Data-driven Realized Kernels and Further Analysis Using a Semi-FI-Log-ACD Model

Dissertation

zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
-Doctor rerum politicarum-
vorgelegt an der Fakultät für Wirtschaftswissenschaften
der Universität Paderborn

von

Chen Zhou M.Sc.

April 2018

Foreword

The presented thesis was written based on my a few years research as a research assistant at the Professorship of Econometrics and Quantitative Methods, Department of Economics, Paderborn University. The aim of this thesis is to estimate and predicate the volatility in the financial market. The proposed algorithms for estimating volatility are described in detailed. A comparison study among several volatility estimators are investigated. Further forecasting financial market activity using the Semi-FI-Log-ACD is developed. The robust practical performances of all proposals are illustrated by application to the high-frequency financial data of several European stocks.

I would like to express my gratitude to all those who helped me during the writing of this thesis. My deepest gratitude goes first and foremost to Prof. Dr. Yuanhua Feng, who introduced me to the subject I treat in this thesis and provided me with patience and professional instructions. As this dissertation consists of two joint works with him. I gratefully acknowledge the insightful discussions we had. I am also thankful for his academic and constructive advices to improve the quality of this thesis. Grateful acknowledgement is also made to Prof. Dr. Thomas Gries, who without hesitation, agreed to serve as my second referee. Special thanks go to Dr. John Riach for polishing the language of this dissertation. I would also to thank all my colleagues, who have given me advice and kind support.

Last but not least my special thanks to my husband Lei and my lovely son Bowen. This thesis would not be finished without the great support of them. I am also forever indebted to my parents for their great love, unselfish support, endless patience and strong encouragement through my life.

Contents

List of Figures **IV**

List of Tables **VII**

List of Abbreviations **IX**

1 Introduction **1**

1.1 Volatility models 1

1.2 Forecasting based on the Semi-FI-Log-ACD model 11

1.3 An iterative plug-in algorithm for realized kernels 14

1.4 RK under dependent noise using different sampling frequencies . . . 18

2 Forecasting financial market activity using a semiparametric fractionally integrated Log-ACD **22**

2.1 Introduction 22

2.2 A semiparametric multiplicative long memory model 24

2.2.1 Origin of long memory in aggregated financial data 24

2.2.2 Simultaneously modeling long memory and scale change . . 25

2.3 Properties and estimation of the models 27

2.3.1 The stationary solutions 27

2.3.2 Properties under the log-normal assumption 29

2.3.3 Estimation of the models 31

2.4 Forecasting based on the Semi-FI-Log-ACD 32

2.4.1 Extrapolation of the trend function 32

2.4.2 The best linear and approximately best linear predictors . . 33
2.4.3 Approximate forecasting intervals 35
2.5 Application 37
2.6 Final remarks 47
Appendix to Chapter 2 48

3 An iterative plug-in algorithm for realized kernels 53
3.1 Introduction 53
3.2 Realized volatility and realized kernels 56
3.2.1 Effect of MS noise on realized volatility 56
3.2.2 Realized kernels 58
3.3 Bandwidth selection for realized kernels 61
3.4 Application 64
3.4.1 Summary of the general findings 65
3.4.2 Detailed analysis of two challenging cases 70
3.5 Further analysis using the Semi-FI-Log-ACD 71
3.6 Final remarks 73

4 A comparison study of realized kernels using different sampling frequencies 75
4.1 Introduction 75
4.2 Realized measures 78
4.2.1 Different sampling schemes 78
4.2.2 The effect of microstructure noise 80
4.3 Bandwidth selection for realized kernels 85
4.3.1 Realized kernels 85
4.3.2 Bandwidth selection under i.i.d. noise 86
4.3.3 Bandwidth selection under dependent noise 88
4.4 Application 90
4.4.1 Implementation of algorithms 90
4.4.2 Comparison of realized estimators based on Value-at-Risk . . 98
4.5 Final remarks 100

Appendix to Chapter 4 113

5 Conclusion **121**

Bibliography **125**

List of Figures

figure 2.1 Histograms of the standardized residuals of the SEMIFAR model for the log-data and their exponential transformation for all examples. 41
figure 2.2 Estimation and forecasting results for daily trading numbers of AF from Jan. 02, 2006 to Jun. 30, 2012, obtained by the Semi-FI-Log-ACD model. 43
figure 2.3 Similar results as given in Figure 2.2 but for daily trading volumes of BMW. 44
figure 2.4 Similar results as given in Figure 2.2 but for daily average durations of Peugeot. 45
figure 2.5 Similar results as given in Figure 2.2 but for realized volatility of Metro. 46

figure 3.1 Examples of ACFs of high-frequency returns on four selected days 57
figure 3.2 Realized kernels against H obtained on the four selected days 60
figure 3.3 Histograms of selected bandwidth for all examples. 66
figure 3.4 Histograms of the number of iterations for all examples. . . . 67
figure 3.5 Logarithmic transformation of all realized volatility estimators for Air France 68
figure 3.6 Logarithmic transformation of realized kernels for the other three companies 69
figure 3.7 ACF of the high-frequency returns on the two challenging days 70

figure 3.8 Estimated trend by ESEMIFAR together with the log-data . 72

figure 4.1 ACFs of Thyssenkrupp on 20. Jan. 2012. 81
figure 4.2 ACFs of Siemens on 13. Jun. 2008. 82
figure 4.3 ACFs of Peugeot on 24. May 2011. 82
figure 4.4 ACFs of Schneider Electric on 24. Jan. 2010. 83
figure 4.5 Process of algorithms for four selected days. 91
figure 4.6 Histograms of selected bandwidth and interation number for Deutsche Bank. 92
figure 4.7 Histograms of selected bandwidth and interation number for Michelin. 93
figure 4.8 Loss and 95% VaR based on the lnRV-SEMIFAR model for Air France. 103
figure 4.9 Loss and 95% VaR based on the lnRV-SEMIFAR model for Allianz. 104
figure 4.10 Loss and 95% VaR based on the lnRV-SEMIFAR model for BMW. 105
figure 4.11 Loss and 95% VaR based on the lnRV-SEMIFAR model for Deutsche Bank. 106
figure 4.12 Loss and 95% VaR based on the lnRV-SEMIFAR model for Michelin. 107
figure 4.13 Loss and 95% VaR based on the lnRV-SEMIFAR model for Peugeot. 108
figure 4.14 Loss and 95% VaR based on the lnRV-SEMIFAR model for RWE. 109
figure 4.15 Loss and 95% VaR based on the lnRV-SEMIFAR model for Schneider Electric. 110
figure 4.16 Loss and 95% VaR based on the lnRV-SEMIFAR model for Siemens. 111
figure 4.17 Loss and 95% VaR based on the lnRV-SEMIFAR model for Thyssenkrupp. 112

figure A.4.1 Histograms of selected bandwidth and interation number for Air France. 113
figure A.4.2 Histograms of selected bandwidth and interation number for Allianz. 114
figure A.4.3 Histograms of selected bandwidth and interation number for BMW. 115
figure A.4.4 Histograms of selected bandwidth and interation number for Peugeot. 116
figure A.4.5 Histograms of selected bandwidth and interation number for RWE. 117
figure A.4.6 Histograms of selected bandwidth and interation number for Schneider Electric. 118
figure A.4.7 Histograms of selected bandwidth and interation number for Siemens. 119
figure A.4.8 Histograms of selected bandwidth and interation number for Thyssenkrupp. 120

List of Tables

table 2.1 Results of ESEMIFAR models for the four data sets 39

table 3.1 Numbers of days in different cases for the four companies . . . 65
table 3.2 Statistics of RV_0, RV_Z and RK; t between RV_0 & RV_Z, and RV_Z & RK 67
table 3.3 Results of ESEMIFAR for realized kernels of Air France . . . 73

table 4.1 Detailed results of different realized estimators ($*10^4$) for selected examples 83
table 4.2 The number of different cases for the ten companies 96
table 4.3 Mean ($*10^4$) and standard deviation ($*10^4$) of RV_0 and RK based on the different sampling frequencies for the ten companies . 97
table 4.4 "Points over 95% VaR" (benchmark: 111) by means of the SEMIFAR model based on the different logarithmic realized measures for the ten companies 101
table 4.5 Summary of "Points over 95% VaR" based on the lnRV-SEMIFAR 102

List of Abbreviations

ACD	Autoregressive conditional duration
ACF	Autocorrelation function
AF	Air France
ALV	Allianz SE (german)
AMSE	Asymptotic mean squared error
AR	Autoregressive
ARCH	Autoregressive conditional heteroskedasticity
ARMA	Autoregressive moving average
BIC	Bayesian information criterion
BMW	Bayerische Motoren Werke AG (german)
BTS	Business time sampling
CCC	Constant conditional correlation
CGARCH	Component GARCH
CTS	Calendar time sampling
DBK	Deutsche Bank (german)
DCC	Dynamic conditional correlation

DN	Dependent noise
ECCC	Extended constant conditional correlation
EFARIMA	Exponential fractional autoregressive integrated moving average
FARIMA	Fractional autoregressive integrated moving average
FIACD	Fractionally integrated ACD
FIGARCH	Fractionally integrated GARCH
FI-Log-ACD	Fractionally integrated log transformed ACD
GARCH	Generalized autoregressive conditional heteroskedasticity
i.i.d	independent and identically distributed
IQ	Integrated quarticity
IN	Independent noise
IV	Integrated volatility
LMGARCH	Long memory GARCH
Log-ACD	Long memory ACD
Log-FARIMA	Log transformed FARIMA
Log-SEMIFAR	Log transformed SEMIFAR
MA	Moving average
MD	Mean duration
MEOG	Metro
MIC	Michelin

MISE	Mean integrated squared error
MN	Microstructure noise
MSE	Mean squared error
PoV	Points over Value-at-Risk
PSA	Peugeot
QRV	Quantile based realized variance
RK	Realized kernels
RV	Realized volatility
RV$_Z$	Realized volatility calculated by Zhou (1996)'s method
RWE	Rhenish-Westphalian Power Plant
SE	Schneider Electric
SEMIFAR	Semiparametric fractional autoregressive
Semi-ACD	Semiparametric ACD
Semi-FI-Log-ACD	Log-transformed semiparametric fractionally integrated ACD
Semi-GARCH	Semiparametric GARCH
Semi-Log-ACD	Semiparametric log transformed ACD
SIE	Siemens
SV	Stochastic volatility
THK	Thyssenkrupp
tick	tick-by-tick
TkTS	Tick time sampling

TrN	Trading number
TrTS	Transaction time sampling
TrV	Trading volume
VaR	Value-at-Risk
VEC	Vector error correction
WACD	Weibull distributed ACD

Chapter 1

Introduction

1.1 Volatility models

Estimating, modelling and forecasting volatility has been the subject of extensive research among academics and practitioners over the last twenty years. Financial market volatility is a statistical measure of the price fluctuation over time for a given security and can be measured by using the standard deviation or variance of returns. It is indispensable for the theory and practice of portfolio selection, asset pricing and risk management. The higher the volatility, the riskier the security. The traditional financial econometrics models of risk are generally regarded as variance independent, identically distributed constants. Since the 1960's, a large number of empirical research have confirmed that the variance varies over time. Meanwhile, researcher and academicians have found four main characteristics of financial asset volatility, which are important for analysis the volatility in the financial market.

1. Volatility clustering: Time series of financial asset returns often exhibit the volatility clustering property, which means large changes tend to be followed by large changes and small changes tend to be followed by small changes. A quantitative manifestation of this fact is that returns themselves are uncorrelated, while absolute returns and their squares display a positive, significant and slowly decaying autocorrelation function. To describe this phenomenon, in 1982 Robert

Engle proposed the ARCH model, which is the first framework for modeling conditional volatility. It is assumed in the ARCH model that the returns depend on the past information with a specific form. Four years after the ARCH model was proposed, Bollerslev (1986) addressed a generalized ARCH model called the GARCH model, which is one of the most well-known extension models based on the ARCH. The main idea behind these two models is the same. However, the GARCH model overcomes some disadvantages of the ARCH model. In particular, in empirical applications, the ARCH model commonly requires a relative long lag in the conditional variance equation which leads to a higher order and makes the model more complicated. Besides, the GARCH can better capture the volatility process of an asset return by adding the conditional heteroskedasticity moving items. In addition, all of the ARCH process can be extended to the GARCH process, i.e. the ARCH process is a special case of the GARCH process, which is why the study of the ARCH was replaced by the GARCH model, since it was proposed. A GARCH (p,q) model is defined by

$$Y_t \mid F_{t-1} = \eta_t \sqrt{h_t}, \qquad Y_t | F_{t-1} \sim N(0, h_t)$$

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

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, F_{t-1} denotes past information. η_t are the i.i.d random variables and can be expressed as a standard normal distribution with zero mean and unit variance. The GARCH model has greater applicability for easy computation. Using maximum log-likelihood method can estimate the parameters in the GARCH model.

2. Leverage effects: However, standard GARCH models assume that positive and negative error terms have a symmetric effect on the volatility. From an empirical point of view the volatility increases more after bad news than after good news. In order to measure the rate of return volatility asymmetry, Ding et al. (1993) proposed an asymmetric power ARCH also called APARCH model, in which the volatility reacts asymmetrically to negative and positive returns. A

time series Y_t follows an APARCH (p,q) model is defined by,

$$Y_t = \eta_t \sigma_t, \quad \sigma_t^\delta = \alpha_0 + \sum_{i=1}^{p} \alpha_i (\mid Y_{t-i} \mid -\gamma_i Y_{t-i})\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta,$$

where $\sigma_t = \sqrt{h_t}$ is the conditional variance, $0 < \delta \leq 2$ is a power index for this model and $0 \leq \gamma_i < 1$ is the "leverage effect". By changing the parameters the APARCH model nests at least seven ARCH-type models.

- The ARCH model with $\delta = 2$, $\gamma_i = 0$ and $\beta_j = 0$.
- The GARCH model with $\delta = 2$, $\gamma_i = 0$.
- The TS-GARCH (Taylor, 1986, Schwert, 1990) with $\delta = 1$, $\gamma_i = 0$.
- The GJR (Glosten et al., 1993) with $\delta = 2$.
- The TARCH of Zakoian (1994) with $\delta = 1$.
- The NARCH (Higgins and Bera, 1992) with $\gamma_i = 0$ and $\beta_j = 0$.
- The Log-ARCH (Geweke, 1986, Pantula 1986) with $\delta \to 0$.

3. Fat tail: Financial asset returns often possess distributions with a fat tail, which has the property that there is a (relatively) high probability of some "unusual" events. Using the normal distribution assumption can underestimate the probability of unusual events and thus affect the accuracy of risk management estimates-VaR (Value at risk). Therefore, such as the GARCH model with conditional t-distribution after standardization (Bollerslev, 1987) and the GARCH model with skewed innovation were proposed.

4. Long memory: Long memory is considered as one of the most important statistical properties of time series. It implies there are correlations between two long distanced observations and its ACF decays slowly at hyperbolic rate. AR, MA, ARMA, ARIMA models represent only short memory features. Hence, these models are inadequate. The fractionally integrated GARCH model (FIGARCH)

was proposed by Baillie et al. (1996) and can be used to model long memory phenomenon in the volatility. However, the unconditional variance in the FIGARCH model does not exist, which is why it can not be used to analyze long memory in a usual sense. This problem was solved by the proposal of the LMGARCH model (Karanasos et al., 2004).

In the literature, a large number of extensions of the GARCH model build up a GARCH family, such as the general exponential GARCH (EGARCH, Nelson, 1991) model, where the conditional variance is an explicit multiplicative function of lagged innovations; the component-GARCH model (CGARCH, Engle and Lee, 1999). Mikosch and Stăaricăă (2004) showed that in a fitted GARCH (1,1) model a non-stationarity phenomenon exhibited ($\alpha_1+\beta_1 \approx 1$), a piecewise GARCH model was hence proposed. The Semi-GARCH model with a slowly changing scale function for modelling conditional heteroscedasticity and time heteroscedasticity simultaneously was proposed by Feng (2004). If an existing nonparametric scale function is not considered, it will be misinterpreted as very strong long memory. The multivariate GARCH class of models was first introduced and estimated empirically by Bollerslev et al. (1988). Their model is generalized directly from the univariate GARCH model to multivariate case and called the Vector Error Correction (VEC) model. Bollerslev (1990) investigated the Constant Conditional Correlation (CCC) model, in which all conditional correlations are constant and the conditional variances can be modeled as univariate GARCH processes. Engle (2002) extended the CCC model to a Dynamic Conditional Correlation (DCC) model, where the correlations change with time.

