligned}$$

if n is large enough. Furthermore, we have

$$P\{|\tilde{g}(\tau_t,\lambda) - B[\tilde{g}(\tau_t,\lambda)] - g(\tau_t,\lambda)| > g(\tau_t,\lambda)/2\} = P\{|Z_n| > n^{\eta_2/2} g(\tau_t,\lambda)/(2\tilde{\sigma})\}.$$

Defining $z_n^o = n^{\eta_2/2} g(\tau_t,\lambda)/(2\tilde{\sigma})$, we have $n = L_g (z_n^o)^{2/\eta_2}$, where $L_g = [2\tilde{\sigma}/g(\tau_t,\lambda)]^{2/\eta_2}$. Furthermore let $Z \sim N(0,1)$. Then

$$\begin{aligned} nP\{|Z_n| > z_n^o\} &\approx nP\{|n| > z_n^o\} \\ &= 2L_g (z_n^o)^{2/\eta_2} \int_{z=z_n^o}^{\infty} e^{-\frac{z^2}{2}} dz \\ &= 2L_g \int_{z=z_n^o}^{\infty} (z_n^o)^{2/\eta_2} e^{-\frac{z^2}{2}} dz \to 0, \end{aligned}$$

as $n \to \infty$, because all moments of Z are finite. Lemma 3.1 is proved. ◇

To prove the results of Theorem 3.1, the following assumptions are required.

A1. The scale function $g(\tau_t,\lambda)$ is strictly positive, bounded, and at least twice

continuously differentiable on [0, 1].

A2. The kernel $K(u)$ is a symmetric density with compact support [-1, 1].

A3. The bandwidth b satisfies $b \to 0$ and $nb \to \infty$ as $n \to \infty$.

A4. $\{\zeta_t\}$ is a stationary process with unit mean and absolutely summable autocovariance.

A5. The stationary process $\{\zeta_t\}$ can be represented as $\zeta_t = 1 + \sum_{i=0}^{\infty} \lambda_i \nu_{t-i}$, where $\{\nu_t\}$ is a sequence of uncorrelated zero-mean innovations with finite variance, $\sum_{i=0}^{\infty} \lambda_i \neq 0$ and $\sum_{i=0}^{\infty} |\lambda_i| < \infty$.

Assumptions A1 to A3 are the regular nonparametric regression conditions. A4 is the requirement of the GARCH model. A5 is a sufficient regularity condition which ensures that the sample means of ζ_t and ξ_t are both asymptotically normal.

Proof of Theorem 3.1

Following Lemma 3.1, we can conclude that $B[\hat{g}(\tau_t, \lambda)] = B[\tilde{g}(\tau_t, \lambda)] + o_p(n^{-1/2})$ and $\text{Var}\,[\hat{g}(\tau_t, \lambda)] = \text{Var}\,[\tilde{g}(\tau_t, \lambda)] + o_p(n^{-1})$. The proof will hence simply be given for the unrestricted local linear estimator $\tilde{g}(\tau_t, \lambda)$.

i) Bias: Since $\tilde{g}(\tau_t, \lambda)$ is a linear smoother, the bias $B[\tilde{g}(\tau_t, \lambda)] = E[\tilde{g}(\tau_t, \lambda)] - g(\tau_t, \lambda)$ is the same as in the nonparametric regression with i.i.d. errors. This is the formula given i).

ii) Variance: The local linear estimator $\tilde{g}(\tau_t, \lambda)$ is a linear estimator $\tilde{g}(\tau_t, \lambda) = \sum_{i=1}^{T} w_i^x y_i$. It is well known that the weights w_i^x are asymptotically equivalent to those defined by the equivalent kernel, i.e.

$$w_i^\tau = \frac{K_{\tau_t}\left(\frac{\tau_i - \tau}{b_\tau}\right)}{\sum_{i=1}^{n} K_\tau\left(\frac{\tau_i - \tau}{b_\tau}\right)}[1 + o(1)] = \frac{1}{nb_\tau} K_\tau\left(\frac{\tau_i - \tau}{b_\tau}\right)[1 + o(1)]$$

for $|\tau_i - \tau| \leq b$ and zero otherwise.

Note that the autocovariances of ξ_t and ζ_t are the same. Furthermore, let K be an integer such that $K \to \infty$ and $K/nb_\tau \to 0$, as $n \to \infty$. For instance, we may choose $K = [\sqrt{nb_\tau}]$ where $[\tau]$ denotes the integer part of τ. Defining $b_K = K/n$

we have $b_K/b_\tau \to 0$ as $n \to \infty$. The variance of $\tilde{g}(\tau_t, \lambda)$ is given by

$$\begin{aligned}
\mathrm{Var}\left[\tilde{g}(\tau_t,\lambda)\right] &= \sum_{|\tau_i-\tau|\le b_\tau} \sum_{|\tau_j-\tau|\le b_\tau} w_i^\tau w_j^\tau \mathrm{Cov}\left[g(\tau_i,\lambda)\xi_i, g(\tau_j,\lambda)\xi_j\right] \\
&= \sum_{|\tau_i-\tau|\le (b_\tau-b_K)} \sum_{|\tau_j-\tau|\le b_\tau} w_i^\tau w_j^\tau \mathrm{Cov}\left[g(\tau_i,\lambda)\xi_i, g(\tau_j,\lambda)\xi_j\right] \\
&+ \sum_{|\tau_i-\tau|> (b_\tau-b_K)} \sum_{|\tau_j-\tau|\le b_\tau} w_i^\tau w_j^\tau \mathrm{Cov}\left[g(\tau_i,\lambda)\xi_i, g(\tau_j,\lambda)\xi_j\right] \qquad (1) \\
&=: V_1 + V_2,
\end{aligned}$$

where V_1 indicates the contribution of the observations in the middle part of the window and V_2 the contribution in the boundary of the window. The definition of K and b_K ensures that $V_2 = o(V_1)$, i.e. $\mathrm{Var}\left[\tilde{g}(\tau_t, \lambda)\right] \approx V_1$. Note that the condition $|\tau_i - \tau| \le b_\tau - b_K$ ensures that τ_i, τ_j with $|i-j| \le K$ are all within the window. This will simplify the analysis in the next part. Denote by V_{1i} the ith sum over τ_j in V_1 for given τ_i . Then

$$\begin{aligned}
V_{1i} &= \sum_{|\tau_j-\tau|\le b_\tau} w_i^\tau w_j^\tau \mathrm{Cov}\left[g(\tau_i,\lambda)\xi_i, g(\tau_j,\lambda)\xi_j\right] \\
&= \sum_{|i-j|\le K} w_i^\tau w_j^\tau \mathrm{Cov}\left[g(\tau_i,\lambda)\xi_i, g(\tau_j,\lambda)\xi_j\right] \\
&+ \sum_{|i-j|> K} w_i^\tau w_j^\tau \mathrm{Cov}\left[g(\tau_i,\lambda)\xi_i, g(\tau_j,\lambda)\xi_j\right] \\
&=: V_{1i}^C + V_{1i}^T, \qquad (2)
\end{aligned}$$

where V_{1i}^C denotes the contribution of the covariances in the central part with lags $|k| \le K$, whereas V_{1i}^T is the contribution of the covariances in the tail part. For the first term in (2) we have

$$\begin{aligned}
V_{1i}^C &= \sum_{|i-j|\le K} w_i^\tau w_j^\tau \mathrm{Cov}\left[g(\tau_i,\lambda)\xi_i, g(\tau_j,\lambda)\xi_j\right] \\
&= \frac{1}{(nb_\tau)^2} K_\tau^2\left(\frac{\tau_i-\tau}{b_\tau}\right)[1+o(1)]g^2(\tau_i,\lambda)[1+O(b_K)]\sum_{|k|\le K}\gamma(k) \\
&\approx \frac{2\pi c_f}{(nb_\tau)^2} K_\tau^2\left(\frac{\tau_i-\tau}{b_\tau}\right) g^2(\tau_i,\lambda). \qquad (3)
\end{aligned}$$

The second term in (2) is asymptotically negligible, because

$$\begin{aligned} V_{1i}^T &= \sum_{|i-j|>K} w_i^\tau w_j^\tau \mathrm{Cov}\left[g(\tau_i,\lambda)\xi_i, g(\tau_j,\lambda)\xi_j\right] \\ &\leq \sum_{|i-j|>K} |w_i^\tau w_j^\tau \mathrm{Cov}\left[g(\tau_i,\lambda)\xi_i, g(\tau_j,\lambda)\xi_j\right]| \\ &\leq \frac{C_i}{(nb_\tau)^2}[1+o(1)] \sum_{|k|>K} |\gamma(k)| \\ &= o\left(\frac{1}{(nb_\tau)^2}\right), \end{aligned} \tag{4}$$

where

$$C_i = \sup_{|i-j|>K} \left| K_\tau\left(\frac{\tau_i-\tau}{b_\tau}\right) K_\tau\left(\frac{\tau_j-\tau}{b_\tau}\right) g(\tau_i,\lambda) g(\tau_j,\lambda)\right|.$$

This leads to

$$\begin{aligned} \mathrm{Var}\left[\tilde{g}(\tau_t,\lambda)\right] &= \left\{ \sum_{|\tau_i-\tau|\leq(b_\tau-b_K)} \sum_{|\tau_j-\tau|\leq b_\tau} w_i^\tau w_j^\tau \mathrm{Cov}\left[g(\tau_i,\lambda)\xi_i, g(\tau_j,\lambda)\xi_j\right] \right\}[1+o(1)] \\ &= \left\{ \sum_{|\tau_i-\tau|\leq(b_\tau-b_K)} \sum_{|i-j|\geq K} V_{1i}^C \right\}[1+o(1)] \\ &= \frac{2\pi c_f g^2(\tau_i,\lambda)}{(nb_\tau)} \left\{ \sum_{|\tau_i-\tau|\leq(b_\tau-b_K)} \frac{1}{(nb_\tau)} K_\tau^2\left(\frac{\tau_i-\tau}{b_\tau}\right) \right\}[1+o(1)] \\ &= \frac{2\pi c_f g^2(\tau_i,\lambda) R(K_\tau)}{(nb_\tau)}[1+o(1)] \end{aligned} \tag{5}$$

as given in Theorem 3.1 ii).