In the last decade, owing to the rapid development in computer technology, methods of data processing and collection have made swift progress. These make the observations with small time intervals–high-frequency data obtainable. High-frequency data records the real-time transaction and provides many more details of the events in the financial market. Hence, it is of great significance for the understanding and research of financial market microstructure. Based on the idea of the ARCH and GARCH models, a new model is developed to investigate the irregularly time-spaced characteristic of high frequency data, namely the autoregressive

conditional duration model (ACD, Engle and Russell, 1998). Let $X_t = t_i - t_{i-1}$ be the i^{th} duration between two events which occur at times t_{i-1} and t_i. The sequence $\{x_1, x_2, \ldots, x_n\}$ has non-negative elements, since $t_1 \leq t_2 \leq t_3 \leq \ldots t_n$. Let $\psi_t \equiv E(X_t|\digamma_{t-1})$ be the conditional mean (expected) durations, where $\digamma_{t-1}$ denotes past information set of durations available at time $t-1$. A general model for conditional duration is:

$$X_t = \psi_t \varepsilon_t,$$

where $\varepsilon_t > 0$ are i.i.d. with $E(\varepsilon_t) = 1$ and $var(\varepsilon_t) = \sigma^2$. Following the GARCH idea the ACD(p,q) model is defined by:

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

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$. It means that the conditional mean duration ψ_t depends on p previous durations and q previous mean durations. The restrictions of the ACD and (G)ARCH models (i.e, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$) are to ensure ψ_t and h_t to be positive. As a matter of fact, the ACD model is an application of the GARCH model to the duration data. Define $\eta_t = X_t - \psi_t$, which are uncorrelated with mean zero, then like the GARCH model, the ACD(p,q) model can also be treated as an ARMA (max(p,q),q) model:

$$X_t = \alpha_0 + \sum_{i=1}^{p} X_{t-i} + \sum_{j=1}^{q} \beta_j X_{t-j} - \sum_{j=1}^{q} \beta_j \eta_{t-j} + \eta_t.$$

The strong similarity between the ACD and (G)ARCH models has nurtured the rapid expansion of alternative specifications of conditional durations. One of the extensions of the ACD model is to change the distribution of ε_t. Engle and Russell (1998) proposed WACD model with standardized weibull distribution. Gramming and Maurer (2000) used a burr distribution which contains the exponential, weibull and log-logistic as special cases.

In the ACD (p,q) model, it is assumed that all coefficients (α_0, α_j, β_j) are positive in order to ensure the positivity of durations. Due to such a restriction

it is not allowed to add some variables taken from the microstructure literature with possible negative coefficients in the autoregressive equation, which may lead to a negative duration. To avoid this, some nonlinear ACD models were developed. The Log-ACD model was proposed by Bauwens and Giot (2000). Following the idea of the linear ACD model, the general form of the Log-ACD model is introduced by

$$X_t = e^{\phi_t}\varepsilon_t, \qquad \ln\psi_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i g(\varepsilon_{t-i}) + \sum_{j=1}^{q}\beta_j \ln\psi_{t-j},$$

where ϕ is proportional to the logarithm of the conditional expection of X_t, ε_t are i.i.d and follow a Weibull $(1,\gamma)$ distribution with $E(\varepsilon_t) = 1$ and $var(\varepsilon_t) = \sigma^2$. Let $g(\varepsilon_{t-i})$ be $\ln\varepsilon_{t-i}$ or ε_{t-i}, which correspond to the Log-ACD$_1$ model or Log-ACD$_2$ model, respectively. In this thesis we will only discuss the Log-ACD$_1$ model, which actually can be considered as the ACD model using logarithmic data. The Log-ACD$_1$ model is defined by

$$\ln\psi_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i \ln X_{t-i} + \sum_{j=1}^{q}\beta_j \ln\psi_{t-j}.$$

This model is close to the Log-GARCH model proposed by Geweke (1986). Compared with the original ACD model it is more flexible and the only constraint on the coefficients is $\beta < 1$. However, one drawback of Log-ACD is that it cannot accommodate durations, which are equal to zero. Allen et al. (2008) showed that the Log-ACD model can be rewritten as a linear ARMA process using the logarithmic data (namely the Log-ARMA model). Define $\mu_t = \ln X_t - \ln\psi_t$, where $E(\mu_t) = 0$. The Log-ACD model can be rewritten as a Log-ARMA model as follows:

$$\begin{aligned}\ln X_t &= \alpha_0 + \sum_{i=1}^{r}\alpha_i \ln X_{t-i} + \sum_{j=1}^{r}\beta_j \ln\psi_t + \ln X_t - \ln\psi_t \\ &= \alpha_0 + \sum_{i=1}^{r}(\alpha_i+\beta_i)\ln X_{t-i} - \sum_{j=1}^{r}\beta_j(\ln X_t - \ln\psi_{t-j}) + \mu_t.\end{aligned}$$

If $\sigma_i = \alpha_i + \beta_i$ and $\theta = -\beta_j$, then

$$\ln X_t = \alpha_0 + \sum_{i=1}^{r} \sigma_i \ln X_{t-i} + \sum_{j=1}^{r} \theta_j \mu_{t-j} + \mu_t.$$

Reversely, it is easy to show that the exponential transformation of a linear ARMA process can be written as a Log-ACD model. Bayesian information critterion (BIC) can be used to select a suitable model and the parameter of the Log-ACD model will also be obtained through the Log-ARMA model. Comparing with the multiplicative form of the ACD model, the Log-ACD model is a additional model, which is easier to estimate. A long memory extension of the Log-ACD model can be thought of as the exponential fractional autoregressive integrated moving average model (EFARIMA), which was proposed by Beran et al. (2015), and can simply be thought of as an application of the well known FARIMA model to the log transformed data.

Following the idea of the Semi-GARCH, a Semi-ACD model with a slowly varying scale function was proposed. The estimation of the Semi-ACD model is better than that of the ACD model, if local stationary phenomenon exists. The Semi-ACD model can be defined as a multiplicative model and it can be considered as an ACD(p,q) model with a smooth scale-function $\nu(\tau_t)$. Let X_t be the duration between two events, $\tau_t = t/n$ be the rescaled time, ψ_t be the conditional mean duration after standardization and $\eta_t \geq 0$ are i.i.d. random variables with unit mean, then the Semi-ACD model is proposed by

$$X_t = \nu(\tau_t)\psi_t\varepsilon_t,$$

where $\psi_t\varepsilon_t = \zeta_t$ is the conditional dynamics of the stationary process. After removing the scale function $\nu(\tau_t)$, the ACD (p, q) model can be employed to analyze ζ_t. The estimation of the scale function ν is a nodus in the estimation of semiparametric models. A data-driven algorithm with a good asymptotic properties for estimating the nonparametric scale function was proposed by Beran and Feng (2002b, 2002c). This approach does not have any assumptions of parametric

model, except that ζ_t is assumed to be stationary. After removing the nonparametric part- scale function, the approximations of ζ_t can be derived by means of fitting some suitable parametric models. The selection of suitable model can follow some well known model selection criteria like BIC. A long memory extension of the Semi-Log-ACD model was proposed for modelling a nonnegative process with long memory and a nonparametric scale function simultaneously. It can be thought of as the semiparametric version of the EFARIMA model (SEMIFAR). Beran and Ocker (2001) found significant trend in volatility by fitted SEMIFAR models.

Taylor (1982, 1986) proposed stochastic volatility model (SV). Both GARCH family and SV models take into account the important volatility clustering phenomenon and are driven by the past information. The main difference is that the GARCH family models assume deterministic volatility states, while in the SV model the volatility is a latent variable with unexpected noise. The introduction of the additional error term makes the SV model more flexible than the GARCH family models. Carnero et al. (2004) showed that compared with the GARCH models, the SV model captures the main empirical properties often observed in daily series of financial returns in a more appropriate way. However, the empirical applications of the SV model have been very limited. The main reason is that the GARCH model is easily estimated via Maximum Likelihood estimator, while the likelihood of SV models is not directly available. The use of simulation techniques, like simulated maximum likelihood, the generalized method of moments and Markov chain Monte Carlo are required. Ghysels et al. (1996) and Shephard (1996) surveyed the SV literature, and Broto and Ruiz (2004) reviewed the estimation methods for the SV models. The first multivariate SV model proposed by Harvey et al. (1994). Asai et al. (2006) reviewed the substantial literature on specification, estimation and evaluation of multivariate SV models. An alternative approach is implied volatility, which invoke option pricing models to invert observed derivatives prices over a fixed future period. Such procedures remain model-dependent and further incorporate a potentially time-varying volatility risk premium in the measure so they generally do not provide unbiased forecasts of

the volatility of the underlying asset.

Increased availability of high-frequency data in the last decade have resulted in development of the new model-free volatility approach called realized volatility (Andersen et al., (2001a, 2001b)), which exploits the information in high frequency returns and is constructed as the sum of the squared intraday returns that are sampled at very short intervals. In contrast to all GARCH and SV models, realized volatility estimator does not need parametric assumptions. Koopman et al. (2005), Martens (2002) and Martens et al. (2007) among others demonstrated the superiority of realized volatility models over GARCH family for volatility forecasting. In the ideal case, increasing the sampling frequency towards to make more accurate estimates of volatility on any given day. This implies that daily volatility becomes almost observable via realized volatility. However, a perfect estimate of realized volatility can be obtained only under the assumption that prices are observed in continuous time and without measurement. In reality the sampling frequency is limited by actual quotation or transaction frequency and observed prices are contaminated by market microstructure (MS) noise, which leads to the bias problem, refer to e.g. Andersen et al., (2001a); Barndorff-Nielsen and Shephard (2002). Market MS noise could, for example, be induced by the discreteness of price changes, bid-ask bounce, latency, and asymmetric information of traders. Due to the market MS noise, the intraday returns are correlated (normally negative). If this negative correlation is not considered, the realized volatility will be overestimated. To solve this bias problem, different approaches have been introduced into the literature. Bandi and Russel (2006, 2008, 2011) and Oomen (2006) investigated a method of selecting the optimal sampling frequency based on a trade-off between the variance and bias. Hansen et al. (2008) investigated correction of MS bias using moving average-based estimators. Podolskij et al. (2009) provided the pre-averaging approach. The subsampling approach originally suggested by Zhang et al. (2005), which builds on the use of a realized volatility estimator with two time scales under dependent MS noise. A generalization of Zhang et al. (2005) was introduced by Aït-Sahalia et al. (2011) and Zhang (2006), which is consistent and asymptotically unbiased under dependent

noise. Zhou (1996) proposed to use the first order correlation to correct the bias. This is the first kernel based method and was generalized in Hansen and Lunde (2006a). Nevertheless, the estimators are inconsistent. Barndorff-Nielsen et al. (2008) proposed the realized kernels (RK), which are a generalization of Zhou (1996) and are consistent estimators of the IV under given conditions. Barndorff-Nielsen et al. (2009) provided the practical application of the non-flat-top realized kernels. Barndorff- Nielsen et al. (2011a) considered further refinements of the realized kernels in the spirit of the subsampling approach adopted in the TSRV estimators by using averaged covariance terms in the realized kernel estimators. Ikeda (2015) proposed two-scale RK, which is a convex combination of two realized kernels with different bandwidths and converges to the daily integrated volatility in the presence of dependent MS noise. Moreover, some alternative realized variance measures based on the quadratic variation that exploit other aspects of high frequency financial data were investigated. Christensen and Podolskij (2007) proposed a realized range-based estimator, that replaces the squared intraday returns by normalized squared ranges. Christensen et al. (2010b) introduced the quantile based realized variance (QRV) estimator. Andersen et al. (2009) introduced the duration based approach based on a localization argument and the theory of Brownian passage times. In the multivariate case Barndorff-Nielsen et al. (2008) provided the multivariate realized kernel estimator, which guarantees consistency, positive semi-definiteness and handles non-synchronous trading simultaneously. Christensen et al. (2010a) introduced the adjusted modulated realized covariance. Lunde et al. (2016) proposed the composite kernel.

Financial risk managers often report the risk of investments using the concept of Value-at-Risk (VaR), which estimates the maximum loss at given confidence interval in a certain period. The Basel Committee demands banks use the VaR in establishing the minimum capital necessary at investments in order to reduce the fragility of international active banks. VaR is widely used by investors and regulators in the financial industry to measure the amount of assets needed to cover possible losses and has been considered as a expectable banking risk measure. The VaR is obtained by $\sigma_t F_L^{-1}(\alpha)$ at a given confidence interval $(1-\alpha)$, where $F_L^{-1}(\alpha)$

is loss distribution and σ_t is the standard deviation (volatility) of investments at time t. By calculating VaR, σ_t play a crucial role. Accurate volatility estimates are vital in risk management for calculating such as VaR. Further analyzing and forecasting the estimated volatility is also of paramount importance. In this thesis the forecasting of long memory and a nonparametric scale function in nonnegative financial processed based on the Semi-FI-Log-ACD model is proposed. A linear predictor based on the truncated AR(∞) form of the logarithmic process is proposed. The proposals are applied to forecasting such as realized volatility, trading volume. Furthermore, we propose new consistent estimators for realized kernels under independent and dependent noise. The comparison of proposed estimators with several other volatility estimates are reported. All practical studies in this thesis are based on high-frequency financial data from several European stocks.

1.2 Forecasting based on the Semi-FI-Log-ACD model

Modeling and forecasting of short- and long memory, and a possible nonparametric scale function in financial time series is of great interest. Well known models with short memory are e.g. the ARCH and GARCH for returns and the ACD for transaction durations. The ACD can also be used for modeling trading volume (Manganelli, 2005). Models based on logarithmic transformation are also proposed, including the Log-GARCH (Geweke, 1986, and Pantula, 1986) and the Log-ACD$_1$ (Bauwens and Giot, 2000, Bauwens et al., 2008, Karanasos, 2008). Now, the estimates are always nonegative and the log-data can be modeled by known linear time series approaches. Modeling of a smooth scale function in volatility caused by changing macroeconomic environment was investigated by Feng (2004) and Engle and Rangel (2008). Well-known long memory volatility models are the FIGARCH, the LM-GARCH, the FIACD (Jasiak, 1998) and the LM-ACD (Karanasos, 2004). In the literatures, the estimation of a nonparametric scale function in volatility models with long memory is not yet well studied.

Most recently, an EFARIMA (exponential fractional autoregressive integrated moving average) for nonnegative processes with long memory is proposed by Beran et al. (2015), called a Log-FARIMA in this thesis. The Log-FARIMA is defined as follows. Let $X_t^* = \lambda_t \eta_t$, where $\eta_t \geq 0$ are i.i.d. random variables and λ_t is the conditional mean. Let $Z_t = \ln(X_t^*)$ and $\varepsilon_t = \ln(\eta_t)$. It is assumed that $Z_t = \ln(X_t^*)$ is a FARIMA

$$(1-B)^d \phi(B) Z_t = \psi(B)\varepsilon_t,$$

where $d \in (-0.5, 0.5)$ is the fractional differencing parameter, $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $\psi(B) = 1 + \psi_1 B + \ldots + \psi_q B^q$ are the MA- and AR-polynomials with no common factor and all roots outside the unit circle. This model can be extended to a Log-SEMIFAR (logarithmic semiparametric fractional autoregressive) $X_t = \nu(\tau_t) X_t^* = \nu(\tau_t)\lambda_t \eta_t$, where $\tau_t = t/n$ and $\nu(\tau_t) > 0$ is a smooth scale function. It can be shown that

$$\ln \lambda_t = \sum_{i=1}^{\infty} \pi_i \ln \lambda_{t-i} + \sum_{j=1}^{\infty} \omega_j \ln(\eta_{t-j}),$$

where π_i are coefficients of $\pi(B) = (1-B)^d \phi(B) = 1 - \sum_{i=1}^{\infty} \pi_i B^i$. The Log-FARIMA is hence a fractionally integrated extension of the Log-ACD$_1$ (Bauwens et al., 2008) and is also called a FI-Log-ACD. The Log-SEMIFAR will hence be called a Semi-FI-Log-ACD. Both models are useful tools for modeling nonnegative financial data and can be estimated using existing software packages.

In chapter 2 necessary and sufficient conditions for the existence of a stationary solution of the FI-Log-ACD are first obtained. It is shown that these conditions are fulfilled, if η_t are log-normal innovations with $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Further examples, which fulfill those conditions are the log-logistic and log-Laplace innovations with suitable restriction on the parameters. In contrast it is shown that, if $d > 0$ those conditions can never be fulfilled by some well known nonnegative distributions, such as the weibull and the generalized gamma distributions. For those distributions a stationary FI-Log-ACD process with long memory in Z_t does not exist. Detailed properties of the FI-Log-ACD under the log-normal assumption as

obtained in Beran et al. (2015) are summarized. Now, all of the processes X_t^*, Z_t, λ_t as well as $\zeta_t = \ln(\lambda_t)$ have long memory with the same memory parameter, if $d > 0$. If $d \leq 0$, X_t^* and λ_t have short memory, which cannot be antipersistent.

Forecasting using the Semi-FI-Log-ACD is then discussed in detail. In chapter 2 we propose to use a simple, truncated linear predictor based on the AR(∞) form of Z_t. This idea is often employed to carry out forecasting based on an ARMA model, when the sample size is large. See e.g. Brockwell and Davis (2006, p. 184). In practice, the approximately best linear predictor based on the truncated part of the AR(∞) representation of Z_t is defined by

$$\hat{Z}_{n+k} = \sum_{j=1}^{k-1} \hat{\beta}_j \hat{Z}_{n+k-j} + \sum_{j=k}^{n+k-1} \hat{\beta}_j \hat{Z}_{n+k-j},$$

where $\hat{\beta}_j$ are the estimated coefficients in the AR(∞) form of Z_t, $\hat{Z}_{n+k-j}$ for $j = k, ..., n+k-1$, are the residuals obtained; and $\hat{Z}_{n+k-j}$, $j = 1, ..., k-1$, are the values predicted previously. To our knowledge, in the literature the above-defined approximately best linear predictor has not been proposed in the presence of long memory. Properties of this proposal are investigated in detail. It is shown that, in the presence of long memory the proposed predictor is still an approximately best linear predictor. Asymptotic variances of the prediction errors for an individual observation and for the conditional mean are obtained. Calculation of approximate forecasting intervals under log-normal assumption is discussed. Effect of the errors in the estimated trend on the asymptotic properties of the proposed predictor is also investigated.