iii) Here a more general result $\sqrt{nb}[\tilde{g}(\tau_t,\lambda) - B[\tilde{g}(\tau_t,\lambda)] - g(\tau_t,\lambda)] \xrightarrow{\mathcal{D}} N(0,V)$ can be proved, with V defined in (3.11). This leads to $\sqrt{nb}[\tilde{g}(\tau_t,\lambda) - g(\tau_t,\lambda)] \xrightarrow{\mathcal{D}} N(0,V)$, when $b_\tau = o(b_A)$, because $\sqrt{nb}B[\tilde{g}(\tau_t,\lambda)] \to 0$. Define $\delta = \tilde{g}(\tau_t,\lambda) - B[\tilde{g}(\tau_t,\lambda)] - g(\tau_t,\lambda)]$. Note that

$$\begin{aligned} \delta(\tau_t) &= \sum_{t=1}^n w_t^\tau [g(\tau_t,\lambda)\xi_t] \\ &= \sum_{t=1}^n w_t^{\tau *} \xi_t, \end{aligned} \tag{6}$$

where $w_t^{\tau *} = w_t^\tau g(\tau_t,\lambda)$. It is easy to check that the regularity conditions (4.2)

and (4.3) in Beran and Feng (2002) are jointly fulfilled by $w_t^{\tau *}$ and λ_t. Hence, following Theorem 1 in Beran and Feng (2002), $\delta(\tau_t)$ is asymptotically normal, if the sample mean of ξ_t is. The latter is guaranteed by A1 and A5. Hence, Theorem 3.1 follows. $\diamond$

Proof of Theorem 3.2

i) The formula for the MSE of $\tilde{g}(\tau_t, \lambda)$ is the sum of the square bias and the variance. The bias and variance follow i) and ii) in Theorem 3.1.

ii) The MISE can be calculated on the whole support $[0, 1]$, because the contribution of the estimated values in the boundary area is asymptotically negligible. $\diamond$

Proof of Theorem 6.1

Define $b_A = C_{\mathrm{A}} n^{-1/5}$, where C_{A} is the constant in b_A. We have $\hat{b} = \hat{C}_{\mathrm{A}} n^{-1/5}$ and

$$(\hat{b} - b_{\mathrm{A}})/b_{\mathrm{A}} = C_{\mathrm{A}}^{-1}(\hat{C}_{\mathrm{A}} - C_{\mathrm{A}}). \tag{7}$$

Taylor expansion leads to

$$\hat{C}_{\mathrm{A}} - C_{\mathrm{A}} \doteq O_p(\hat{I}(g_\lambda^2) - I(g_\lambda^2)) + O_p(\hat{I}((g_\lambda'')^2) - I((g_\lambda'')^2)) + O(\hat{c}_f - c_f). \tag{8}$$

It is well known that

$$\hat{I}((g_\lambda'')^2) - I((g_\lambda'')^2) \doteq O_p(n^{-2/7}). \tag{9}$$

Following the results in chapter 6.2 of Priestley (1981), the error in the lag-window estimator of c_f using the Bartlett-window and a bandwidth $K = O(n^{1/3})$ is

$$\hat{c}_f - c_f = O_p(n^{-1/3}). \tag{10}$$

Moreover, the first term $O_p(\hat{I}(g_\lambda^2) - I(g_\lambda^2))$ is neglectable (see Feng, 2004). $\diamond$