The Semi-FI-Log-ACD is then applied for modeling and forecasting daily trading volumes, daily trading numbers and realized volatility. The data are from four European stocks, namely AF, BMW, PSA, MEOG. Application to real datasets shows that the proposed linear predictor works very well in practice. It is also confirmed that the Semi-FI-Log-ACD model is very useful for modeling different kinds of financial data, in particular aggregated financial data. Furthermore, note that the nonparametric trend and the long memory error process should be esti-

mated simultaneously, because suitable estimation of both parts will improve the forecasting. The more important reasons for this are as follows. On the one hand, if possible long memory in the conditional mean of a process is not considered, the selected bandwidth will be much smaller than it should be and the formula for calculating the asymptotic variance is also wrong. This will lead to a significant trend, even if the underlying process is indeed stationary. On the other hand, if an existing nonparametric scale function is not considered, it will be misinterpreted as very strong long memory. The results indicate that this model is widely applicable and the proposed linear predictor works very well in practice. It is also shown that the log-normal distribution is a suitable choice for different kinds of aggregated financial data.

1.3 An iterative plug-in algorithm for RK

One of the most important concepts in financial econometrics and financial mathematics introduced in the last two decades is the realized volatility (RV) (Andersen et al., 2001a,b), which is a model-free estimator of the daily integrated volatility (IV) based on high-frequency financial data. Let p_i^* be the latent logarithmic asset price observed at $0 = \tau_0 < \tau_1 < \ldots < \tau_n < \tau_{n+1} = T$, and p_i^* determined by the stochastic differential equation $dp^*(\tau) = \sigma(\tau)dW(\tau)$, where $W(\tau)$ is a standard Brownian motion and $\sigma(\tau)$ is the volatility process. Furthermore, assume that the $\sigma(\tau)$ and $W(\tau)$ processes are uncorrelated. The estimation of the daily integrated volatility is defined by $\text{IV} = \int_0^T \sigma^2(\tau)d\tau$. RV is constructed as the square root of the sums of intraday squared returns. However, it is found that, if the data exhibit microstructure noise u_i, most of the simple definitions of RV are now inconsistent estimators of the IV, and the observed log-price p_i consists of two components $p_i = p_i^* + u_i$, where u_i is i.i.d independent of p_i^* with zero mean and $var(u_i) = \omega^2$. Let $r_i^* = p_i^* - p_{i-1}^*$ be the returns with n observations per day. The corresponding noise contaminated returns is given by $r_i = r_i^* + e_i$, where $e_i = u_i - u_{i-1}$, which indicates an MA(1) error structure in observed returns.

In the presence of the MS noise, the commonly used biased estimator of IV is

$$\mathrm{RV}_0 = \sum_{i=1}^{n}(r_i^*)^2 + 2\sum_{i=1}^{n} r_i^* e_i + \sum_{i=1}^{n} e_i^2,$$

where its bias $\mathrm{B}(\mathrm{RV}_0) = 2n\omega^2$ and its variance is $var(\mathrm{RV}_0) = 4nE(u_i^4) + \frac{2}{n}\int_0^T \sigma_t^4(\tau)d\tau$.

Under the i.i.d noise assumption, Zhou (1996) proposed to use the first order correlation to correct the bias in RV_0. Following the original idea of Zhou the past one and the further one values are required. In order to facilitate practical use, we modify Zhou's approach based on the observed returns

$$\mathrm{RV}_{\mathrm{Z}} = \sum_{i=2}^{n-1}(r_i^2 + r_i r_{i+1} + r_i r_{i-1}).$$

It can be easily proved that volatility calculated by this approach is unbiased, while its variance tends to be infinite, if the number of the observation is large enough. There is an optimal observation frequency which trades off between bias and the variance of RV_{Z}.

In the literature, Bandi and Russel (2006), as well as Oomen (2006) investigate the optimal lower frequency, such as 5 minute and 15 minute. Zhang et al. (2005) and Aït-Sahalia et al. (2008) proposed to use two time scales estimator to solve the MS bias. Hansen and Lunde (2006a) and Oomen (2005) proposed a simple kernel based estimator. Barndorff-Nielsen et al (2008) proposed an approach to get the optimal observation frequency. Besides, Hansen et al. (2008) investigate microstructure bias based on MA filter. Most recently, Barndorff-Nielsen et al. (2008, 2009 and 2011b) introduced the realized kernels (RK), which is a consistent positive semi-definite estimators of the time-varying volatility. The realized kernels are given by

$$\mathrm{RK} = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \quad \gamma_h = \sum_{j=|h|+1}^{n} r_j r_{j-|h|},$$

where $k(u)$ is a kernel weight function, H is the selected bandwidth and γ_h is the

h-th realized autocovariance. We use the Parzen kernel, which is one of the kernels that can guarantee a nonnegative realized kernel estimate. Asymptotic properties of RK are studied by Barndorff-Nielsen et al. (2008, 2011b), see also Ikeda (2015).

A crucial problem when applying RK is the selection of the bandwidth. The main purpose of chapter 3 is to propose a simple, fast and fully data-driven consistent bandwidth selector for RK based on the iterative plug-in idee (Gasser et al., 1991). The asymptotically optimal bandwidth (Barndorff-Nielsen et al., 2009, 2011b), which minimizes the dominating part of the MSE (mean squared error) of an RK is given by

$$H_{\mathrm{A}} = c_0 \xi^{4/5} n^{3/5} \quad \text{with} c_0 = \left\{ \frac{k''(0)^2}{k_{\bullet}^{0.0}} \right\}^{1/5} \quad \text{and} \quad \xi^2 = \frac{\omega^2}{\sqrt{T \int_0^T \sigma(\tau)^4 d\tau}}.$$

For the Parzen kernel $c_0 = 3.5134$. We see the optimal bandwidth for an RK with a non-flat top kernel is of the order $O(n^{3/5})$. If a bandwidth of this order is employed, the resulted RK will achieve its optimal convergence rate of the order $O(n^{-1/5})$. In chapter 3, to simplify the estimation procedure we use a biased version of the asymptotically optimal bandwidth of RK, called H_{B} by replacing $T \int_0^T \sigma^4(\tau) d\tau$ with IV^2. The reason is that the former is not far from the latter, if $\sigma(\tau)$ does not vary too much. The bandwidth H_{B} is obtained by

$$H_{\mathrm{B}} = c_0 \xi_{\mathrm{B}}^{4/5} n^{3/5} \quad \text{with} \quad c_0 = \left\{ \frac{k''(0)^2}{k_{\cdot}^{0.0}} \right\},$$

where $\xi_B^2 = \hat{\omega}^2 / \hat{\mathrm{IV}}$ and $\hat{\omega}^2 = \frac{\mathrm{RV}_0 - \hat{\mathrm{IV}}}{2n}$. The term "consistent" is used in a relative sense that $(\hat{H}_{\mathrm{B}} - H_{\mathrm{B}})/H_{\mathrm{B}} \to 0$, as $n \to \infty$. The selected bandwidth $\hat{H}_{\mathrm{B}}$ is computed by means of an iterative procedure. In each iteration, the resulting RK is used as an estimate of the IV, and the variance of the MS noise ω^2 is estimated based on the difference between RV_0 and RK. In the first iteration RV_{Z} is used as the initial value of RK. It is shown that $\hat{\omega}^2$ defined in this way is $\sqrt{n}$-consistent in each iteration. Both of RK and $\hat{H}_{\mathrm{B}}$ become consistent form in the third iteration, while their rate of convergence can still be improved in the fourth iteration. Thereafter, RK achieves its optimal rate of convergence of the order

$O(n^{-1/5})$ and this rate of convergence is also shared by $(\hat{H}_{\mathrm{B}} - H_{\mathrm{B}})/H_{\mathrm{B}}$.

An R code is developed for practical implementation of the proposed bandwidth selector. The algorithm will be stopped, if $\hat{H}_j$ is equal to $\hat{H}_{j-1}$. It is also found that the selected bandwidths sometimes take two consequent integers alternatively. The procedure will also be ended in this case and the larger of the two selected bandwidths will be used. Both mentioned cases will be considered as regular cases. Note that in our implementation, $\hat{H}$ is obtained by truncating the integer part of selected optimal bandwidth and plus 1. The end effect of realized kernels can be eliminated in the computation of the realized kernels (Barndorff-Nielsen et al., 2011b). Furthermore, except for regular cases there are also three cases, which should be specially treated. The first special case (Sp. Case 1) is that RV_{Z} is smaller than 0. In this case the noise may be very strong. We manually set $\hat{\xi}^2 = 100/(2n)$, which can select a big starting bandwidth. After j iterations H_j will converge to the optimal bandwidth H_{B}. The second special case (Sp. Case 2) is that RV_{Z} is bigger than RV_0, which indicates probably there is no noise in those days. In this case, we use estimator RV_0, because in the case of no noise RV_0 is unbiased and we manually set $\hat{H}^* = 0$. The third special case (Sp. Case 3) is that RK is bigger than RV_0, which means there may be positive noise in returns on those days. In this case, we still use estimator RK.

In chapter 3 the nice practical performance of the proposal is illustrated by application to data of two German and two French firms within a period of several years. The proposal runs very quickly and works usually very well in practice. Detailed analysis of two challenging cases are reported. In Sp. Case 2 the ACF at lag 1 is positive, however, some ACFs at higher lags can be negative so that RV_0 is still biased. One possible problem, which can arise is that the dependent MS noise appears. In Sp. Case 3 the ACF at lag 1 is negative, however some other ACFs are clearly positive so that the sum of the ACF is positive. This indicated the existence of possible dependent MS noise again, which could cause a negative bias in RV_0. The effect of this kind of possibly dependent MS noise can however not be corrected by the proposed data-driven RK and it is worthy to development a data-driven RK by taking possibly dependent MS noise into account.

Furthermore, the ESEMIFAR model, which can simultaneously investigated long memory, nonparametric trends and possible structural breaks, is used to analyze the RK. Possible structure breaks cased by the financial crisis in 2008 may have a clear effect on the estimation results. Using piecewise ESEMIFAR model can improve the quality of estimation results.

1.4 RK under dependent noise using different sampling frequencies

Further improving the fully automatic iterative plug-in algorithm is the focus of chapter 4. A data-driven RK under dependent noise assumption is proposed and now the algorithm works for all cases. Moreover, we extent the proposed algorithm for the different sampling frequencies and then compare them with several other realized estimators. In total, we have 4 sampling frequencies (tick-by-tick, 1-minute, 5-minute and 15-minute), 2 types of realized estimators (RV_0 and RK) and two algorithms (bandwidth selectors under independent and dependent noise assumptions) of a given transaction price series.

Realized volatility using different sampling frequencies is obtained by

$$\mathrm{RV}_0^s = \sum_{i=0}^{n_s} r_{t,i}^2,$$

where n_s denotes the number of observations using different sampling frequencies. The version of RV_Z with different sampling frequencies is given by

$$\mathrm{RV}_\mathrm{Z}^s = \sum_{i=2}^{n_s-1} (r_{t,i}^2 + r_{t,i}r_{t,i+1} + r_{t,i}r_{t,i-1}).$$

By displaying the ACFs of returns on several selected days it can be found that in most cases $\rho_r(1)$ is significantly negative and there are possible dependent MS noise in tick-by-tick returns and for some cases there may be still MS noise in 1-minute and 5-minute returns. The correlations reduce with the decrease of sample

frequencies. The reason is that the diminished number of observations reduces the bias and extends the two ACF bounds, which makes the daily returns uncorrelated. However, decreasing the sampling frequency toward to increase the variance and reduce the accuracy of volatility estimators. Hence, a realized estimator with an optimal frequencies is necessary to investigate, which can remove the effect of MS noise and grantee the accuracy of the estimation at the same time.

The proposed asymptotically optimal bandwidth for RK is given by

$$H = c_0 \xi^{4/5} n^{3/5} \quad \text{with} \quad c_0 = \left\{ \frac{k''(0)^2}{k_{\cdot}^{0.0}} \right\}^{1/5} \quad \text{and} \quad \xi^2 = \frac{\Omega}{\sqrt{T\mathrm{IQ}}},$$

where $c_0 = 3.5134$ for the Parzen kernel, IQ$= \int_0^T \sigma_t^4 dt$ is called the integrated quarticity, $\Omega = \Sigma_{h>0}\Omega(h)$ is the long-run variance of $u_{t,i}$ and $\Omega(h) = \Sigma_{i=h+1}^{n} u_{t,i}u_{t,i-h}$. The bandwidth H depends on the unknown quantities Ω and IQ. When $u_{t,i}$ is under i.i.d. noise assumption, Ω reduces to ω^2. We let $\hat{\omega}^2_{s,1} = \frac{\mathrm{RV}_0^{\mathrm{s}}}{2n_s}$ to replace $\hat{\omega}^2_{s,1} = \frac{\mathrm{RV}_0^{s} - \mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}}}{2n_s}$. It aims to avoid $\mathrm{RV}_0^s - \mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}} < 0$. It can be observed that $\frac{\mathrm{RV}_0^{\mathrm{s}}}{2n_s}$ is not far from $\frac{\mathrm{RV}_0^{s} - \mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}}}{2n_s}$, if $n \to \infty$. The algorithm after this adjustment is called the IN algorithm.

However, this IPI algorithm is only for the case under assumption of i.i.d. MS noise. Ikeda (2015) proposed a two-scale RK, which is a convex combination of two realized kernels with different bandwidths. He showed that his estimator converges to the daily integrated volatility in the presence of dependent MS noise. Following Ikeda (2015) we utilize $MK(G)$ to estimate Ω. Define

$$MK(G) = (|k''(0)|nG^{-2})^{-1}\mathrm{RK}(G),$$

where RK(G) are the realized kernels for bandwidth G. The asymptotically optimal bandwidth $G = O(n^{\beta})$ can be obtained by minimizing the AMSE$_G$. When $\beta = 1/3$ only for Parzen kernel, which is allowed for the asymptotic normality of MK(G). According to the related results in Ikeda (2015), we can get the AMSE$_G$ for MK(G)

$$\mathrm{AMSE}_G = \left(\frac{\mathrm{IV}_t}{k''(0)} \right)^2 \frac{G^4}{n^2} + O\left(\frac{G}{n} \right) + O\left(\frac{1}{G^2} \right).$$

Based on Ikeda (2015)'s conclusion, Wang (2014) proposed two IPI algorithms for RK under dependent noise assumption. He proved that his proposals are consistent and that for algorithm A in the second iteration $\hat{H}_2$ has already achieved the rate of convergence of the order $O(n^{3/5})$ and after only a few iterations $\hat{H}_j$ converges to H^{B}. He also showed that $\hat{H}^B$ is consistent even when α is outside the range of $(1/2, 1)$. Algorithm B is a fully automatic data-driven algorithm and with our slight adjustment it is called the DN algorithm. The starting bandwidth in the DN algorithm is more reasonable and end effects in the computation of the RK are considered. The processes of the IN and DN algorithms starting with different bandwidth for four selected days are displayed in Fig. 4.5. It shows that both algorithms work very well in practice and only a few iterations are required. In addition, no matter what the starting bandwidths, the selected bandwidths as of the second iteration are very close and that the final selected bandwidths are indeed the same. Normally, the required iteration number becomes smaller, if a suitable starting bandwidth is chosen. The histograms of selected bandwidths for RK-IN-tick, RK-IN-1min, RK-IN-5min and RK-DN-tick of all companies are illustrated. The selected bandwidths for RK-IN-tick are smaller than those for RK-DN-tick. The reason is that the dependent MS noise is taken into account. Meanwhile, the selected bandwidths for 1-minute RK-IN are smaller than those for tick-by-tick RK-IN, but larger than those for 5-minute RK-IN. This corresponds to their ACFs results, that the correlations in 5-minute returns are smaller than the correlations in 1-minute and tick-by-tick returns. The commonly required number of iterations for RK-IN-tick, RK-IN-1min, RK-IN-5min and RK-DN-tick are 3, 3, 3, 5, respectively. These confirm that both IN and DN algorithms work very well in practice and only a few iterations are required. Please note that the IN algorithm does not work for 15-minute RK of all ten stocks, because on some days for these stocks the number of observations are smaller than the selected bandwidths ($n < \hat{H}^{\mathrm{B}}$). Meanwhile, the number of three cases mentioned in Feng and Zhou (2015a) ($\mathrm{RV_Z} < 0$, $\mathrm{RV_Z} > \mathrm{RV_0}$ and $\mathrm{RK} > \mathrm{RV_0}$) for RK-IN-tick, RK-IN-1min, RK-IN-5min and RK-DN-tick of the ten companies are listed, which shows the necessity of adjusting the old algorithms to the proposed IN and DN

algorithms.

Further comparison of these realized estimators is carried out by assessing their performances in the computation of Value-at-Risk (VaR) based on the lnRV-SEMIFAR (also called the Semi-FI-Log-ACD model). The corresponding VaR calculation is obtained by

$$\mathrm{VaR}_{1-\alpha}^{t} = s(\tau_t)\sigma_t Z_{1-\alpha},$$

where $s(\tau_t)$ is local variance, σ_t is the volatility of stocks at time t and the loss distribution $Z_{1-\alpha}$ is assumed normal. In the SEMIFAR model the total means in the original data $s(\tau_t)\sigma_t$ is obtained through the exponential transformation of the estimated deterministic trend and the estimated conditional mean.