Appendix B: POT plots of VaR and ES

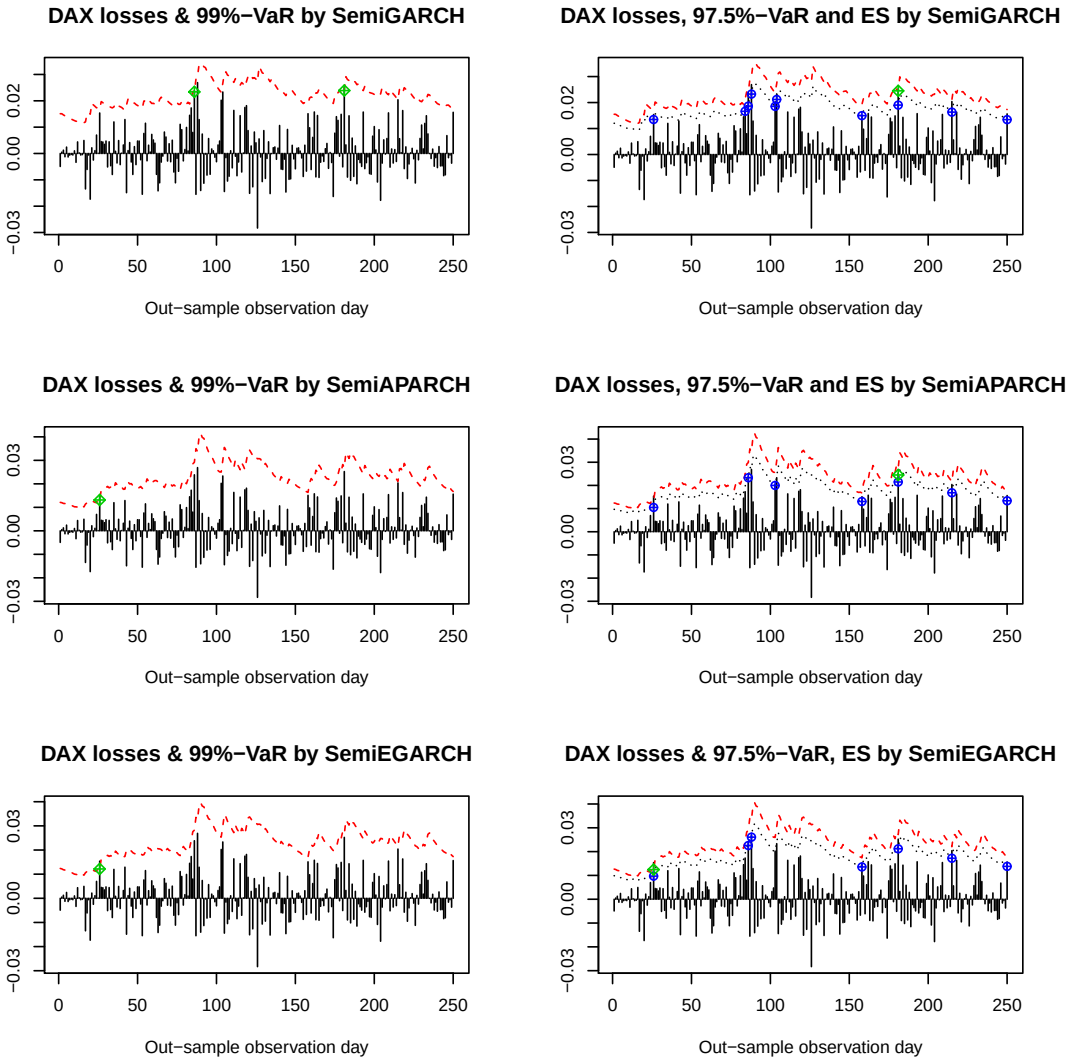

Figure A.1: DAX POT of VaR and ES with semiparametric models

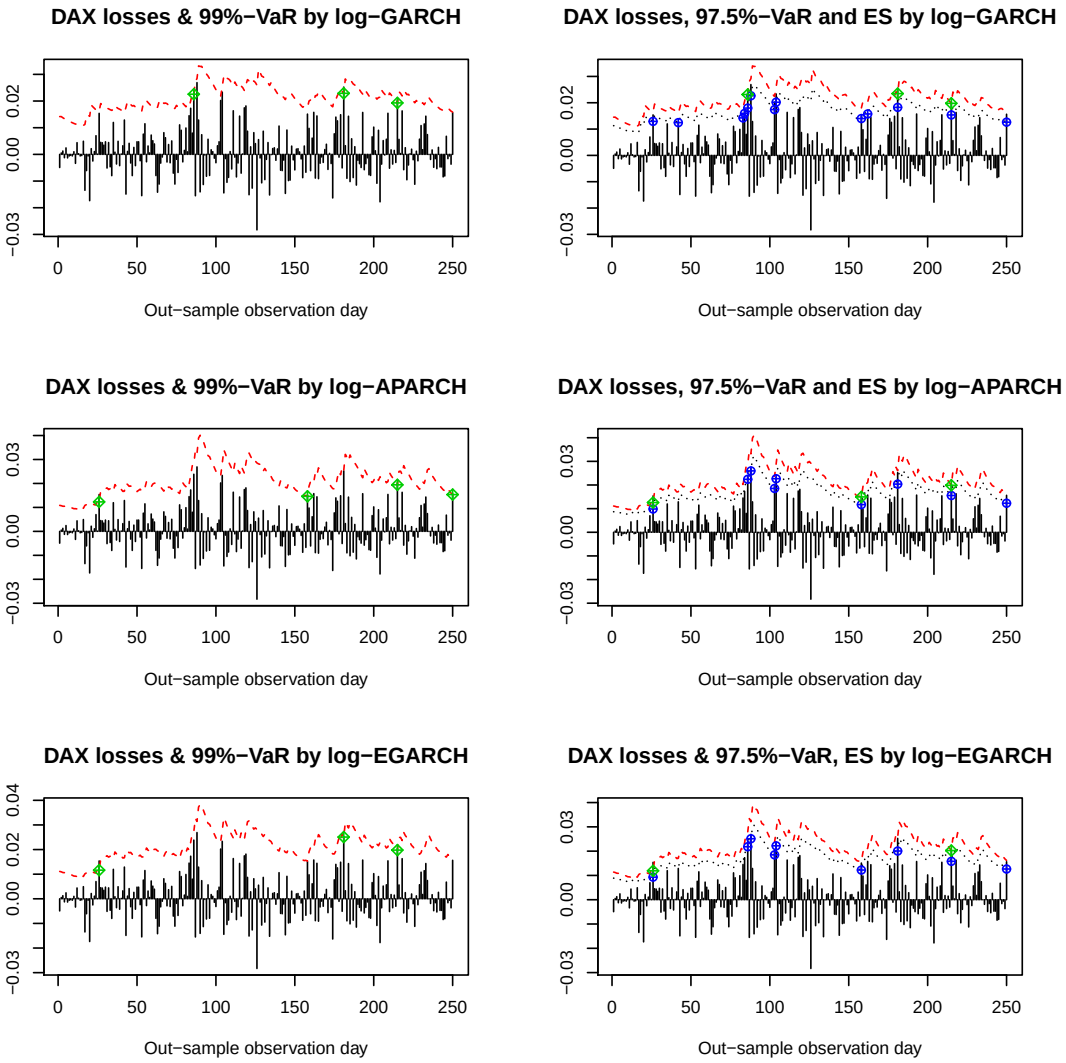

Figure A.2: DAX POT of VaR and ES with log-models

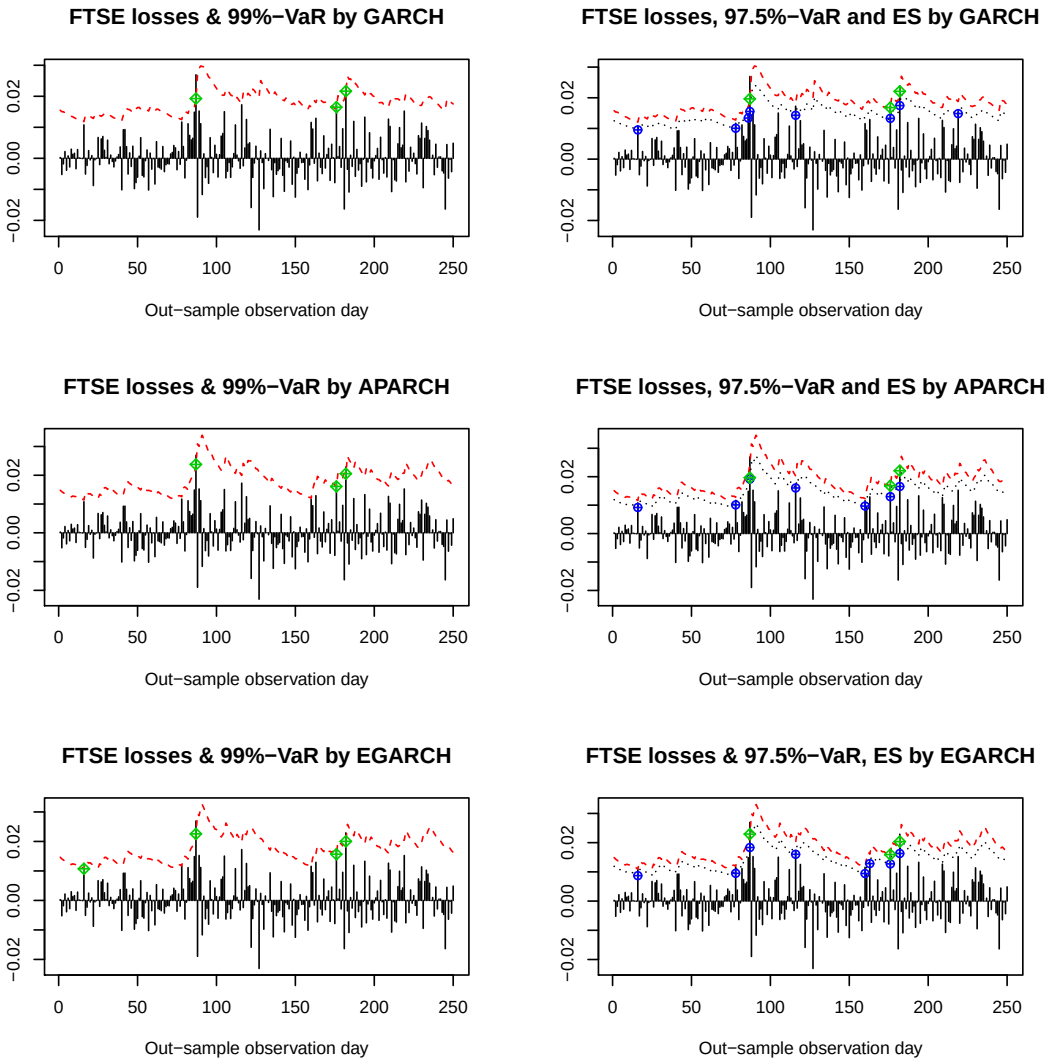

Figure A.3: FTSE POT of VaR and ES with parametric models

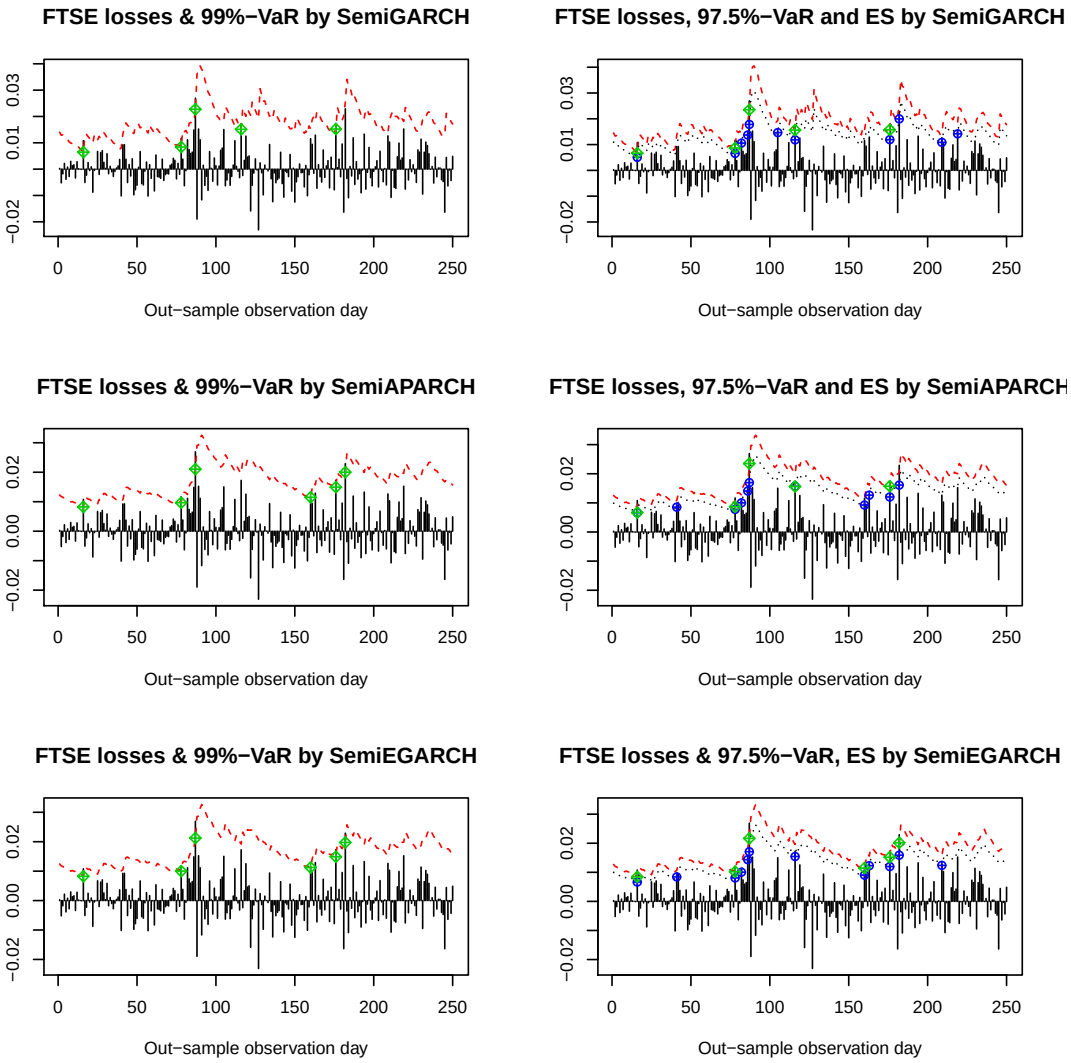

Figure A.4: FTSE POT of VaR and ES with semiparametric models

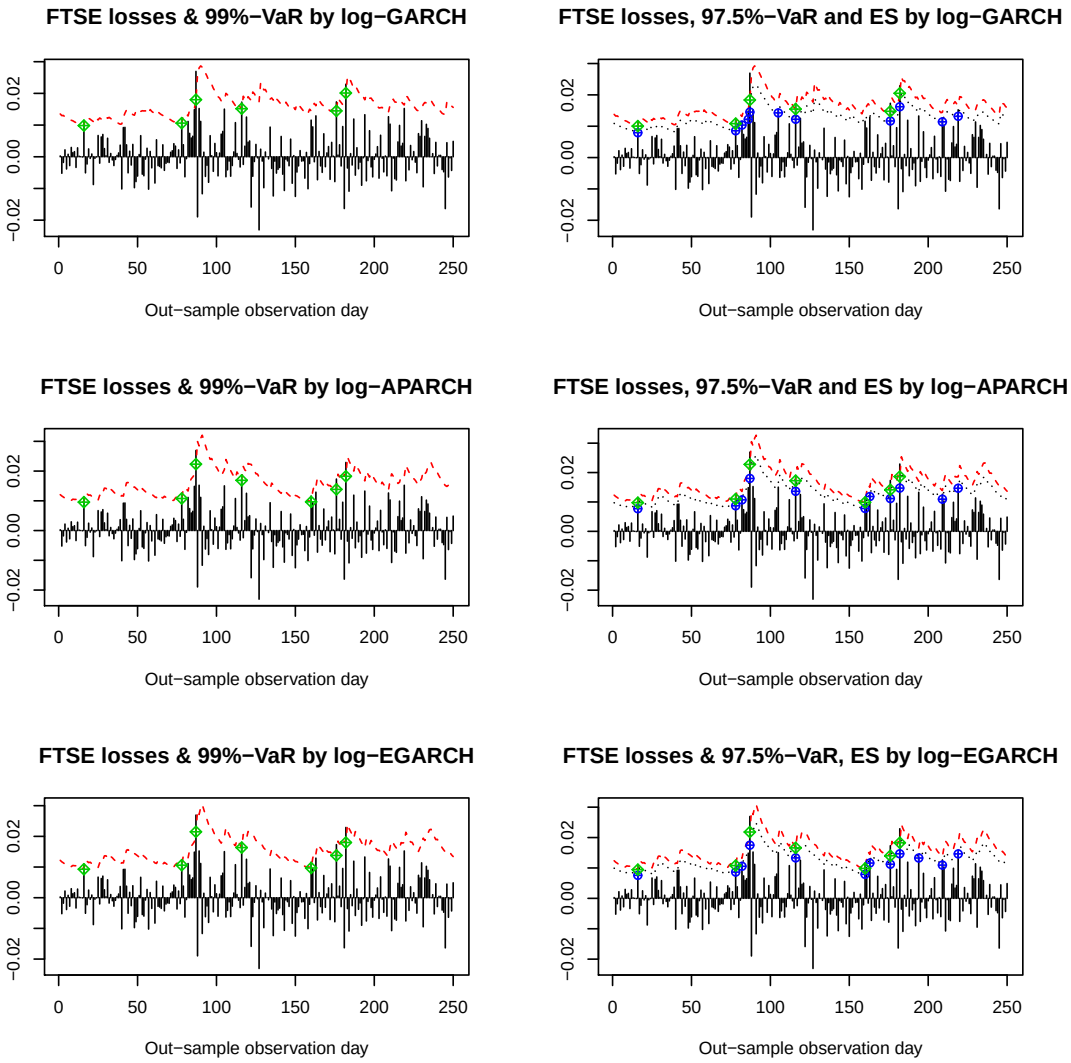

Figure A.5: FTSE POT of VaR and ES with log-models

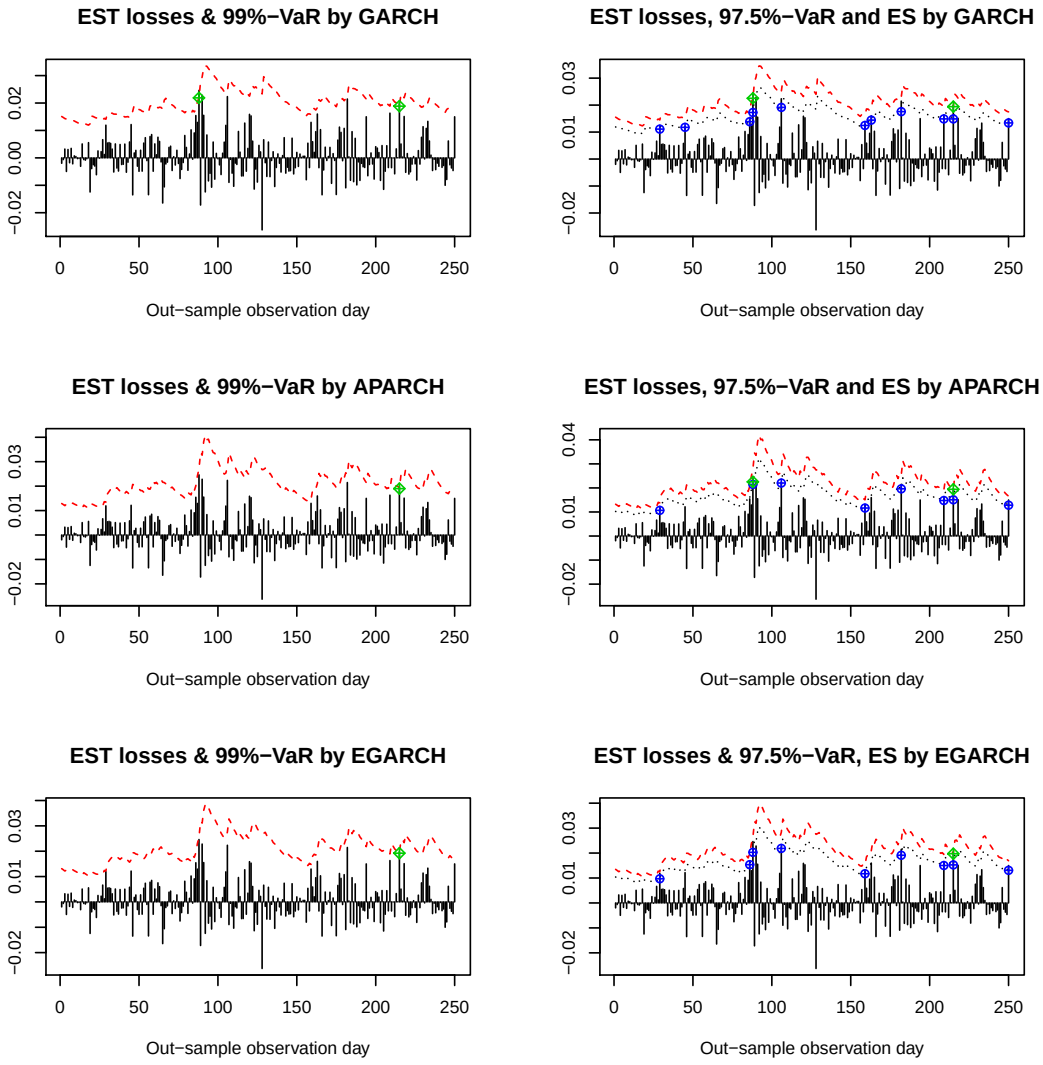

Figure A.6: EST POT of VaR and ES with parametric models

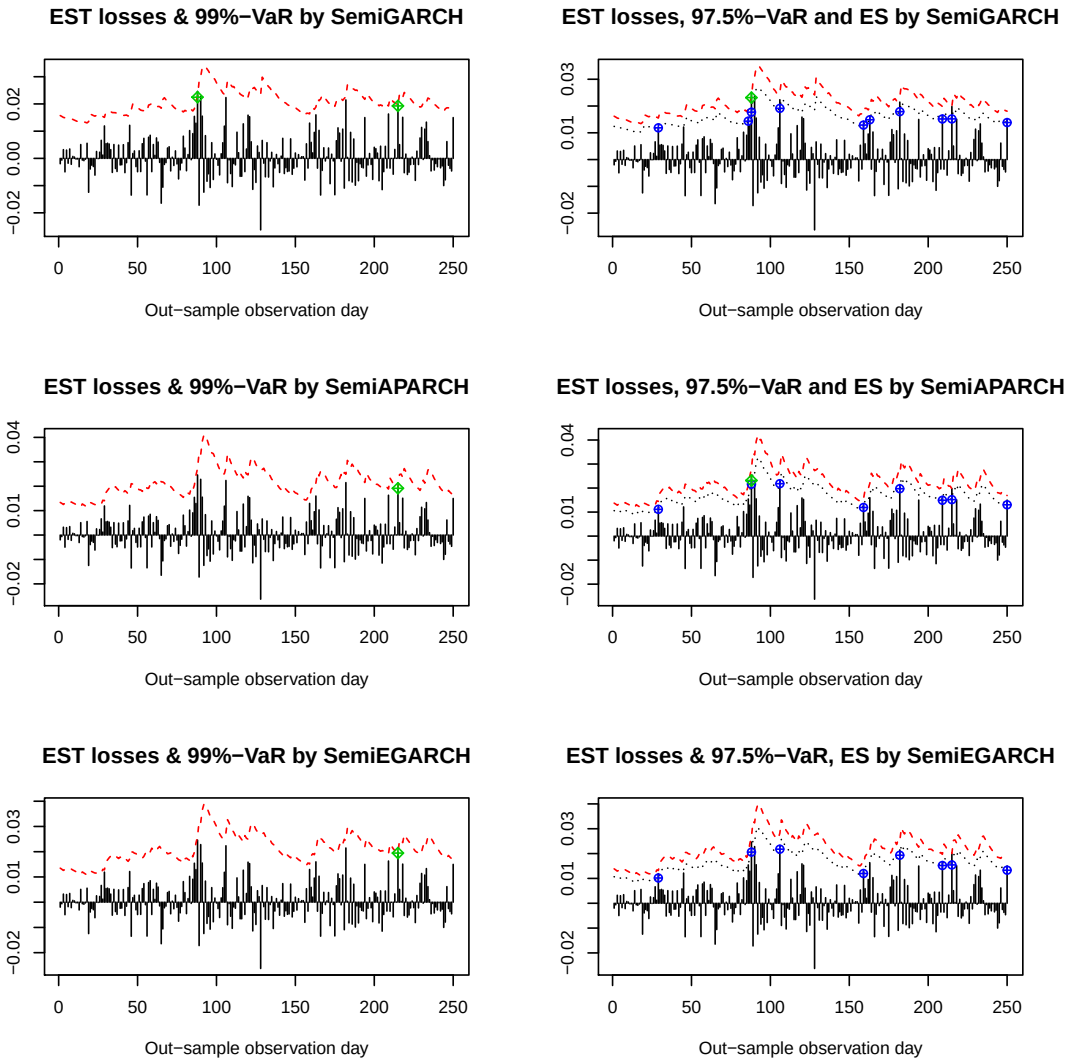

Figure A.7: EST POT of VaR and ES with semiparametric models

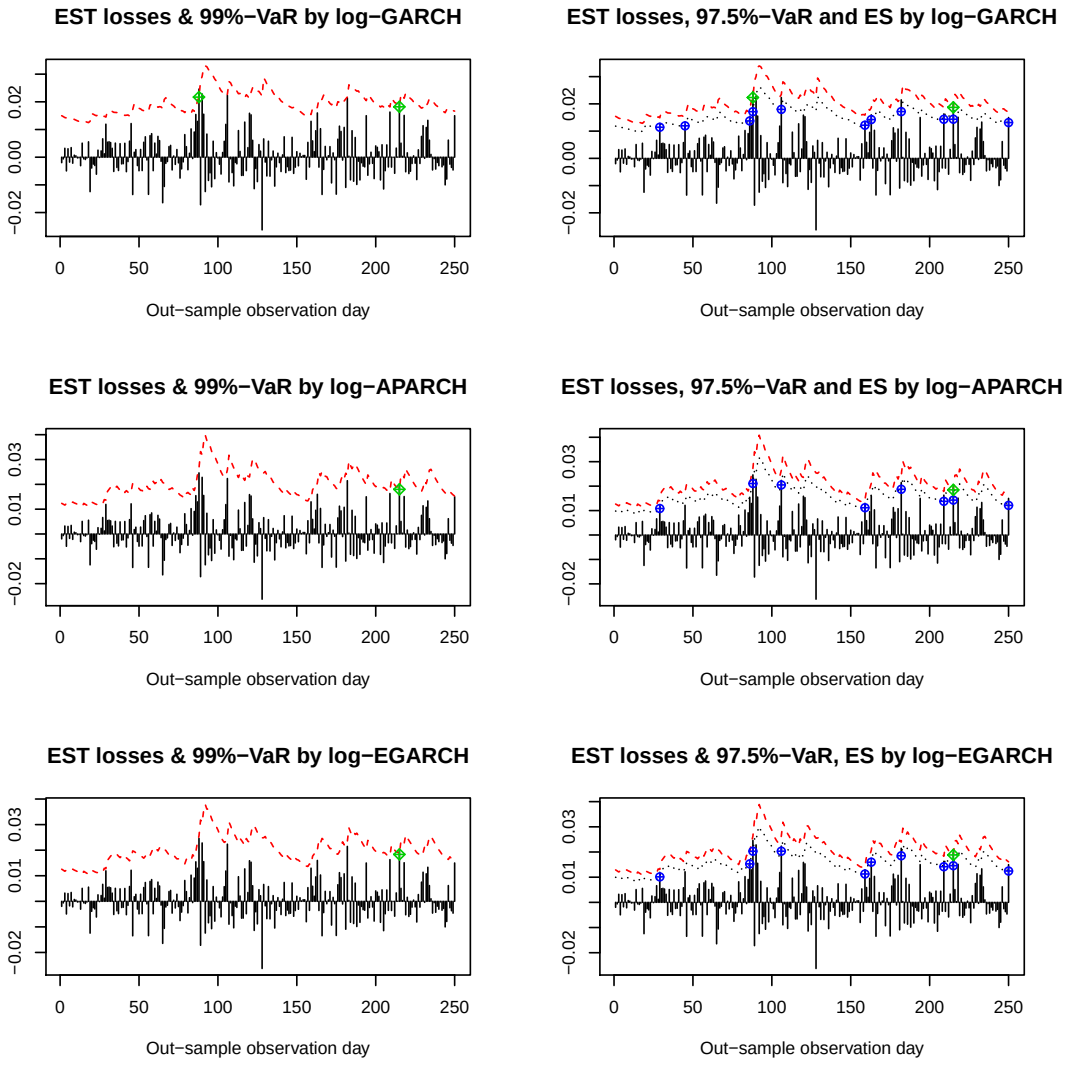