A Backtesting for comparing the observed amount of exceptions (points over VaR) with the expected amount/benchmark is utilized. The one day dynamic ahead 95% VaR based on the lnRV-SEMIFAR model together with the losses for the ten companies are illustrated in Figure 4.8 to Figure 4.17. The corresponding numerical results for points over VaR and the deviations from the benchmark are listed in Table 4.4 and Table 4.5. It is found that RK-IN-tick and RK-DN-tick estimators have good performances by computing VaR and are hence recommended using as the estimators of IV in practice.

Chapter 2

Forecasting financial market activity using a semiparametric fractionally integrated Log-ACD

This chapter is based on joint work with Yuanhua Feng and published with slight differences in the International Journal of Forecasting 31 (2015) 349-363.

2.1 Introduction

This chapter considers forecasting of long memory and a smooth scale function in financial time series aggregated from high-frequency data, such as (daily) trading volumes, trading numbers, average transaction durations as well as realized volatility. Here, long memory is probably caused by aggregation. Well known short-memory models in the current context are the ARCH (autoregressive conditional heteroskedasticity, Engle, 1982) and GARCH (generalized ARCH, Bollerslev, 1986) for returns and the ACD (autoregressive conditional duration, Engle and Russell, 1998) for transaction durations. The ACD can also be applied to trading volumes (Manganelli, 2005) and other quantities. Furthermore, the (first type) Log-ACD (Bauwens and Giot, 2000, Bauwens et al., 2008, Karanasos, 2008, Allen et al., 2008) is also proposed, where the log-data are modeled by an ARMA.

The idea ensures that the estimates of the original data are always nonnegative. Modeling a smooth scale function in volatility caused by changing macroeconomic environment was investigated by Feng (2004), and Engle and Rangel (2008).

Well known long memory volatility models are the FIGARCH (fractionally integrated GARCH, Baillie et al., 1996), the LM-GARCH (long memory GARCH, Karanasos et al., 2004), the FIACD (Jasiak, 1998) and the LM-ACD (Karanasos, 2004). Baillie and Morana (2009) proposed an adaptive FIGARCH for modeling long memory and structural breaks in volatility. As far as we know, estimating a smooth scale function in volatility models with long memory is not yet well studied. Most recently, Beran et al. (2015) proposed to model short memory, long memory and a nonparametric scale function in nonnegative financial time series based on the log-transformation. They assumed that the process under consideration has a log-normal marginal distribution and proposed to model the stochastic component of the log-data by a Gaussian FARIMA (fractional autoregressive integrated moving average, Hosking, 1981, Beran et al., 2013). The log-data themselves are analyzed by a SEMIFAR (semiparametric fractional autoregressive, Beran and Feng, 2002a). Their proposals are hence called an EFARIMA (exponential FARIMA) and an ESEMIFAR, respectively, which can be easily estimated using existing software packages.

In this chapter, a possible origin of long memory in the data under consideration is discussed. The EFARIMA and ESEMIFAR models are extended to the case with general marginal distributions and represented as FI-Log-ACD and Semi-FI-Log-ACD models, respectively. Necessary and sufficient conditions for the existence of a stationary solution of the FI-Log-ACD are obtained. Detailed properties of this model under the log-normal assumption are investigated. In particular, the long memory parameter is now not affected by the log-transformation (see also Dittmann and Granger, 2002). Our focus is on forecasting using the Semi-FI-Log-ACD based on an improved data-driven SEMIFAR algorithm. The forecasting is carried out using a truncated linear predictor based on the AR(∞) form, which is more simple and runs faster than the best linear SEMIFAR predictor of Beran and Ocker (1999). Properties of the proposal are investigated in

detail. For an ARMA model the proposed predictor is an approximately best linear predictor (Brockwell and Davis, 2006, p. 184). We show that this is still true in the presence of long memory. Asymptotic variances of the prediction errors for an individual observation and for the conditional mean are obtained. The calculation of approximate forecasting intervals under the log-normal assumption is discussed. The effect of the errors in the estimated trend on the asymptotic properties of the proposed predictor is investigated. Application to (daily) trading volumes, trading numbers, average durations and realized volatility shows that the proposals work very well in practice and the log-normal distribution is quite reasonable. Finally, our empirical results reveal that, besides long memory and smooth scale change, realized volatility may also exhibit structural breaks.

Motivations for this study and definitions of the models are given in section 2.2. section 2.3 describes the properties and estimation of the proposed models. The linear predictor is proposed and studied in section 2.4. section 2.5 reports the empirical results. Final remarks in section 2.6 conclude the chapter. Proofs of results are put in the appendix.

2.2 A semiparametric multiplicative long memory model

2.2.1 Origin of long memory in aggregated financial data

Nonnegative financial time series often exhibit long memory. Long memory in realized volatilities and trading volumes has been studied by Andersen et al. (2001a), and Fleminga and Kirbyb (2011). Deo et al. (2010) revealed that transaction durations, trading numbers, squared returns and realized volatility may exhibit long memory at potentially the same level. Beran et al. (2015) found evidence of long memory in average durations.

A well known theoretical origin of long memory in economic time series is the cross-sectional aggregation of microeconomic data. For instance, Granger (1980) and Leipus et al. (2014) showed that the aggregation of random-coefficient

ARMA processes under certain conditions will result in a long-memory process. Zaffaroni (2007) discussed the aggregation of GARCH models and indicated that this does not lead to long memory in volatility. See also Beran et al. (2013) and references therein. Time series considered in this chapter can be thought of as special cross-sectional aggregates of high-frequency data, i.e. aggregation of micro-financial data in some sense. For instance, define the realized volatility based on 1-minute (log-) returns as the sum of squared returns (Andersen et al., 2001a). If returns at a given time point on a trading day follow a GARCH model, realized volatility will be an aggregate of squared GARCH processes. Although a squared GARCH process corresponds to a nonlinear ARMA model, the stationary conditions for such (squared) processes are quite different to those for linear ARMA models. Hence, results in Granger (1980) and Leipus et al. (2014) do not apply to realized volatility. Results of Zaffaroni (2007) on the aggregation of GARCH models also do not apply to the aggregation of squared returns. Realized volatility used in the application in section 2.5 is indeed defined as the sum of the squared ultra-high-frequency returns (without considering the effect of the microstructure noise). Now, discussing the origin of long memory in such time series would be even more difficult. To our knowledge, theoretical models to explain the origin of long memory in financial time series aggregated from high-frequency data are still unknown. But we believe that it is mainly caused by aggregation.

2.2.2 Simultaneously modeling long memory and scale change

A well known model for a stationary nonnegative financial time series, X_t, $t = 1, ..., n$, is the MEM (multiplicative error model, Engle, 2002) defined as

$$X_t = \nu \lambda_t \eta_t, \tag{2.2.1}$$

where $\nu > 0$ is a scale parameter, $\lambda_t > 0$ is the conditional mean of $X_t^* = X_t/\nu$ and $\eta_t \geq 0$ are i.i.d. random variables. In order to model long memory and a slowly changing scale function simultaneously, we propose the use of a semiparametric MEM model by replacing ν in Eq. (2.2.1) with a nonparametric scale function

$\nu(\tau) > 0$:

$$X_t = \nu(\tau_t)X_t^* = \nu(\tau_t)\lambda_t\eta_t, \tag{2.2.2}$$

where $\tau_t = t/n$ denotes the rescaled time. Let $Y_t = \ln(X_t)$, $\mu(\tau_t) = \ln[\nu(\tau_t)]$, $\zeta_t = \ln(\lambda_t)$, $\varepsilon_t = \ln(\eta_t)$ and $Z_t = \zeta_t + \varepsilon_t$, where $\mu(\tau_t)$ and ζ_t are the local and conditional means of Y_t, respectively. Following Beran et al. (2015), we assume that $E(\varepsilon_t) = 0$, $var(\varepsilon_t) = \sigma_\varepsilon^2$, and the stochastic component Z_t follows a FARIMA

$$\pi(B)Z_t = \psi(B)\varepsilon_t, \tag{2.2.3}$$

where $\pi(B) = (1-B)^d\phi(B) = 1 - \Sigma_{i=1}^{\infty}\pi_i B^i$ with $\pi_i \approx c_\pi i^{d-1}$ for large i, $d \in (-0.5, 0.5)$ is the fractional differencing parameter, $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $\psi(B) = 1 + \psi_1 B + \ldots + \psi_q B^q$ are the AR- and MA-characteristic polynomials with no common factor and all roots outside the unit circle. The model defined by Eqs. (2.2.1) and (2.2.3) is called an exponential FARIMA (EFARIMA), which is a nonnegative process whose log-transformation follows a FARIMA. The model defined by Eqs. (2.2.2) and (2.2.3) will be called an ESEMIFAR, because $Y_t = \ln(X_t) = Z_t + \mu(\tau_t)$ follows a SEMIFAR (Beran and Feng, 2002a) with the integer integration parameter $m = 0$ and an additional MA part. Because of Eq. (2.2.3), we have

$$\pi(B)\zeta_t = [\psi(B) - \pi(B)]\varepsilon_t.$$

Beran et al. (2015) indicated that the EFARIMA model can be written as a fractionally integrated generalization of the (first type) Log-ACD model. The reason is that, similar to Eq. (7) of Bauwens et al. (2008), the conditional mean of Z_t can be represented as

$$\zeta_t = \ln\lambda_t = \sum_{i=1}^{\infty}\pi_i \ln\lambda_{t-i} + \sum_{j=1}^{\infty}\omega_j \ln(\eta_{t-j}), \tag{2.2.4}$$

where $\omega_j = \pi_j + \psi_j$ for $1 \le j \le q$, and $\omega_j = \pi_j$ for $j > q$. The model defined by Eqs. (2.2.1) and (2.2.4) will be called a FI-Log-ACD. And Eqs. (2.2.2) and (2.2.4) define a semiparametric generalization of the FI-Log-ACD, called a Semi-

FI-Log-ACD. The Log-ACD (p, q) model is the special case with $d = 0$. Moreover, according to Eq. (2.2.3), ζ_t can also be rewritten as

$$\begin{aligned}\zeta_t &= Z_t - \varepsilon_{t-j} = Z_t - \varepsilon_{t-j} - \pi(B)Z_t + \psi(B)\varepsilon_t \\ &= [1 - \pi(B)]Z_t + [\psi(B) - 1]\varepsilon_t.\end{aligned} \tag{2.2.5}$$

The relationship between the FI-Log-ACD and the EFARIMA models is given below.

Proposetion 2.2.1. *The EFARIMA model defined by Eqs.* (2.2.1) *and* (2.2.3)*, and the FI-Log-ACD model defined by Eqs.* (2.2.1) *and* (2.2.4) *are equivalent to each other.*

Proof of Proposition 2.2.1 is omitted. This result means that the proposed models are the application of the well known FARIMA and SEMIFAR models to the log-process. Hence, the log-transformation of a nonlinear (nonnegative) process following the FI-Log-ACD is assumed to be a linear process. The original process X_t^* is hence a log-linear process.

2.3 Properties and estimation of the models

2.3.1 The stationary solutions

In the following, some results in Beran et al.'s (2015) under the log-normal assumption are extended to more general distributions. Let $\alpha(B) = (1-B)^{-d}\phi^{-1}(B)\psi(B) = 1 + \sum_{i=1}^{\infty}\alpha_i B^i$. Note that the stationary solution of the FARIMA process Z_t is given by

$$Z_t = \sum_{i=0}^{\infty}\alpha_i\varepsilon_{t-i} \tag{2.3.1}$$

with $\alpha_i \approx c_\alpha i^{d-1}$ for large i, and, for large k, the autocorrelation (ACF) of Z_t is given by

$$\rho_Z(k) \approx c_\rho^Z |k|^{2d-1}, \tag{2.3.2}$$

where c_ρ^Z is a constant. Note that $c_\rho^Z > 0$ for $d > 0$ and now Z_t has long memory. Let $\alpha_{\max} = \max(\alpha_i)$ and $\alpha_{\min} = \inf(\alpha_i)$, where $\alpha_{\max} \geq 1$ and $\alpha_{\min}$ may be negative.

Conditions for the existence of a stationary solution of X_t^* in the current case with $2u$-th finite moment are similar to those given by Karanasos (2008).

A1. Z_t is a stationary and invertible FARIMA process as defined in Eq. (2.2.3).

A2. Both $E(\eta_t^{2u\alpha_{\max}})$ and $E(\eta_t^{2u\alpha_{\min}})$ are finite for some $u > 0$.

Now, the stationarity solution of X_t^* is given by

$$X_t^* = \prod_{i=0}^{\infty} \eta_{t-i}^{\alpha_i}. \tag{2.3.3}$$

Lemma 2.3.1. *The solution of X_t^* given in* (2.3.3) *is strictly stationary with finite $2u$-th moment, if and only if A1 and A2 hold. If A2 holds for $u \geq 1$, X_t^* is also weakly stationary.*

The proof is similar to that of Lemmas and from Karanasos (2008), and is omitted.

A2 ensures that all of the terms in the product in Eq. (2.3.3) exist. A1 implies that $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$ and $E(\varepsilon_t) = 0$. This together with A2 ensures the convergence of X_t^* defined in Eq. (2.3.3). The condition $E(\varepsilon_t) = 0$ is different to the typical assumption $E(\eta_t) = 1$ used in an ACD model. For instance, if η_t is exponentially distributed with the density $f(u) = \mu_\eta^{-1} \exp(-u/\mu_\eta)$, we have $\mu_\eta = \exp(\gamma) \approx 1.781$, where γ is the Euler constant. However, the restriction $E(\varepsilon_t) = 0$ is now necessary. Otherwise, the mean in Z_t and the scale in X_t^* are not well defined, because α_i are not summable. $E(\varepsilon_t) = 0$ is fulfilled, for example by:

Example 1. The log-normal innovations η_t with $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ and $\sigma_\varepsilon^2 > 0$,

Example 2. The log-logistic innovations η_t with $\varepsilon_t \sim Lo(0, b)$ and $b > 0$ or

Example 3. The log-Laplace innovations η_t with $\varepsilon_t \sim La(0, b)$ and $b > 0$.

Note that A2 may or may not be affected by d. The question of whether A2 is fulfilled or not is determined jointly by the distribution of η_t, the value of u and

the FARIMA coefficients. In Example 1, A2 is always fulfilled and X_t^* is strictly and weakly stationary. In Examples 2 and 3, X_t^* is only weakly stationary if b is small enough.

Furthermore, the stationary solutions of the conditional mean of Z_t and that of X_t^* are

$$\zeta_t = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{t-i} \text{ and } \lambda_t = \prod_{i=1}^{\infty} \eta_{t-i}^{\alpha_i}. \tag{2.3.4}$$

The forecasts of the FARIMA process Z_t and its conditional mean ζ_t to be proposed later are based on their AR(∞) representations, respectively. For Z_t we have

$$Z_t = \sum_{j=1}^{\infty} \beta_j Z_{t-j} + \varepsilon_t, \tag{2.3.5}$$

where β_j are the coefficients of $\beta(B) = (1-B)^d \phi(B) \psi^{-1}(B) = 1 - \sum_{j=1}^{\infty} \beta_j B^j$. For large j, we have $\beta_j \approx c_\beta j^{-d-1}$ with $c_\beta > 0$. This yields the representation of ζ_t based on Z_t:

$$\zeta_t = \sum_{j=1}^{\infty} \beta_j Z_{t-j}. \tag{2.3.6}$$

The stationary solutions of X_t^* and λ_t, respectively, can be rewritten as

$$X_t^* = \eta_t \prod_{j=1}^{\infty} (X_{t-j}^*)^{\beta_j} \text{ and } \lambda_t = \prod_{j=1}^{\infty} (X_{t-j}^*)^{\beta_j}. \tag{2.3.7}$$

2.3.2 Properties under the log-normal assumption

Beran et al. (2015) found that when aggregated financial data are considered, the log-normal assumption is usually a suitable choice. They hence studied the properties of the proposed models with log-normally distributed innovations in detail. In the following, their results will be first summarized briefly. Then we will focus on discussing the application of the Semi-FI-Log-ACD model. We will see that now all of Z_t, X_t^*, ζ_t and λ_t exhibit long memory. In particular, the authors showed that the stationary process X_t^* is also log-normally distributed, $X_t^* \sim LN(0, \sigma^2)$ with $\sigma^2 = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \alpha_i^2$, if ε_t are i.i.d. $N(0, \sigma_\varepsilon^2)$ random variables.

The closed form formula of the ACF of X_t^* can be obtained. Furthermore, in the presence of long memory it holds

$$\rho_{X^*}(k) \approx c_\rho^{X^*} |k|^{2d-1}$$

for large k, where $0 < c_\rho^{X^*} < c_\rho^Z$. We see that X_t^* is a long memory process with the same memory parameter d, if Z_t has long memory. This confirms the well known fact that the long memory parameter in Z_t and that in X_t^* under the log-normal assumption is the same (see Dittmann and Granger, 2002). The reason is that the Hermite rank of the exponential function is one. However, the constant in the asymptotic formula of $\rho_{X^*}(k)$ is smaller than that in $\rho_Z(k)$. If Z_t is a FARIMA with $-0.5 < d \leq 0$, it can be shown that $\sum \rho_{X^*}(k) > 0$. We see that X_t^* does not have antipersistence, even if Z_t is antipersistent (see also Dittmann and Granger, 2002). This leads to the very interesting fact:

Proposetion 2.3.1. *A FI-Log-ACD process X_t^* with log-normal marginal distribution cannot exhibit antipersistence.*

In financial econometrics, study on the long-memory property of the conditional means ζ_t and λ_t in Z_t and X_t^*, respectively, is also of great interest. The ACF of ζ_t with $d > 0$ is given by

$$\rho_\zeta(k) \approx c_\rho^\zeta |k|^{2d-1} \tag{2.3.8}$$

for large k, where $c_\rho^\zeta > c_\rho^Z$. From (2.3.8) we see that ζ_t also has long memory with the same memory parameter d. However, the constant in the asymptotic formula of $\rho_\zeta(k)$ is larger than that in $\rho_Z(k)$. And the ACF of λ_t for large k is given by

$$\rho_\lambda(k) \approx c_\rho^\lambda |k|^{2d-1},$$

where $0 < c_\rho^\lambda < c_\rho^\zeta$. Again, the long memory parameter in λ_t is d. But the constant in the asymptotic formula of the ACF after the exponential transformation is reduced.