Figure A.8: EST POT of VaR and ES with log-models

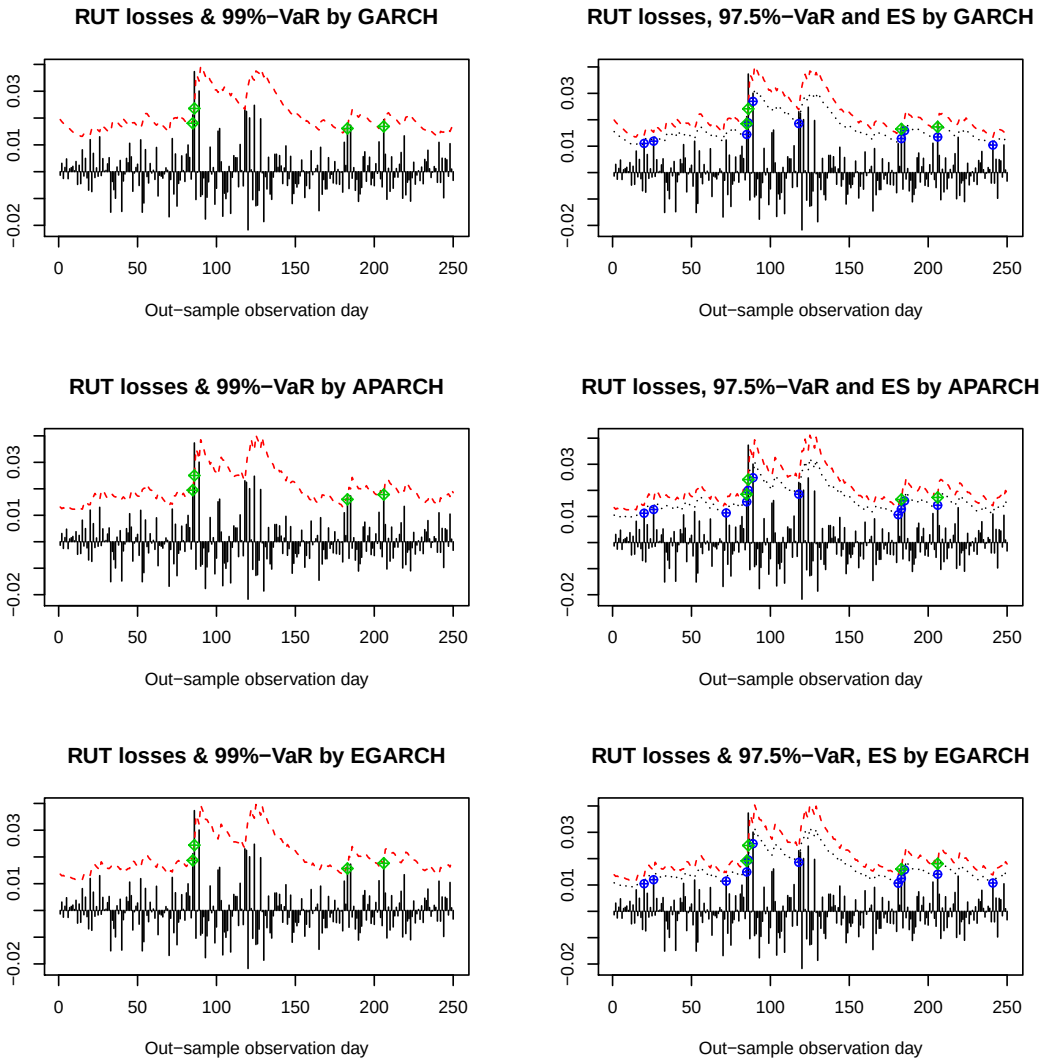

Figure A.9: RUT POT of VaR and ES with parametric models

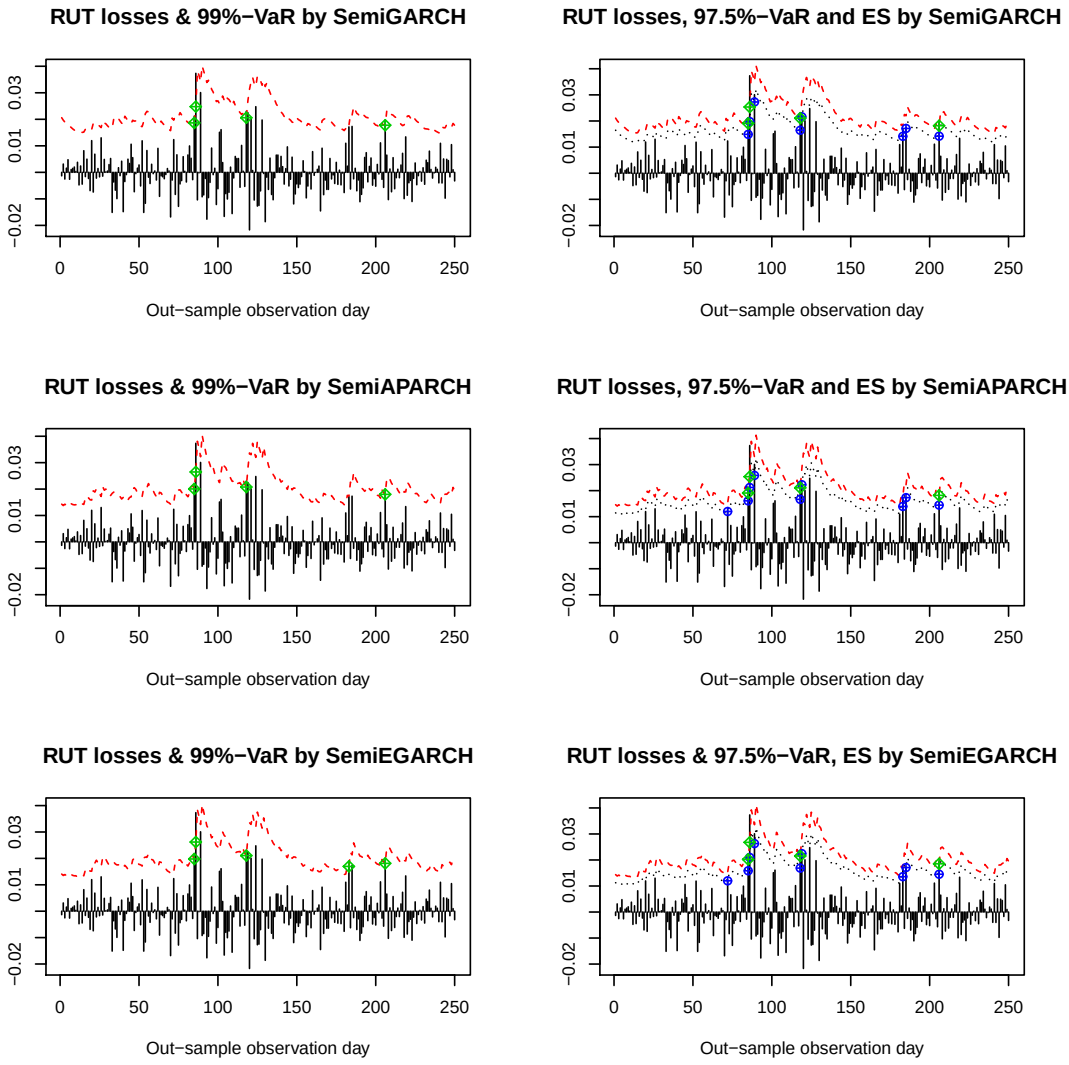

Figure A.10: RUT POT of VaR and ES with semiparametric models

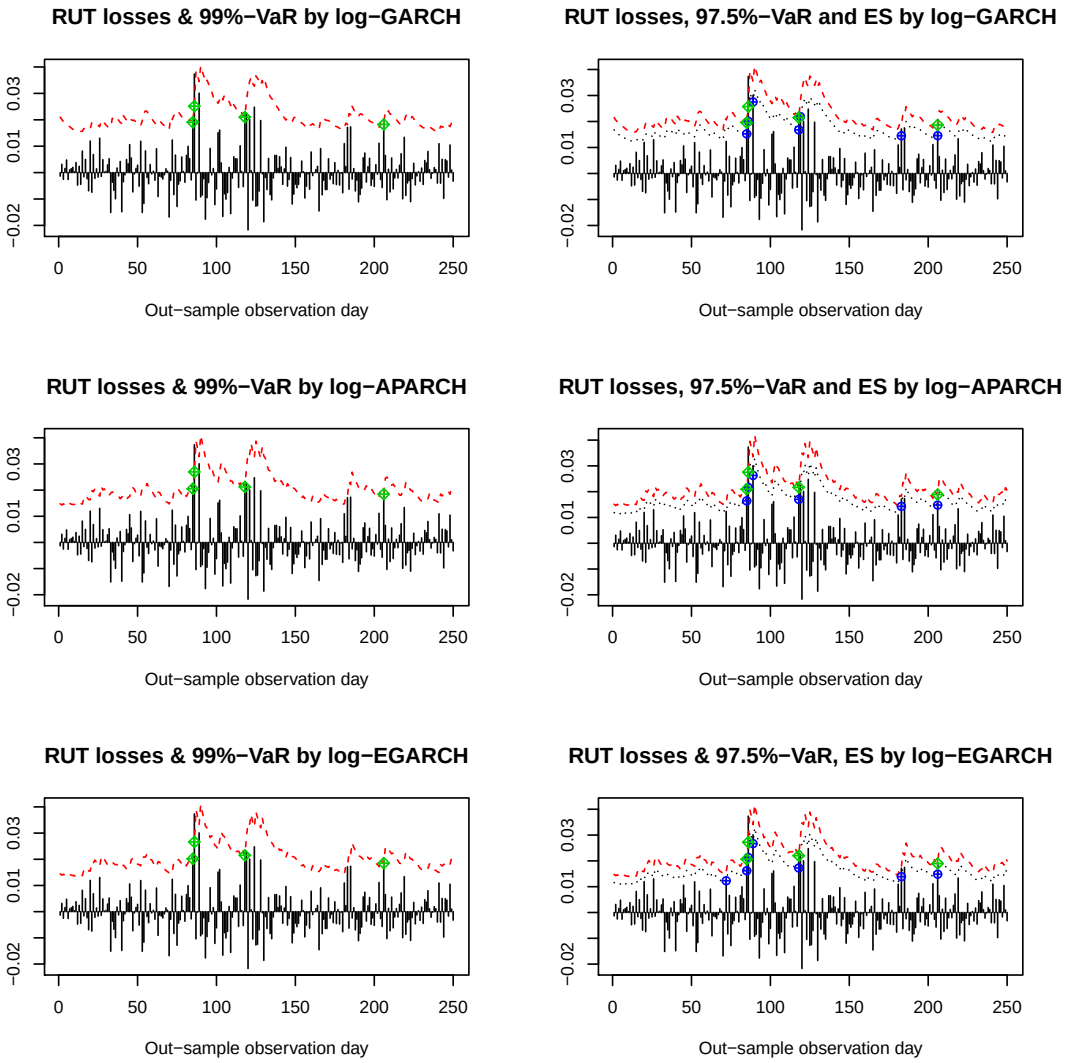

Figure A.11: RUT POT of VaR and ES with log-models

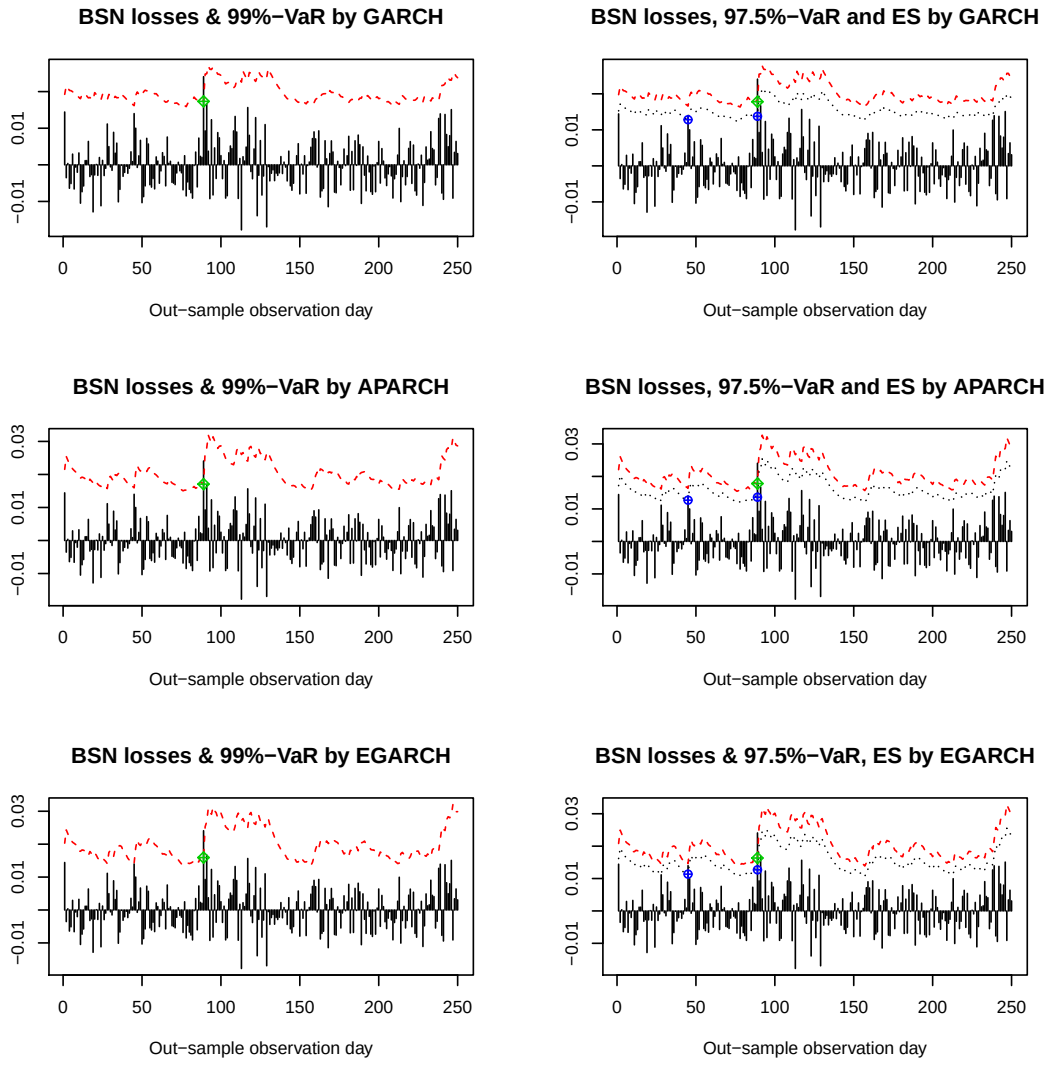