2.3.3 Estimation of the models

From now on we will mainly consider the model of Z_t without the MA part for simplicity. But the theoretical discussion holds in the case, when Z_t follows a general FARIMA model. Now, let $\theta = (\sigma_\varepsilon^2, d, \phi_1, \ldots, \phi_p)^{\mathrm{T}}$ denote the unknown parameter vector of the SEMIFAR model. Under the normal assumption of ε_t, θ can be estimated from $\hat{Z}_t = Y_t - \hat{\mu}(\tau_t)$ by the approximate Gaussian maximum likelihood method (MLE). The AR order p can be selected consistently by the BIC. The trend will be estimated by a p_μ-th order local polynomial (Beran and Feng, 2002b) with a weighting function $K(u)$ and the bandwidth h, which does not share the boundary problem, if p_μ is odd. We mainly consider the use of $p_\mu = 1$ and $p_\mu = 3$, and put $l = p_\mu + 1$. The asymptotically optimal bandwidth is given by

$$h_A = C_A n^{(2d-1)/(2l+1-2d)} \text{ with } C_A = \left[\frac{1-2d}{2l\beta^2}\frac{(l!)^2 V}{I(\mu^{(l)})}\right]^{1/(2l+1-2d)}, \qquad (2.3.9)$$

where $I(\mu^{(l)}) = \int_0^1 [\mu^{(l)}(\tau)]^2 d\tau$ and V is a constant as defined in Beran and Feng (2002b).

An iterative plug-in algorithm for kernel estimator of $\mu(\tau)$ developed by Beran and Feng (2002a) following Gasser et al. (1991) is built-in S+ FinMetrics. In this chapter the algorithm of Beran and Feng (2002b) for local polynomial regression will be used, where $\mu^{(l)}$ is estimated by a $(l+1)$-th local polynomial with the bandwidth $h_{l,j}$, inflated from the selected bandwidth h_{j-1} in the $(j-1)$-th iteration. This algorithm processes as follows:

Step 1. Start with the bandwidth $h_0 = O(n^{-b})$ with $b = 1/(2l+1)$.

Step 2. In the jth iteration with $j > 1$, estimate μ using h_{j-1}, and d and V from $\hat{Y}_t$, where $\hat{Y}_t = Z_t - \hat{\mu}(\tau_t)$ and the AR order p is selected by BIC. Estimate $I(\mu^{(l)})$ using $h_{l,j} = (h_{j-1})^\alpha$ with $\alpha = (2l+1-2\hat{d})/(2l+3-2\hat{d})$.

Step 3. Improve h_{j-1} by

$$h_j = \left\{ \frac{1-2\hat{d}}{2l\beta^2} \frac{(l!)^2 \hat{V}}{\hat{I}[\mu^{(l)}(\tau; h_{l,j})]} \right\}^{1/(2l+1-2\hat{d})} \cdot n^{(2\hat{d}-1)/(2l+1-2\hat{d})} \tag{2.3.10}$$

Step 4. Increase j by one and repeatedly carry out Steps 2 and 3 repeatedly until convergence is reached or until a maximal number of iterations is achieved.

The starting bandwidth in Step 1 is roughly estimated under independent errors, which works very well in practice. As was shown by Beran and Feng (2002b), the α used in Step 2 is the asymptotically optimal choice, called α_{opt}. Two further reasonable choices of α are $\alpha_0 = (2l+1-2d)/(2l+5-2d)$, to minimize the MISE of $\hat{\mu}^{(l)}$, and $\alpha_{\text{var}} = 1/2$, to achieve the most stable bandwidth selector. The simulation study of Beran and Feng (2002c) showed that various different iterative plug-in SEMIFAR algorithms based on the kernel regression work very well. This should be the same in the current context, because the local polynomial regression and the kernel regression are asymptotically equivalent.

2.4 Forecasting based on the Semi-FI-Log-ACD

Now, we will discuss forecasting based on the Semi-FI-Log-ACD model, which is equivalent to the ESEMIFAR. Note that the ESEMIFAR is a SEMIFAR applied to the log-transformed data, the ESEMIFAR forecasting hence consists of two stages: 1) The forecasting based on the SEMIFAR model applied to the log-data, and 2) The calculation of the forecasts for the original data through exponential transformation. The former consists again of two parts: the extrapolation of the estimated trend function $\hat{\mu}(\tau_n)$ and the prediction of the stochastic part Z_{n+k}.

2.4.1 Extrapolation of the trend function

For the purpose of forecasting, we will propose the use of the local linear regression, because the local cubic approach is sometimes instable at the endpoint. The great advantage of the local linear estimator compared to a kernel estimator is that $\hat{\mu}$

has automatic boundary correction, i.e. the bias of $\hat{\mu}(\tau_n)$ at the endpoint τ_n is still of the order $O(h^2)$, while the bias of a kernel estimator at τ_n is of the order $O(h)$. We propose to forecast the trend $\mu(\tau_{n+k})$ in the future by linear extrapolation of $\hat{\mu}(\tau_n)$. Let $\Delta\mu = \hat{\mu}(\tau_n) - \hat{\mu}(\tau_{n-1})$. By means of the linear extrapolation, the forecasted trend $\hat{\mu}(\tau_{n+k})$ is given by

$$\hat{\mu}(\tau_{n+k}) = \hat{\mu}(1) + k\Delta\mu. \tag{2.4.1}$$

The following assumptions are required for further discussion.

A3. In A1, assume further that $d \in (0, 0.5)$, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, and $q = 0$ for simplicity.

A4. The weighting function $K(u)$ is a symmetric density on the compact support $[-1, 1]$.

A5. The trend function μ is at least four times continuously differentiable.

A6. The bandwidth h is selected by the data-driven algorithm proposed in section 2.3.3.

A4 and A5 are standard assumptions in nonparametric regression. A6 ensures that $\hat{\mu}$ achieves the optimal rate of convergence of the order $O(n^{(2d-1)l/(2l+1-2d)})$.

2.4.2 The best linear and approximately best linear predictors

Let $Z_1, \ldots, Z_n$ denote the past observations. The best linear predictor of Z_{n+k} for SEMIFAR was proposed by Beran and Ocker (1999):

$$\breve{Z}_{n+k} = \sum_{j=1}^{n} \beta_{k,j}^o Z_j, \tag{2.4.2}$$

where $\beta_k^o = (\beta_{k,1}^o, \ldots, \beta_{k,n}^o)^{\mathrm{T}}$ is as given in Theorem 1 of Beran and Ocker (1999), which minimizes the mean squared prediction error (MSE). Furthermore, $\breve{Z}_{n+k}$

satisfies

$$E[(Z_{n+k} - \breve{Z}_{n+k})Z_t] = 0, \quad t = 1, \ldots, n. \tag{2.4.3}$$

Eq. (2.4.3) implies that the prediction error of $\breve{Z}_{n+k}$ is orthogonal to any of the observations.

It is however not easy to use $\breve{Z}_{n+k}$ defined in Eq. (2.4.2), because β_k^o has to be solved repeatedly at each forecasting step. Note that in the current context n is very large. For simplicity, we propose to use an approximately best linear predictor based on the truncated part of the AR(∞) representation of Z_t. This idea is often employed to carry out forecasting based on an ARMA model, when the sample size is large (Brockwell and Davis, 2006, p. 184). Hence, an approximately best linear predictor based on $Z_1, \ldots, Z_n$, by means of the AR(∞) representation Eq. (2.3.5) of the FARIMA process is defined by

$$\hat{Z}^*_{n+k} = \sum_{j=1}^{k-1} \beta_j \hat{Z}^*_{n+k-j} + \sum_{j=k}^{n+k-1} \beta_j Z_{n+k-j}, \tag{2.4.4}$$

where $\hat{Z}^*_{n+k-j}$ are the previously predicted values. For the practical implementation, we propose to use the following linear predictor

$$\hat{Z}_{n+k} = \sum_{j=1}^{k-1} \hat{\beta}_j \hat{Z}_{n+k-j} + \sum_{j=k}^{n+k-1} \hat{\beta}_j \hat{Z}_{n+k-j}, \tag{2.4.5}$$

where $\hat{\beta}_j$ are the estimated coefficients in the AR(∞) form of Z_t, $\hat{Z}_{n+k-j}$ for $j = k, ..., n+k-1$, are the residuals obtained; and $\hat{Z}_{n+k-j}$, $j = 1, ..., k-1$, are the values predicted previously. The linear predictor Eq. (2.4.5) is what we propose to use in practice. To the best of our knowledge, the above-defined approximately best linear predictor has not previously been proposed in the literature in the presence of long memory. The relationship between $\hat{Z}_{n+k}$ and $\hat{Z}^*_{n+k}$ is given by

Lemma 2.4.1. *Under Assumptions A3 through A6, the two linear predictors $\hat{Z}_{n+k}$ and $\hat{Z}^*_{n+k}$ are asymptotically equivalent to each other.*

Proof of Lemma 2.4.1 is given in the appendix. Lemma 2.4.1 indicates that

the asymptotic properties of $\hat{Z}^*_{n+k}$, defined based on the unobservable quantities β_j and Z_t, are the same as those of $\hat{Z}_{n+k}$, defined using $\hat{\beta}_j$ and $\hat{Z}_t$.

Now, the best linear predictor given infinite past observations $Z_n, \ldots, Z_1, Z_0, Z_{-1}, \ldots$, is introduced. Similar to Eq. (5.5.3) of Brockwell and Davis (2006), this linear predictor based on the AR(∞) form of the FARIMA model in Eq. (2.3.5) is defined by

$$\tilde{Z}_{n+k} = \sum_{j=1}^{k-1} \beta_j \tilde{Z}_{n+k-j} + \sum_{j=k}^{\infty} \beta_j Z_{n+k-j}. \tag{2.4.6}$$

The properties of $\hat{Z}_{n+k}$ are stated in the following theorem.

Theorme 2.4.2. *Under the same assumptions of Lemma 2.4.1, the proposed linear predictor $\hat{Z}_{n+k}$ is an approximately best linear predictor in the following sense:*

i) $E[(\tilde{Z}_{n+k} - \hat{Z}_{n+k})^2] = o(1)$ *and*

ii) $E[(Z_{n+k} - \hat{Z}_{n+k})Z_t] = o(1)$, $t = 1, \ldots n$.

Proof of Theorem 2.4.2 is given in the appendix. Theorem 2.4.2 i) shows that $\hat{Z}_{n+k}$ converges to $\tilde{Z}_{n+k}$ in mean squared error. Theorem 2.4.2 ii) shows that the prediction error of $\hat{Z}_{n+k}$ is approximately orthogonal to all of the observations. Note that $\tilde{Z}_{n+k}$ is the best linear predictor based on infinite past observations. Hence, its MSE is no larger than that of $\breve{Z}_{n+k}$, because the σ-algebra generated by $Z_n, \ldots, Z_1, Z_0, Z_{-1}, \ldots$ includes that generated by $Z_n, \ldots, Z_1$ as a subset. Moreover, the MSE of $\hat{Z}_{n+k}$ is no smaller than that of $\breve{Z}_{n+k}$. Thus, Theorem 2.4.2 i) ensures that the MSE of all of the above mentioned linear predictors are asymptotically the same. This leads to the following corollary.

Corollary 1. *Under Assumptions A3 to A6, the linear predictor $\hat{Z}_t$ is asymptotically equivalent to the (exactly) best linear predictor $\breve{Z}_t$ proposed by Beran and Ocker (1999).*

2.4.3 Approximate forecasting intervals

Next, we will discuss the interval forecasting of an individual observation, the conditional mean and the total mean. Note that the point forecasting for ζ_{n+k} is

the same as that for Z_{n+k}, i.e. $\hat{\zeta}_{n+k} = \hat{Z}_{n+k}$. The variance of $Z_{n+k} - \hat{Z}_{n+k}$ and that of $\zeta_{n+k} - \hat{Z}_{n+k}$ can be easily obtained by adapting known results in the literature.

Theorme 2.4.3. *Under the same conditions of Theorem 2.4.2 we have*

i) $var(Z_{n+k} - \hat{Z}_{n+k}|Z_n, \ldots, Z_1) \approx V_{Z_{n+k}}$, *where* $V_{Z_{n+k}} = \sigma_\varepsilon^2 \sum_{i=0}^{k-1} \alpha_i^2$,

ii) $var(\zeta_{n+k} - \hat{Z}_{n+k}|Z_n, \ldots, Z_1) \approx V_{\zeta_{n+k}}$, *where* $V_{\zeta_{n+k}} = \sigma_\varepsilon^2 \sum_{i=1}^{k-1} \alpha_i^2$.

The proof of Theorem 2.4.3 is given in the Appendix. The result in Theorem 2.4.3 i) is well known. Note however that, in the current case $V_{Z_{n+k}}$ tends to $var(Z_t)$ very slowly. Moreover, as far as we know, the result in Theorem 2.4.3 ii) on the variance of the prediction error for the conditional mean is usually not discussed in the literature. This is however an interesting topic in financial econometrics. For example, it helps us to understand the accuracy of the forecasted volatility or the forecasted conditional mean duration.

The point forecast for an individual future observation is $\hat{Y}_{n+k} = \hat{\mu}(\tau_{n+k}) + \hat{Z}_{n+k}$. The length of the forecasting interval for Y_{n+k} is asymptotically the same as that for Z_{n+k}, because the error in $\hat{\mu}(\tau_{n+k})$ is asymptotically negligible compared to that in $\hat{Z}_{n+k}$. Assume that ε_t are i.i.d. $N(0, \sigma_\varepsilon^2)$. The approximate $100(1-\alpha)\%$-forecasting interval for Y_{n+k} is given by

$$Y_{n+k} \in \left(\hat{\mu}(\tau_{n+k}) + \hat{Z}_{n+k} - q_{\alpha/2}\sqrt{V_{Z_{n+k}}}, \hat{\mu}(\tau_{n+k}) + \hat{Z}_{n+k} + q_{\alpha/2}\sqrt{V_{Z_{n+k}}}\right) \quad (2.4.7)$$

and, for $k > 1$, the approximate $100(1-\alpha)\%$-forecasting interval of ζ_{n+k} is given by

$$\zeta_{n+k} \in \left(\hat{Z}_{n+k} - q_{\alpha/2}\sqrt{V_{\zeta_{n+k}}}, \hat{Z}_{n+k} + q_{\alpha/2}\sqrt{V_{\zeta_{n+k}}}\right), \quad (2.4.8)$$

where $q_{\alpha/2}$ is the upper $\alpha/2$-quantile of $N(0,1)$. Furthermore, let $m(\tau_t) = \mu(\tau_t) + \zeta_t$ and $g(\tau_t) = \exp[m(\tau_t)]$ be the total means in Y_t and X_t, respectively. We have $\hat{m}(\tau_{n+k}) = \hat{Y}_{n+k}$. But the prediction error for $m(\tau_{n+k})$ is approximately equal to that for ζ_{n+k}. Thus, the approximate $100(1-\alpha)\%$-forecasting interval for $m(\tau_{n+k})$,

$k > 1$, is given by

$$m(\tau_{n+k}) \in \left(\hat{\mu}(\tau_{n+k}) + \hat{Z}_{n+k} - q_{\alpha/2}\sqrt{V_{\zeta_{n+k}}}, \hat{\mu}(\tau_{n+k}) + \hat{Z}_{n+k} + q_{\alpha/2}\sqrt{V_{\zeta_{n+k}}}\right). \tag{2.4.9}$$

Note that the prediction errors in $\hat{\zeta}_{n+1}$ and $\hat{m}(\tau_{n+1})$ are both asymptotically negligible.

Our main purpose is to achieve suitable forecasts for X_{n+k}, λ_{n+k} and $g(\tau_{n+k})$. Taking the exponential transformation of $\hat{Z}_{n+k}$ and $\hat{m}(\tau_{n+k}) = \hat{Y}_{n+k}$, respectively, we have

$$\hat{\lambda}_{n+k} = \exp(\hat{Z}_{n+k}) = \prod_{j=1}^{n+k-1} \hat{Z}_{n+k-j}^{\hat{\beta}_j}, \tag{2.4.10}$$

$$\hat{X}_{n+k} = \hat{g}(\tau_{n+k}) = \exp[\hat{\mu}(\tau_{n+k}) + \hat{Z}_{n+k}] = \hat{\nu}(\tau_{n+k})\hat{\lambda}_{n+k}. \tag{2.4.11}$$

The approximate $100(1-\alpha)\%$-forecasting intervals for X_{n+k}, λ_{n+k} and $g(\tau_{n+k})$ can be obtained based on Eqs. (2.4.7) to (2.4.9), respectively, through exponential transformation.

2.5 Application

The forecasting of realized volatility plays an important role in option pricing and risk management. Hence, the proposals are applied to realized volatility and other related quantities aggregated from high-frequency data of different European firms. In what follows, empirical results for four selected examples, namely the (daily) trading numbers of Air France (AF-TrN), trading volumes of BMW (BMW-TrV), average durations of Peugeot (PSA-MD) and realized volatility of Metro (MEOG-RV) from Jan. 2, 2006 to Jun. 30, 2012, are reported. For each example, data-driven estimates were carried out using six sub-methods, namely those with $p_\mu = 1$ and 3, and $\alpha = \alpha_{opt}$, α_0 and α_{var}, denoted by M1, M2 and M3, respectively. This enables us to determine the effect of the bandwidth selection on the parameter estimation. The Epanechnikov kernel is used as the weighting function. To reduce the effect of the large variation in $\hat{\mu}^{(l)}(\tau)$ at boundary points on $\hat{h}$, $\hat{I}$ was calculated on $[\delta, 1-\delta]$ (See Beran and Feng, 2002c), where $\delta = 2.5\%$ and 5% are used for

$p_\mu = 1$ and 3, respectively.

Table 2.1 lists the selected bandwidth, the estimated long memory parameter and the selected AR order in all cases, together with the estimated short memory parameter, if applicable, with the 95%-confidence intervals for the corresponding parameters and results of the significance test of the trend being given as well. Note that, in a given case , $\hat{h}$ with $p_\mu = 3$ will be much larger than $\hat{h}$ with $p_\mu = 1$. For a given example and fixed p_μ, the bandwidths selected by α_{opt}, α_0 and α_{var} may clearly differ from each other. Usually, the bandwidths selected by α_0 and α_{var} are larger than that selected by α_{opt}. For instance, for PSA-MD with $p_\mu = 1$, the bandwidths selected by α_{opt}, α_0 and α_{var} are 0.112, 0.192 and 0.187, respectively. For AF-TrN with $p_\mu = 3$, they are 0.226, 0.227 and 0.280. The trend is significant in all cases, except in the example of MEOG-RV with $p_\mu = 1$. In this last case, the test is significant using the bandwidth selected by α_{opt}, but insignificant using bandwidths selected by α_0 and α_{var}, due to the enlarged bandwidths. The test results for this example might also be affected strongly by possible structural breaks (this will be discussed further a little later). In this chapter, we propose the use of the asymptotically optimal inflation factor α_{opt}, because the number of observations is very large. For choosing p_μ we found that forecasts with $p_\mu = 3$ using the linear extrapolation may be unreasonable at times, due to the instability of the estimate at the right endpoint. Hence, we propose the use of $p_\mu = 1$ for forecasting purposes. It is found that the use of $p_\mu = 3$ with a constant extrapolation will also work, but now the change of the trend cannot be reflected by the forecast.

For any given example, all six sub-methods select the same EFARIMA model. Here, we obtained an EFARIMA (0,d,0) for the AF-TrN and MEOG-RV examples. The relationship between $\hat{d}$ and $\hat{h}$ is obvious. For a given p_μ, the larger $\hat{h}$ is, the higher the $\hat{d}$ that will be obtained from the residuals. However, none of the examples clarify the differences between the values of $\hat{d}$ obtained by the six sub-methods with different bandwidths and different p_μ. The biggest difference occurred for PSA-MD, where $\hat{d}$ has a maximum of 0.294 and a minimum of 0.259. This difference becomes even smaller if we consider only the three sub-methods

Table 2.1: Results of ESEMIFAR models for the four data sets

Series		Bandwidth selection		$\hat{d}$ & 95%-CI	$\hat{p}$	$\hat{\phi}_1$ & 95%-CI	trend
		p_μ	$\hat{h}$				
M1	AF-TrN	1	0.207	0.409 [0.372, 0.447]	0		sign.
		3	0.226	0.397 [0.359, 0.434]	0		sign.
	BMW-TrV	1	0.146	0.299 [0.233, 0.366]	1	0.109 [0.024, 0.193]	sign.
		3	0.252	0.293 [0.226, 0.360]	1	0.114 [0.029, 0.199]	sign.
	PSA-MD	1	0.112	0.271 [0.202, 0.341]	1	0.158 [0.071, 0.246]	sign.
		3	0.228	0.259 [0.189, 0.329]	1	0.169 [0.081, 0.258]	sign.
	MEOG-RV	1	0.127	0.420 [0.382, 0.457]	0	—	sign.
		3	0.203	0.415 [0.377, 0.452]	0		sign.
M2	AF-TrN	1	0.203	0.409 [0.371, 0.447]	0	—	sign.
		3	0.227	0.396 [0.359, 0.434]	0	—	sign.
	BMW-TrV	1	0.196	0.312 [0.246, 0.378]	1	0.097 [0.013, 0.181]	sign.
		3	0.266	0.295 [0.228, 0.361]	1	0.112 [0.028, 0.197]	sign.
	PSA-MD	1	0.192	0.294 [0.226, 0.362]	1	0.137 [0.051, 0.223]	sign.
		3	0.230	0.259 [0.1889, 0.330]	1	0.169 [0.080, 0.257]	sign.
	MEOG-RV	1	0.149	0.425 [0.388, 0.463]	0	—	insign.
		3	0.215	0.417 [0.379, 0.455]	0	—	sign.
M3	AF-TrN	1	0.201	0.409 [0.371, 0.447]	0	—	sign.
		3	0.280	0.401 [0.363, 0.438]	0	—	sign.
	BMW-TrV	1	0.191	0.311 [0.245, 0.377]	1	0.098 [0.015, 0.182]	sign.
		3	0.254	0.293 [0.226, 0.360]	1	0.114 [0.029, 0.199]	sign.
	PSA-MD	1	0.187	0.293 [0.225, 0.362]	1	0.138 [0.052, 0.224]	sign.
		3	0.278	0.268 [0.198, 0.338]	1	0.161 [0.073, 0.248]	sign.
	MEOG-RV	1	0.151	0.426 [0.388, 0.463]	0	—	insign.
		3	0.246	0.420 [0.382, 0.458]	0	—	sign.

with the same p_μ. An interesting empirical finding is that (for a given example with a fixed α) the $\hat{d}$ obtained by $p_\mu = 3$ is always slightly lower than that obtained by $p_\mu = 1$. Theoretically, the large sample properties of the estimated parameters

with $p_\mu = 3$ are slightly better than those with $p_\mu = 1$. Furthermore, if $\hat{p} = 1$, $\hat{\phi}_1$ is also affected less by the selected bandwidths. However, now, due to the existence of short memory, $\hat{d}$ is much lower than in the case of $\hat{p} = 0$. Finally, if $\hat{p} = 1$ the $\hat{\phi}_1$ obtained in a given example with $p_\mu = 1$ is always slightly lower than that obtained with $p_\mu = 3$, a trade-off effect which ensures that the resulting theoretical ACF in a finite sample is not affected as much by the choice of p_μ.

Let $\hat{Z}_t = Y_t - \hat{\mu}(\tau_t)$ denote the residuals of the log-data. Histograms of the standardized values of $\hat{Z}_t$ and those of their exponential values are shown in Figure 2.1 for all examples. We see that the distribution of $\hat{Z}_t$ in all cases is nearly normal. This indicates that the normal assumption on ε_t is a suitable choice for analyzing these time series.

Figure 2.2(a) shows daily trading numbers of Air France together with the point and 95%-forecasting intervals for the next 50 days. We can see that the higher the scale function, the larger the variation in the observations, which reflects the fact that X_t has time varying variance $var(X_t) = \nu^2(\tau_t)var(X_t^*)$. To this end see also the other examples. This provides the evidence for the use of the log-transformation based on a multiplicative nonparametric regression model. Figure 2.2(b) displays the log-transformed data together with the estimated trend $\hat{\mu}(\tau_t)$ (solid line) and the corresponding forecasts of Y_{n+k}. In this example, the point forecasts are much lower than the estimated trend at the right end, but will tend to the average level in the near future. This reflects the well known fact that a long memory process may exhibit spurious local trends and indicates that long memory and the smooth scale change should be investigated together. The estimated conditional means of the log-data together with the corresponding point and interval forecasts for ζ_{n+k} are given in Figure 2.2(c). The estimated conditional means look quite stationary. The difference between the forecasting interval of Y_{n+k} in Figure 2.2(b) and that of ζ_{n+k} in Figure 2.2(c) is that the former is affected by ε_{n+k}, but the latter not. The estimated total means in the original data $\hat{g}(\tau_t)$ together with the point forecasts $\hat{g}(\tau_{n+k})$ and their forecasting intervals are displayed in Figure 2.2(d), which reflect the total dynamics of the daily trading number of Air France caused by past information and slowly changing macroeco-

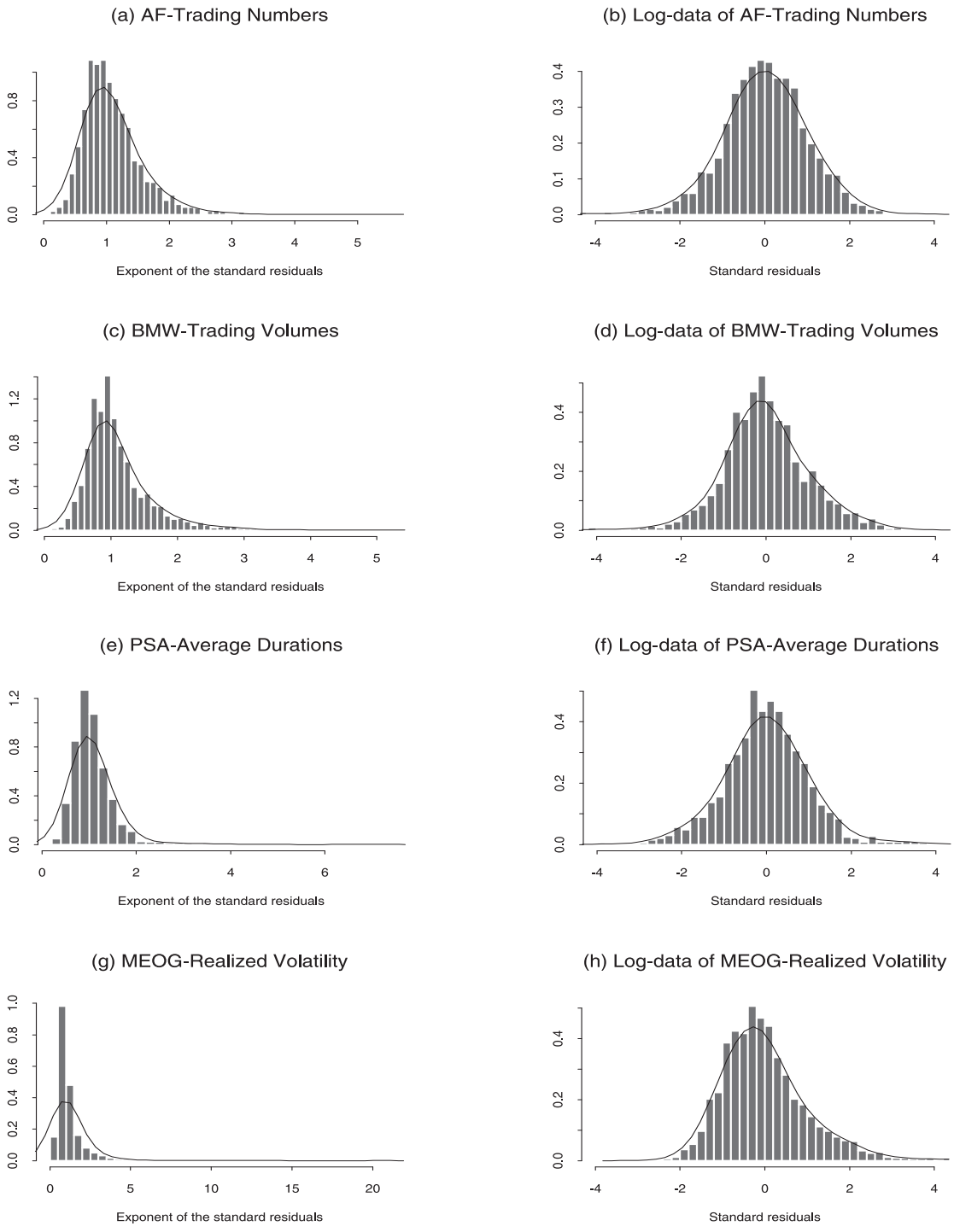

Figure 2.1: Histograms of the standardized residuals of the SEMIFAR model for the log-data and their exponential transformation for all examples.

nomic environment. From Figure 2.2(d) we can see that trading numbers of Air France and their volatility increased strongly in the last years and will possibly increase further in the future. Results on $\hat{\lambda}_t$ and $\hat{m}(\tau_t)$ are omitted.

Similar results for daily trading volumes of BMW and daily average durations of Peugeot are displayed in Figures 2.3 and 2.4. From Figure 2.3(b) we can see that the current BMW trading volumes stay at a relatively low but stable level. The results in Figure 2.5 indicate that the transaction durations of Peugeot reduced clearly in the last years due to the introduction of electronic trading platforms. Figures 2.3(b) and 2.4(b) show in particular that after the log-transformation those data can be well modeled by an additive nonparametric regression and the residual series in Figures 2.3(c) and 2.4(c) are quite stationary.

Finally, empirical results of the Semi-FI-Log-ACD applied to realized volatility of Metro are displayed in Figure 2.5. Comparing Figure 2.5(a) with Figure 2.5(b) we see that analysis of the log-transformed realized volatility is a natural way, as it is usually proposed in the literature. Although the Semi-FI-Log-ACD works in this case, there seems to be a problem. That is the realized volatility during the global financial crisis in 2008 and 2009 and the European debt crisis in 2011 was very high with possible structural breaks. The fact that realized volatility may exhibit long memory and structural breaks at the same time is e.g. reported by Choi et al. (2010), who proposed to improve the estimation of the long memory parameter in realized volatility after removing the detected structural breaks. From Figure 2.5(b) we can see that in addition to long memory and possible structural breaks, this series may also exhibit significant trend, in particular in the subperiod between the two crises. A scrutinizing empirical study shows that the quality of both of the estimated long-memory parameter and the fitted trend can be improved clearly, if possible structural breaks are taken into account. It is better to find out all structural breaks in the scale function of this realized volatility series first using a suitable nonparametric detecting procedure. An ESEMIFAR model can then be fitted to each of the subperiods determined by the detected structural breaks separately. Detailed study on such an approach is beyond the aim of this chapter and will be investigated elsewhere.

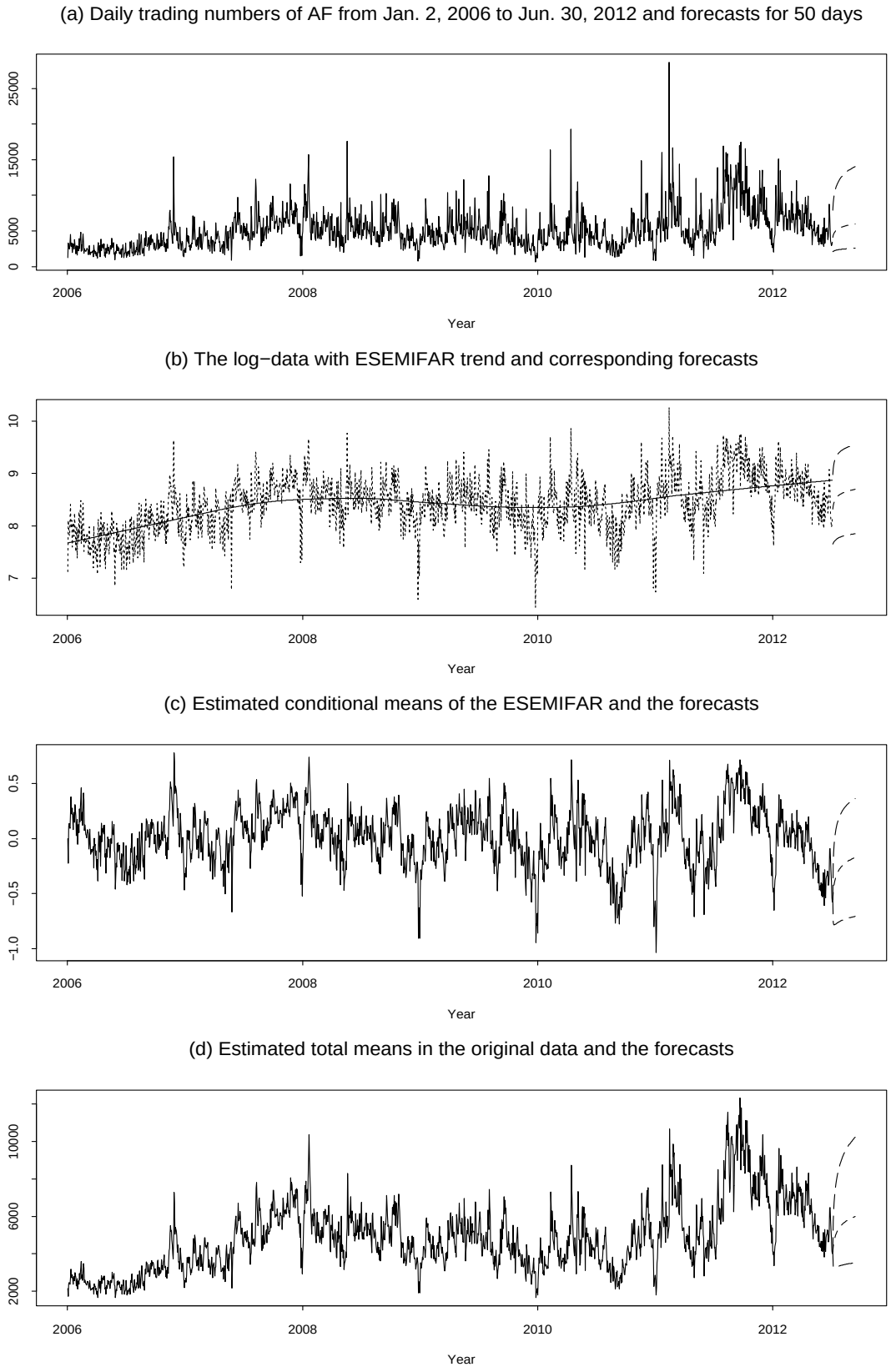

Figure 2.2: Estimation and forecasting results for daily trading numbers of AF from Jan. 02, 2006 to Jun. 30, 2012, obtained by the Semi-FI-Log-ACD model.