Figure A.12: BSN POT of VaR and ES with parametric models

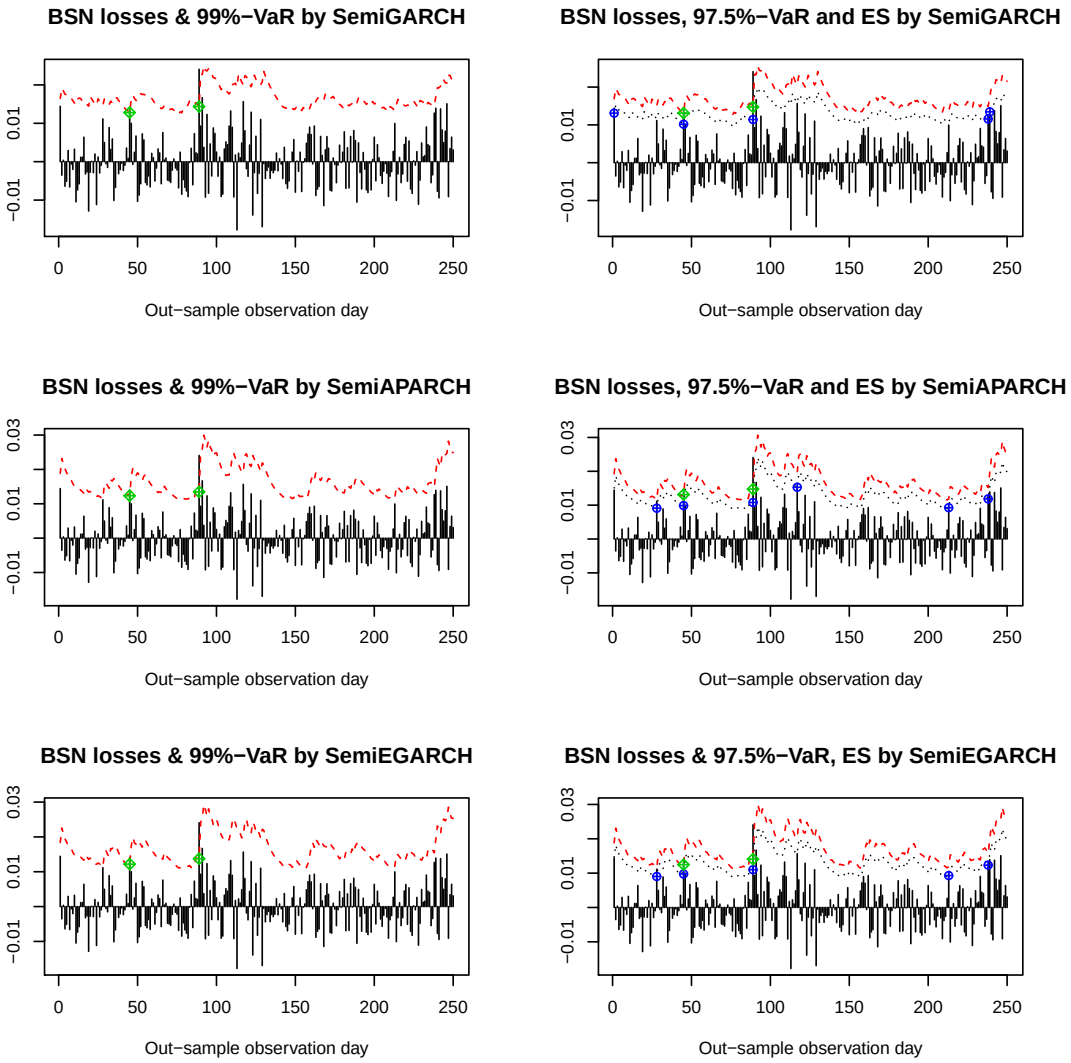

Figure A.13: BSN POT of VaR and ES with semiparametric models

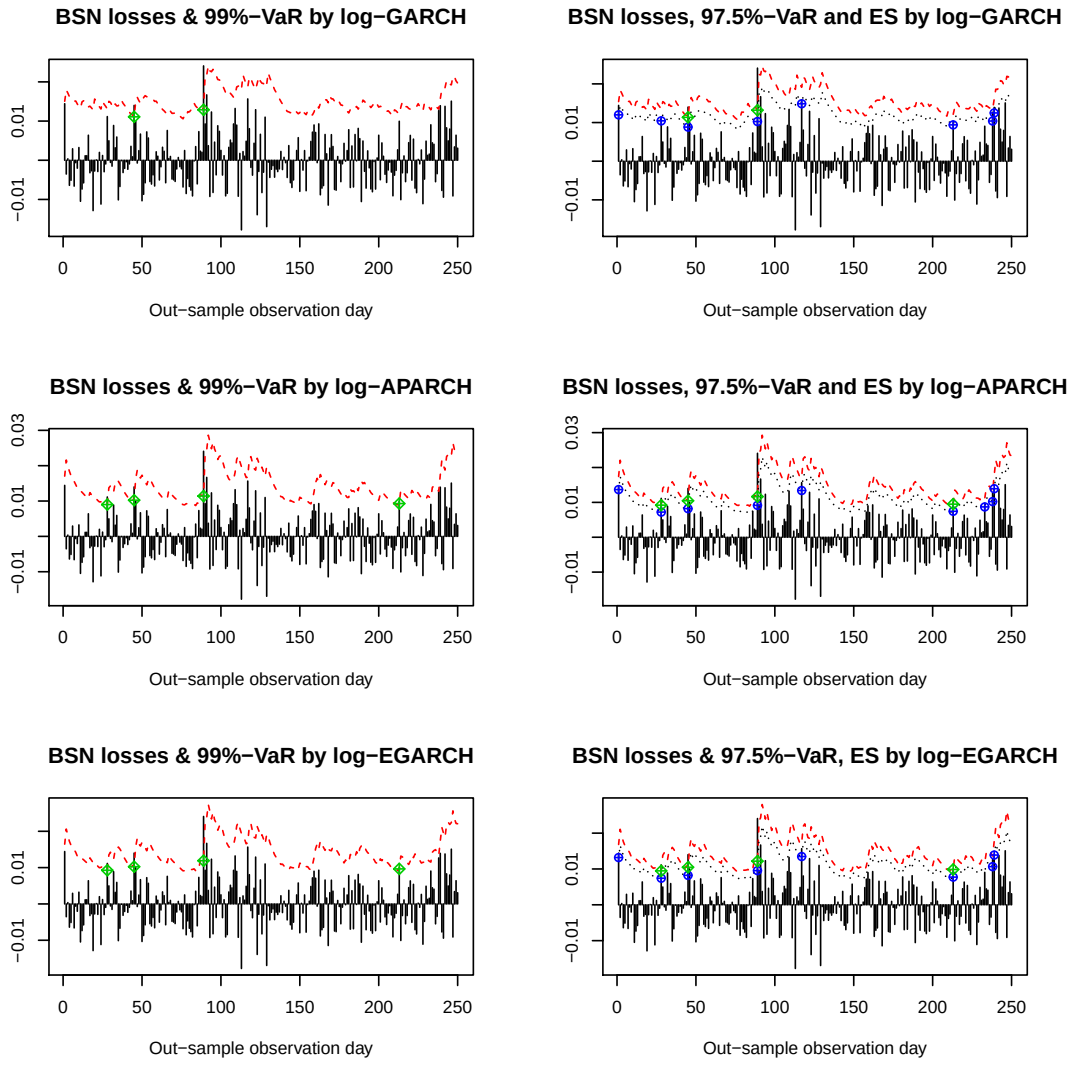

Figure A.14: BSN POT of VaR and ES with log-models

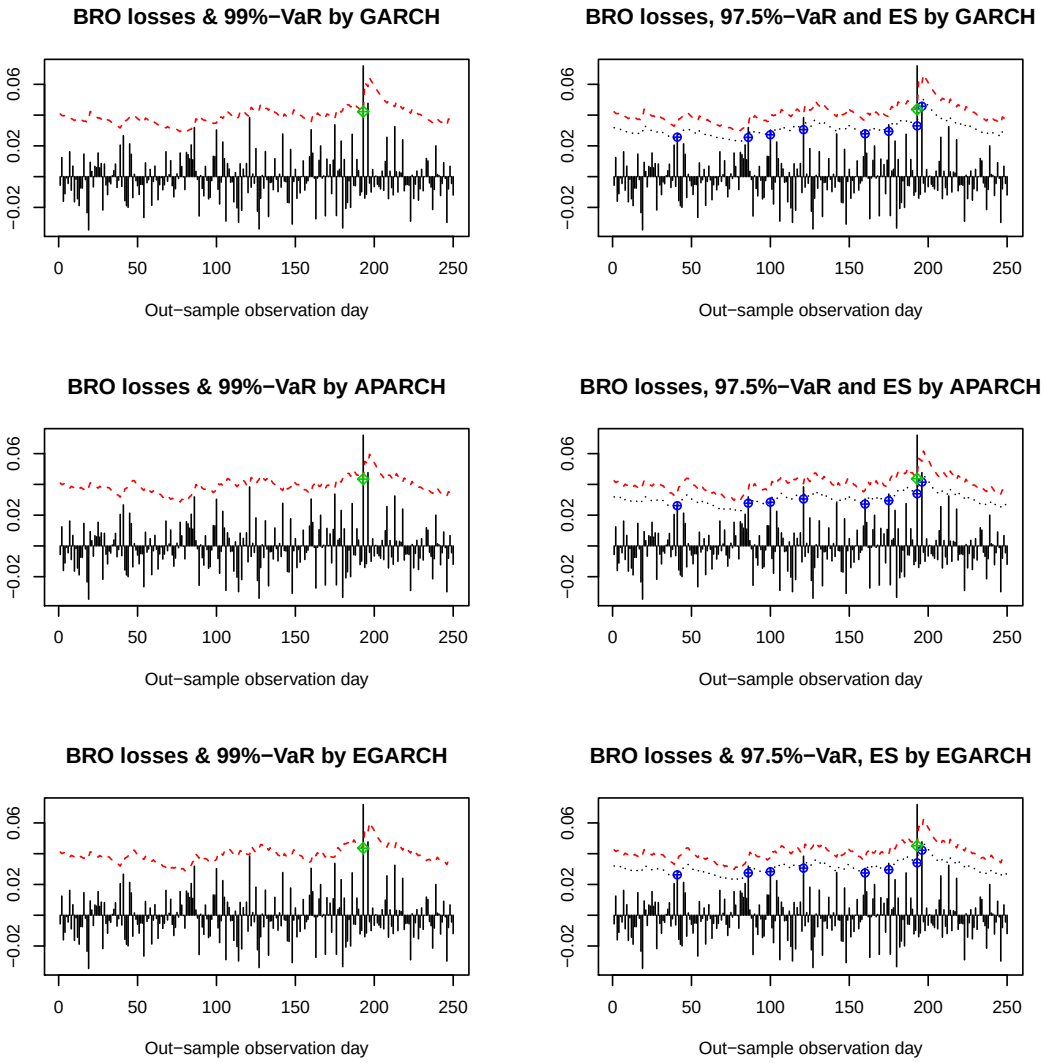

Figure A.15: BRO POT of VaR and ES with parametric models

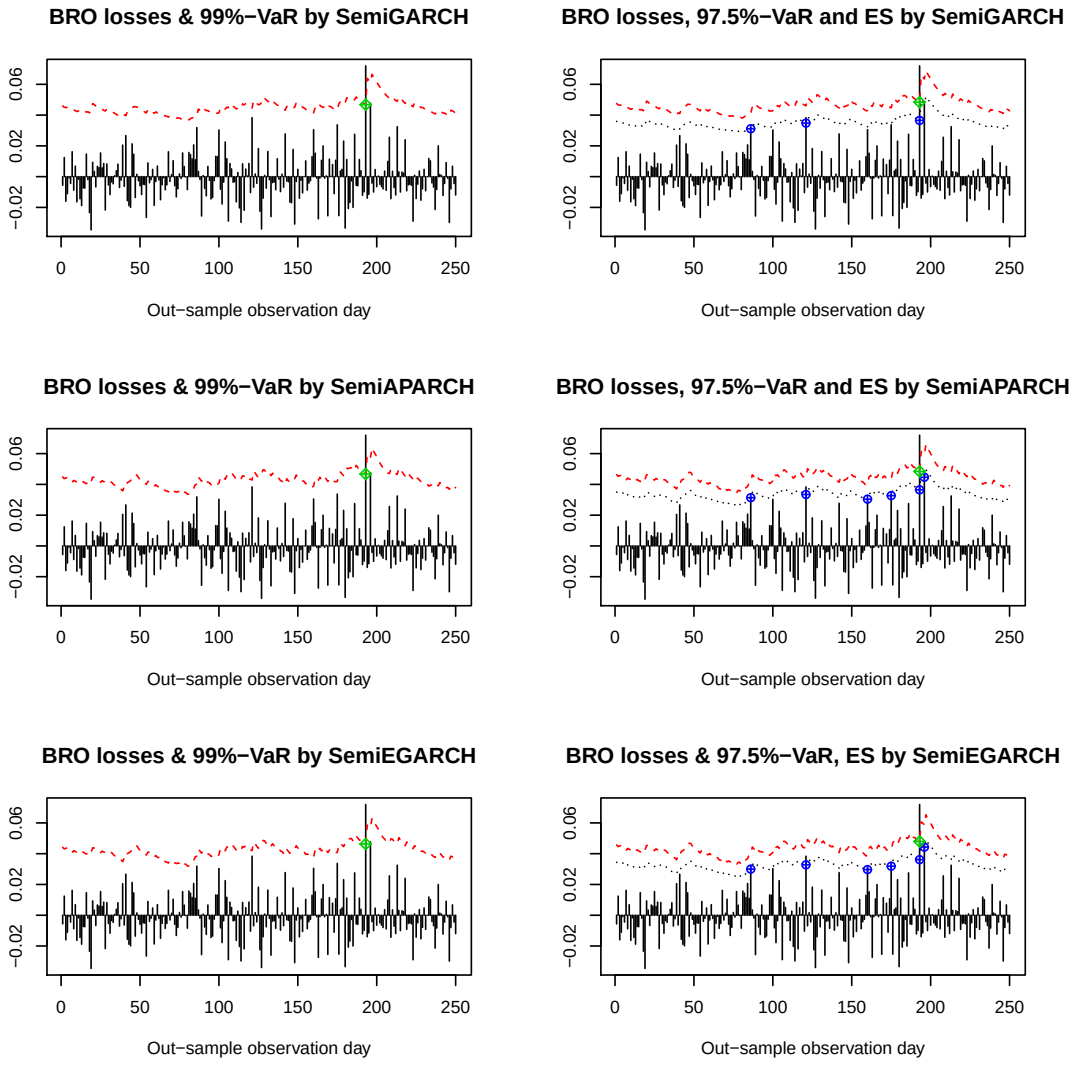