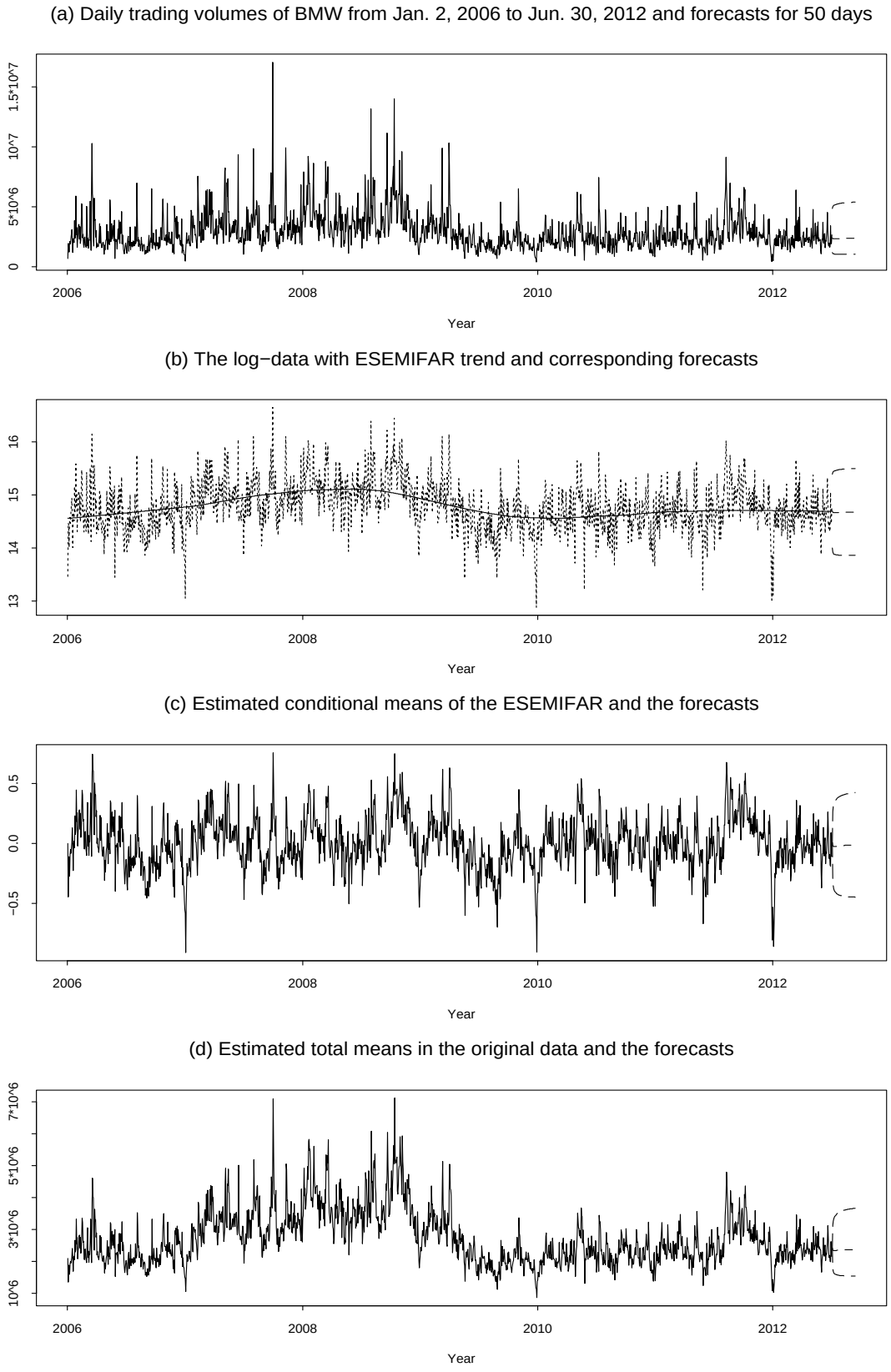

Figure 2.3: Similar results as given in Figure 2.2 but for daily trading volumes of BMW.

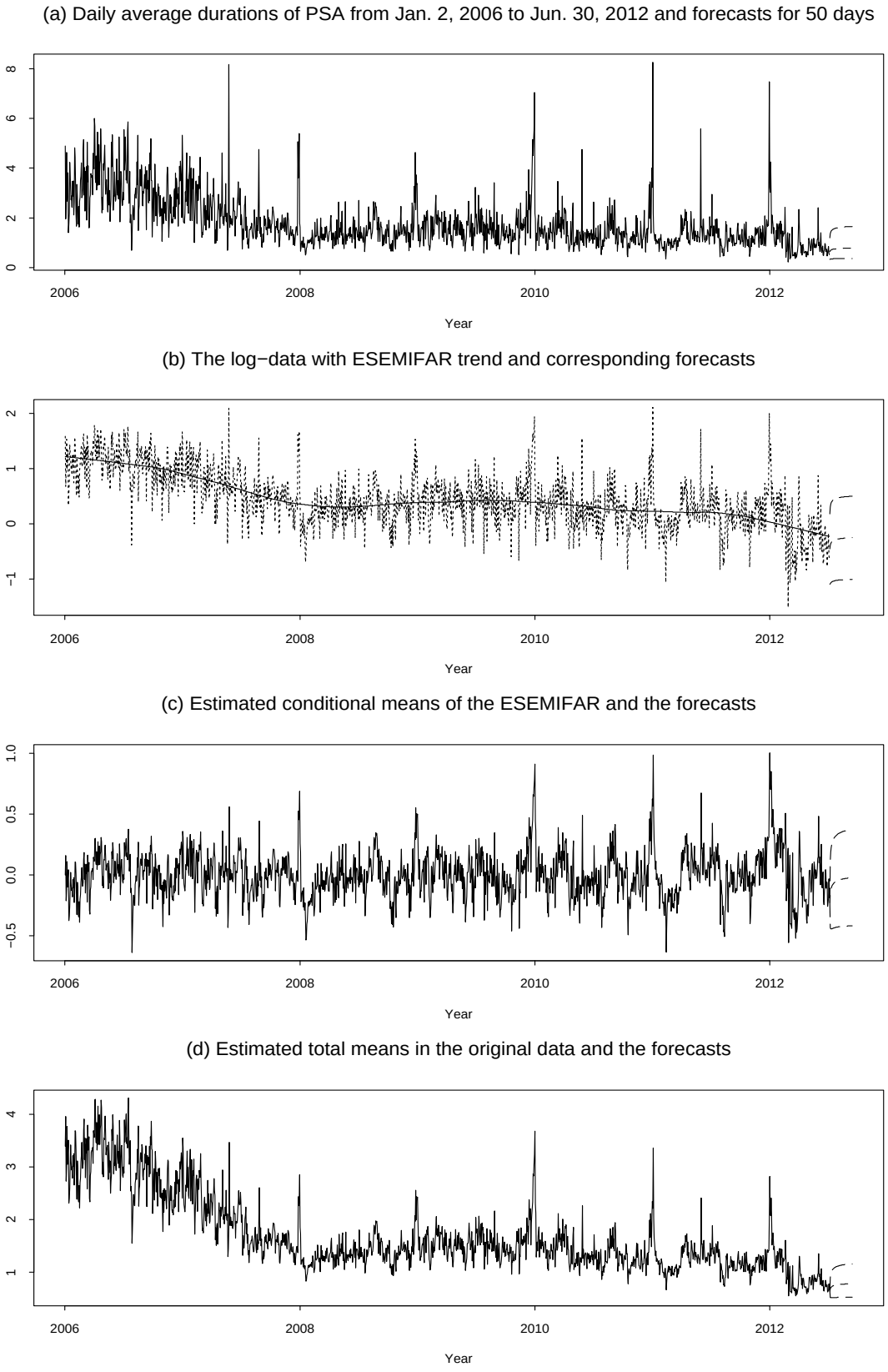

Figure 2.4: Similar results as given in Figure 2.2 but for daily average durations of Peugeot.

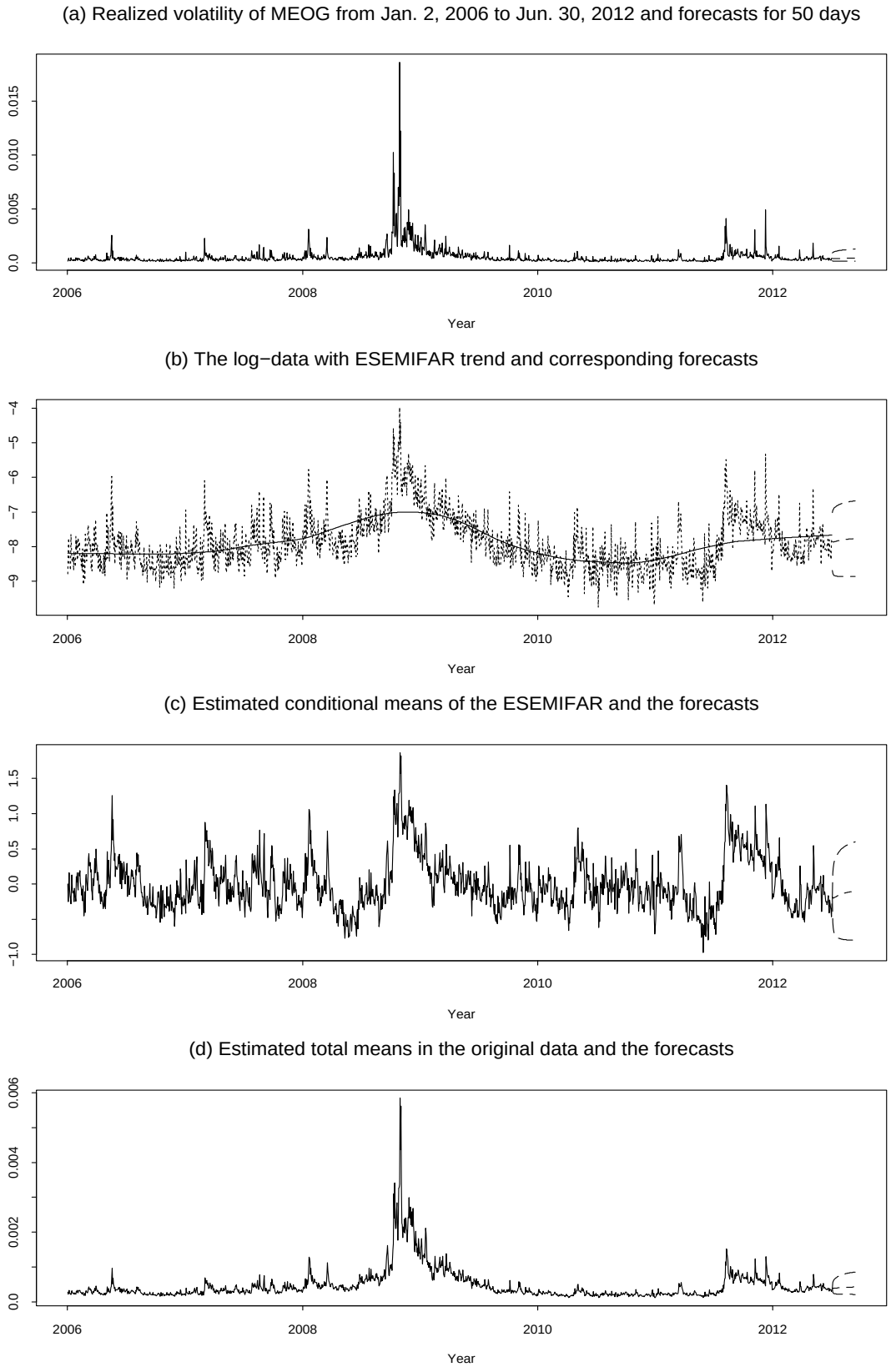

Figure 2.5: Similar results as given in Figure 2.2 but for realized volatility of Metro.

2.6 Final remarks

Some important results and a data-driven algorithm for the Semi-FI-Log-ACD model have been obtained. The short- and middle-term forecasting of a nonnegative process with long memory and a nonparametric scale function based on this model have also been studied. An approximately best linear predictor has been proposed, and an application to a number of financial time series aggregated from high-frequency data has shown that the proposals work very well in practice. Note in particular that the simultaneous estimation of the nonparametric trend and the long memory error process improves the forecast quality. The reasons for this are as follows. On the one hand, if potential long memory in the conditional mean of a process is not considered, the bandwidth selected will be much smaller than it should be, and the formula for calculating the asymptotic variance will also be wrong. This may lead to the estimation of a significant trend, even if the underlying process is in fact stationary. On the other hand, if an existing nonparametric scale function is not considered, it will be misinterpreted as very strong long memory. Finally, possible structural breaks also have a clear effect on the estimated model; however, this is not considered in the current chapter.

Appendix to Chapter 2

Proof of Lemma 2.4.1. It is well known that, under assumptions A3 through A6, the local linear estimator $\hat{\mu}(\tau)$ with fractional times series error achieves the optimal convergence rate of the order $O(n^{-2(1-2d)/(5-2d)})$ (Feng and Beran, 2013). This results in the fact that the difference between $\hat{Z}_t$ and Z_t is also of the order $O(n^{-2(1-2d)/(5-2d)})$. Moreover, when $n \to \infty$ and $d > 0$, the effect of the estimated trend function on the estimation of the unknown parameter vector θ is negligible (Beran and Feng, 2002a), and $\hat{\theta}$ is now still $\sqrt{n}$-consistent. Using Taylor expansion it can be shown that $\hat{\beta}_j - \beta_j = \beta_j O_p(n^{-1/2})$. Detailed discussion on this point is omitted to save space.

In what follows, we will only show the result of Lemma 2.4.1 for $k = 1$ in detail. Note that $\hat{Z}^*_{n+1} = \sum_{j=1}^{n} \beta_j Z_{n+1-j} = O_p(1)$ and $\sum_{j=1}^{n} |\beta_j| < \sum_{j=1}^{\infty} |\beta_j| < \infty$ for $d > 0$. We have

$$\begin{aligned}
\hat{Z}_{n+1} - \hat{Z}^*_{n+1} &= \sum_{j=1}^{n} \hat{\beta}_j \hat{Z}_{n+1-j} - \sum_{j=1}^{n} \beta_j Z_{n+1-j} \\
&= \sum_{j=1}^{n} \hat{\beta}_j \hat{Z}_{n+1-j} - \sum_{j=1}^{n} \hat{\beta}_j Z_{n+1-j} - \sum_{j=1}^{n} \beta_j Z_{n+1-j} + \sum_{j=1}^{n} \hat{\beta}_j Z_{n+1-j} \\
&= \sum_{j=1}^{n} \hat{\beta}_j (\hat{Z}_{n+1-j} - Z_{n+1-j}) - \sum_{j=1}^{n} (\beta_j - \hat{\beta}_j) Z_{n+1-j} \\
&= \sum_{j=1}^{n} \hat{\beta}_j O_p(n^{-2(2d-1)/(5-2d)}) - \sum_{j=1}^{n} \beta_j Z_{n+1-j} O_p(n^{-1/2}) \\
&\approx \sum_{j=1}^{n} \hat{\beta}_j O_p(n^{-2(2d-1)/(5-2d)}) \qquad \text{(A.2.1)} \\
&\leq O_p(n^{2(2d-1)/(5-2d)}) \sum_{j=1}^{n} |\hat{\beta}_j| = o_p(1).
\end{aligned}$$

Similarly, this result can be proved for $k > 1$. Lemma 2.4.1 is proved. $\diamond$

Proof of Theorem 2.4.2. Following Lemma 2.4.1, the results of Theorem 2.4.2 will be proved by replacing $\hat{Z}_{n+k}$ with $\hat{Z}^*_{n+k}$. Under the same conditions of Lemma 2.4.1 we have

i) For $k = 1$:

$$\begin{aligned}
E[(\tilde{Z}_{n+1} - \hat{Z}^*_{n+1})^2] &= E\left[\left(\sum_{j=1}^{\infty} \beta_j Z_{n+1-j} - \sum_{j=1}^{n} \beta_j Z_{n+1-j}\right)^2\right] \\
&= E\left[\left(\sum_{j=n+1}^{\infty} \beta_j Z_{n+1-j}\right)^2\right] \\
&= \sum_{i=n+1}^{\infty} \beta_i \sum_{j=n+1}^{\infty} \beta_j E[Z_{n+1-i} Z_{n+1-j}] \\
&= \sum_{i=n+1}^{\infty} \beta_i \sum_{j=n+1}^{\infty} \beta_j \gamma(i-j) \qquad \text{(A.2.2)} \\
&\approx \sum_{i=n+1}^{\infty} C_\beta i^{-d-1} \sum_{j=n+1}^{\infty} C_\beta j^{-d-1} \gamma(i-j) \\
&\leq \gamma(0) \sum_{i=n+1}^{\infty} |C_\beta| i^{-d-1} \sum_{j=n+1}^{\infty} |C_\beta| j^{-d-1} \\
&= \gamma(0) O[(n+1)^{-2d}] = o(1).
\end{aligned}$$

More details of the proof above will be clarified by the remark given later.