Figure A.16: BRO POT of VaR and ES with semiparametric models

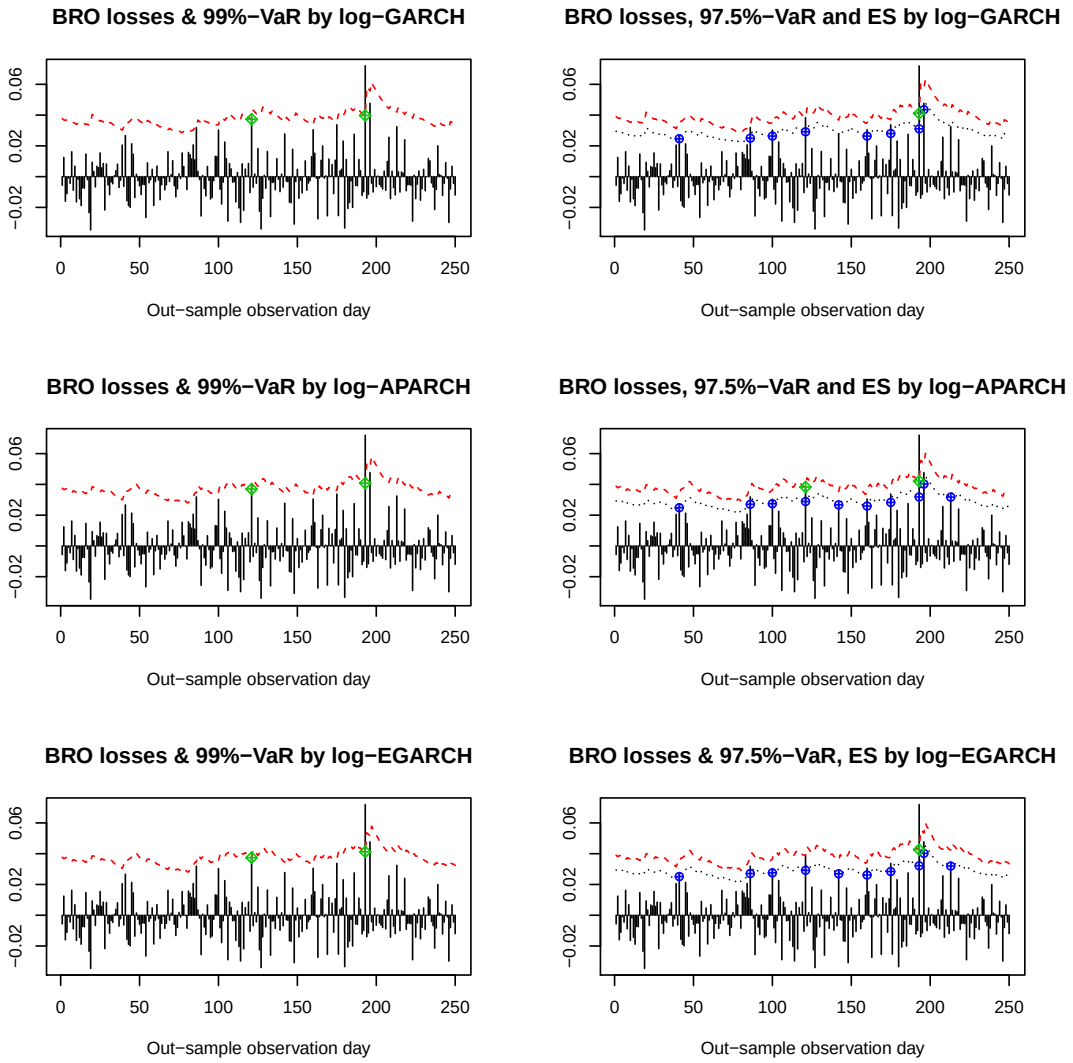

Figure A.17: BRO POT of VaR and ES with log-models

Declaration

I hereby declare that I prepared this thesis (Chapter 3 not included) on my own and have not used outside sources without declaration in the text. Any concepts or quotations applicable to these sources are clearly attributed to them. This thesis has not been submitted in the same or substantially similar version to any other authority for grading and has not been published elsewhere.

Erklärung

Ich versichere, dass ich die Arbeit (mit Ausnahme von Chapter 3) ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen worden ist. Alle Ausführungen, die wörtlich oder sinngemäß übernommen worden sind, sind als solche gekennzeichnet.

Ort, Datum	Unterschrift