Now, let $k > 1$ and assume that the results are proved for $i = 1, \ldots, k-1$. We have

$$\begin{aligned}
E[(\tilde{Z}_{n+k} - \hat{Z}^*_{n+k})^2] &= E\left[\left(\sum_{j=1}^{k-1} \beta_j \tilde{Z}_{n+k-j} + \sum_{j=k}^{\infty} \beta_j Z_{n+k-j} - \sum_{j=1}^{k-1} \beta_j \hat{Z}^*_{n+k-j} - \sum_{j=k}^{n+k-1} \beta_j Z_{n+k-j}\right)^2\right] \\
&= E\left[\left(\left(\sum_{j=1}^{k-1} \beta_j \tilde{Z}_{n+k-j} - \sum_{j=1}^{k-1} \beta_j \hat{Z}^*_{n+k-j}\right) + \sum_{j=n+k}^{\infty} \beta_j Z_{n+k-j}\right)^2\right] \\
&= E\left[T_1^2 + 2T_1T_2 + T_2^2\right] \\
&= E[T_1^2] + 2E[T_1T_2] + 2E[T_2^2],
\end{aligned}$$

where $T_1 = \sum_{j=1}^{k-1} \beta_j (\tilde{Z}_{n+k-j} - \hat{Z}^*_{n+k-j})$ and $T_2 = \sum_{j=n+k}^{\infty} \beta_j Z_{n+k-j}$.

Since all of the terms in T_1 are of the order $o_p(1)$, T_1 is hence an $o_p(1)$ term. Similarly as for $k = 1$, it can be shown that T_2 is also of the order $o_p(1)$. This leads to the conclusion that $E[(\tilde{Z}_{n+k} - \hat{Z}_{n+k})^2] = o(1)$ holds for $k > 1$.

ii) For $k = 1$ and any $t = 1, ..., n$:

$$\begin{aligned}
E[(Z_{n+1} - \hat{Z}^*_{n+1})Z_t] &= E\left[\left(\sum_{i=1}^{\infty}\beta_i Z_{n+1-i} + \varepsilon_{n+1} - \sum_{i=1}^{n}\beta_i Z_{n+1-i}\right)\left(\sum_{j=1}^{\infty}\beta_j Z_{t-j} + \varepsilon_t\right)\right] \\
&= E\left[\left(\sum_{i=n+1}^{\infty}\beta_i Z_{n+1-i} + \varepsilon_{n+1}\right)\left(\sum_{j=1}^{\infty}\beta_j Z_{t-j} + \varepsilon_t\right)\right] \\
&= E\left[\sum_{j=1}^{\infty}\beta_j Z_{t-j}\sum_{i=0}^{\infty}\beta_{n+1+i}Z_{-i} + \varepsilon_{n+1}\sum_{j=1}^{\infty}\beta_j Z_{t-j}\right. \\
&\quad \left. + \varepsilon_t\left(\sum_{i=0}^{\infty}\beta_{n+1+i}Z_{-i} + \varepsilon_{n+1}\right)\right] \\
&= E\left[\sum_{j=1}^{\infty}\beta_j Z_{t-j}\sum_{i=0}^{\infty}\beta_{n+1+i}Z_{-i}\right] + E\left[\varepsilon_{n+1}\sum_{j=1}^{\infty}\beta_j Z_{t-j}\right] \\
&\quad + E\left[\varepsilon_t\left(\sum_{i=0}^{\infty}\beta_{n+1+i}Z_{-i} + \varepsilon_{n+1}\right)\right].
\end{aligned}$$

Since $E\left[\varepsilon_{n+1}\sum_{j=1}^{\infty}\beta_j Z_{t-j}\right] = 0$ and $E\left[\varepsilon_t\left(\sum_{i=0}^{\infty}\beta_{n+1+i}Z_{-i} + \varepsilon_{n+1}\right)\right] = 0$,

$$\begin{aligned}
E[(Z_{n+1} - \hat{Z}^*_{n+1})Z_t] &= E\left[\sum_{j=1}^{\infty}\beta_j Z_{t-j}\sum_{i=0}^{\infty}\beta_{n+1+i}Z_{-i}\right] \qquad \text{(A.2.3)} \\
&= \sum_{i=0}^{\infty}\beta_{n+1+i}\sum_{j=1}^{\infty}\beta_j E[Z_{-i}Z_{t-j}] \\
&\leq \sum_{i=0}^{\infty}|\beta_{n+1+i}|\sum_{j=1}^{\infty}|\beta_j||\gamma(t+i-j)| \\
&\leq \gamma(0)\sum_{i=0}^{\infty}|C_\beta|(n+1+i)^{-d-1} = O[(n+1)^{-d}] = o(1).
\end{aligned}$$

Now, let $k > 1$ and assume that the results are proved for $i = 1, \ldots, k-1$, we have

$$\begin{aligned}
E[(Z_{n+k} - \hat{Z}^*_{n+k})Z_t] &= E\left\{\left[\sum_{i=1}^{k-1}\beta_i(Z_{n+k-i} - \hat{Z}^*_{n+k-i}) + \sum_{i=n+k}^{\infty}\beta_i Z_{n+k-i} + \varepsilon_{n+k}\right]Z_t\right\} \\
&= E[T_3 + T_4 + T_5], \qquad \text{(A.2.4)}
\end{aligned}$$

where $T_3 = Z_t \sum_{i=1}^{k-1} \beta_i (Z_{n+k-i} - \hat{Z}^*_{n+k-i})$, $T_4 = Z_t \sum_{i=n+k}^{\infty} \beta_i Z_{n+k-i}$ and $T_5 = \varepsilon_{n+k} Z_t$ with $E(T_5) = 0$. It is clear that $E(T_3) = o(1)$, because the results hold for $i = 1, \ldots, k-1$. The fact that $E(T_4) = o(1)$ can be proved similarly as for $k = 1$. Insert these results into (A.2.4) we obtain

$$E[(Z_{n+k} - \hat{Z}^*_{n+k}) Z_t] = o(1), t = 1, ..., n, \tag{A.2.5}$$

for any $k > 1$. Theorem 2.4.2 is proved. $\diamond$

Remark 1. *Some techniques used in the proof only apply to $d > 0$, while for $d < 0$ other approaches should be used. It is very common that some conclusions hold only for long memory errors but not for antipersistent errors. For instance, for $d > 0$ we have $\sum_{i=1}^{\infty} \beta_i = 1$. For $d < 0$, β_i are however not summable. Furthermore, the approximate formula of $\gamma(k)$ does not apply to $\gamma(i-j)$ in the fourth line of* (A.2.2)*. The reason is that although both i and j tend to infinity, their difference may be very small. Hence, here the fact that $|\gamma(k)| \leq \gamma(0)$ is simply employed. Detailed analysis of the second sum there may lead to more accurate result. This is however omitted to simplify the proof.*

Proof of Theorem 2.4.3. For a causal stationary and invertible ARMA model, the predictor $\tilde{Z}_{n+k}$ defined in (2.4.6) can be represented as a MA(∞) form (see e.g. Theorem 5.5.1 of Brockwell and Davis, 2006)

$$\tilde{Z}_{n+k} = \sum_{i=k}^{\infty} \alpha_i \varepsilon_{n+k-i}. \tag{A.2.6}$$

It can be shown that this fact also holds, if Z_t is a causal stationary and invertible FARIMA model considered in this chapter. The difference between Z_{n+k} and $\tilde{Z}_{n+k}$ is:

$$Z_{n+k} - \tilde{Z}_{n+k} = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{n+k-i} - \sum_{i=k}^{\infty} \alpha_i \varepsilon_{n+k-i} = \sum_{i=0}^{k-1} \alpha_i \varepsilon_{n+k-1}. \tag{A.2.7}$$

The variance of $Z_{n+k} - \tilde{Z}_{n+k}$ is therefore

$$var(Z_{n+k} - \tilde{Z}_{n+k}) = \sigma_\varepsilon^2 \sum_{i=0}^{k-1} \alpha_i. \tag{A.2.8}$$

Note that the point forecast for the conditional mean, $\tilde{\zeta}_{n+k}$, is the same as $\tilde{Z}_{n+k}$. The difference between ζ_{n+k} and $\tilde{Z}_{n+k}$ is given by

$$\zeta_{n+k} - \tilde{Z}_{n+k} = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{n+k-i} - \sum_{i=k}^{\infty} \alpha_i \varepsilon_{n+k-i} = \sum_{i=1}^{k-1} \alpha_i \varepsilon_{n+k-i} \tag{A.2.9}$$

with the variance

$$var(\zeta_{n+k} - \tilde{Z}_{n+k}) = \sigma_\varepsilon^2 \sum_{i=1}^{k-1} \alpha_i^2. \tag{A.2.10}$$

In Theorem 2.4.2 it is shown that $\hat{Z}_{n+k} \approx \tilde{Z}_{n+k}$. Thus, $Z_{n+k} - \hat{Z}_{n+k} \approx Z_{n+k} - \tilde{Z}_{n+k}$ and $\zeta_{n+k} - \hat{Z}_{n+k} \approx \zeta_{n+k} - \tilde{Z}_{n+k}$. Consequently, $var(Z_{n+k} - \hat{Z}_{n+k}) \approx var(Z_{n+k} - \tilde{Z}_{n+k})$ and $var(\zeta_{n+k} - \hat{Z}_{n+k}) \approx var(\zeta_{n+k} - \tilde{Z}_{n+k})$. Theorem 2.4.3 is proved. $\diamond$

Chapter 3

An iterative plug-in algorithm for realized kernels

This chapter is based on joint work with Yuanhua Feng and published with slight differences as Working Paper (2015-01) in the Working Paper Series of Center for International Economics at Paderborn University.

3.1 Introduction

Estimation of the daily integrated volatility (IV) is an important topic in risk management, portfolio allocation and option pricing. Realized volatility (RV) introduced by Andersen et al. (2001a, b) is a model-free estimator of this quantity based on high-frequency financial data. The most simple definition of the RV, called RV_0, is the sum of the squared intraday returns. It is however found that high-frequency data often exhibit microstructure (MS) noise (Hasen and Lunde, 2006a). Strong evidence for the existence of MS noise is illustrated in Figure 3.1 in the next section using numerical examples. Now, RV_0 is an inconsistent estimator of the IV (Zhang et al., 2005, Bandi and Russel, 2008). Different bias corrected estimators of the IV are introduced into the literature. For instance, Zhou (1996) proposed an improved estimator, called RV_Z, by including the cross-products between two consequent observations, which is unbiased under i.i.d. MS

noise. Bandi and Russel (2006, 2008, 2011) and Oomen (2006) investigated the use of sparse equidistant high-frequency data and studied the choice of the optimal frequency to make a trade-off between the variance and bias of the proposed estimators. Zhang et al. (2005) and Aït-Sahalia et al. (2011) proposed the use of a realized volatility estimator with two time scales to solve the bias problem caused by MS noise. Hansen and Lunde (2006a) and Oomen (2005) proposed a simple kernel based estimator of the IV. Furthermore, Hansen et al. (2008) investigated correction of MS bias using moving average-based estimators.

Recently, Barndorff-Nielsen et al. (2008, 2009, 2011b) introduced the realized kernels (RK), which are consistent estimators of the IV under given conditions. A crucial problem by applying realized kernels is the selection of the bandwidth, because an RK only works well, if the bandwidth is selected properly. This is illustrated in Figure 3.2 in the next section through the above mentioned numerical examples. Barndorff-Nielsen et al. (2009) proposed to select the bandwidth by plugging suitable estimates of two unknowns into a simplified (but biased) formula of the asymptotically optimal bandwidth of the RK. However, their proposal is very complex and not fully data-driven. And the selected bandwidth by this algorithm does not converge to the targeted bandwidth.

In this chapter an iterative plug-in (IPI) bandwidth selector for realized kernels is developed by adapting the idea of Gasser et al. (1991) to the current context. So far as we know, this is the first IPI algorithm for RK. For simplicity, we also adopt the biased targeted bandwidth proposed by Barndorff-Nielsen et al. (2009). The difference between RV_0 and the RK resulted in each iteration is used to estimate the variance of the MS noise. And RV_{Z} is used as an initial value of the RK so that the procedure is fully data-driven. It is shown that the proposed bandwidth selector is consistent in the sense that the relative error with respect to the targeted bandwidth tends to zero, as $n \to \infty$. Furthermore, the proposed bandwidth selection rule is very simple and the algorithm runs very fast, because only a few iterations are required. It is hence suitable to be applied to obtain data-driven RK in a long observation period. Theoretically, both of the resulted RK and the selected bandwidth become consistent from the third iteration, while their

rate of convergence can still be improved in the fourth iteration. Thereafter, the resulted RK achieves its optimal rate of convergence of the order $O(n^{-1/5})$, which is also shared by the relative error in the selected bandwidth. The nice practical performance of the proposal is illustrated by application to data of a few German and French firms. These results show that in most of the cases the procedure converges within four iterations. And the distribution of the selected bandwidths is nearly normal. Empirical analysis showed that the resulted RK performs better than RV_0 and RV_Z. It seems that both of the bias and the standard deviation of RV_0 and RV_Z are clearly reduced by the data-driven RK. But this fact still needs to be confirmed through simulation. The performance of the proposed bandwidth selector on a few so-called 'challenging days' is discussed in detail.

Further analysis of the obtained results is of great interest. Andersen et al. (2001a, 2001b, 2011) and Deo et al. (2006) find that the logarithmic RV may exhibit long memory. Choi et al. (2010) found that the observed long memory may be spuriously generated e.g. by a nonparametric trend or structural breaks. Hence, long memory, nonparametric trends and possible structural breaks should be studied simultaneously. We propose to analyze realized kernels use a piecewise version of the ESEMIFAR (exponential semiparametric fractional autoregressive, Beran et al., 2015). It is found that realized kernels exhibit long memory and a significant nonparametric trend at the same time. Estimation results for the two sub-periods before and after the 2008 financial crisis are clearly different.

The chapter is organized as follows. Some necessary known results are summarized in section 3.2. The data-driven bandwidth selector is proposed and studied in section 3.3. Application to real data is reported in section 3.4. In section 3.5 modeling of realized kernels using the ESEMIFAR model is discussed. Final remarks in section 3.6 conclude the chapter.

3.2 Realized volatility and realized kernels

3.2.1 Effect of MS noise on realized volatility

Let $p^*(\tau)$ denote the logarithmic efficient asset price on a trading day, where $0 \leq \tau \leq T$, and 0 and T denote the opening and closing time. Assume that $p^*(\tau)$ are determined by the stochastic differential equation

$$dp^*(\tau) = \sigma(\tau)dW(\tau), \tag{3.2.1}$$

where $W(\tau)$ is a standard Brownian motion and $\sigma(\tau)$ is the spot volatility process. Furthermore, it is assumed that the $\sigma(\tau)$ and $W(\tau)$ processes are independent of each other. Estimation of the daily integrated volatility

$$\text{IV} = \int_0^T \sigma^2(\tau)d\tau \tag{3.2.2}$$

is of great interest. Realized volatility is introduced as a model free estimator of the IV based on high-frequency financial data. Let p_i be the logarithmic asset prices observed at time points $0 = \tau_0 < \tau_1 < \ldots < \tau_n < \tau_{n+1} = T$, where n is the (random) number of observations happened on that day. It is assumed that $\tau_i - \tau_{i-1} = O_p(n^{-1})$. The intraday returns are given by $r_i = p_i - p_{i-1}$. The most simple definition of the realized volatility is

$$\text{RV}_0 = \sum r_i^2, \tag{3.2.3}$$

which is a consistent estimator of the IV, if there is no MS noise such that $p_i = p_i^*$, where p_i^* stands for $p^*(\tau_i)$. In the presence of MS noise we have however

$$p_i = p_i^* + u_i, \tag{3.2.4}$$

where u_i represents a stationary noise process with mean zero and $\text{var}(u_i) = \omega^2$. It is assumed that u_i is independent of p_i^*. In this chapter we will focus on the case with i.i.d. u_i. Let $r_i^* = p_i^* - p_{i-1}^*$ be the efficient returns. The corresponding

noise contaminated observed returns are given by

$$r_i = p_i - p_{i-1} = r_i^* + e_i, \tag{3.2.5}$$

where $e_i = u_i - u_{i-1}$ is the noise in r_i. The observed returns are correlated to each other, while r_i^* are uncorrelated. Under the i.i.d. assumption on u_i, e_i follow an MA(1) model. It can be shown that, the ACF of r_i at lag 1 is $\rho_r(1) = -\omega^2/(2\omega^2 + \sigma_i^2) \to -0.5$, as $n \to \infty$, where $\sigma_i^2 = \text{var}(r_i^*)$. If ω^2/σ_i^2 is large, RV_0 is clearly overestimated.

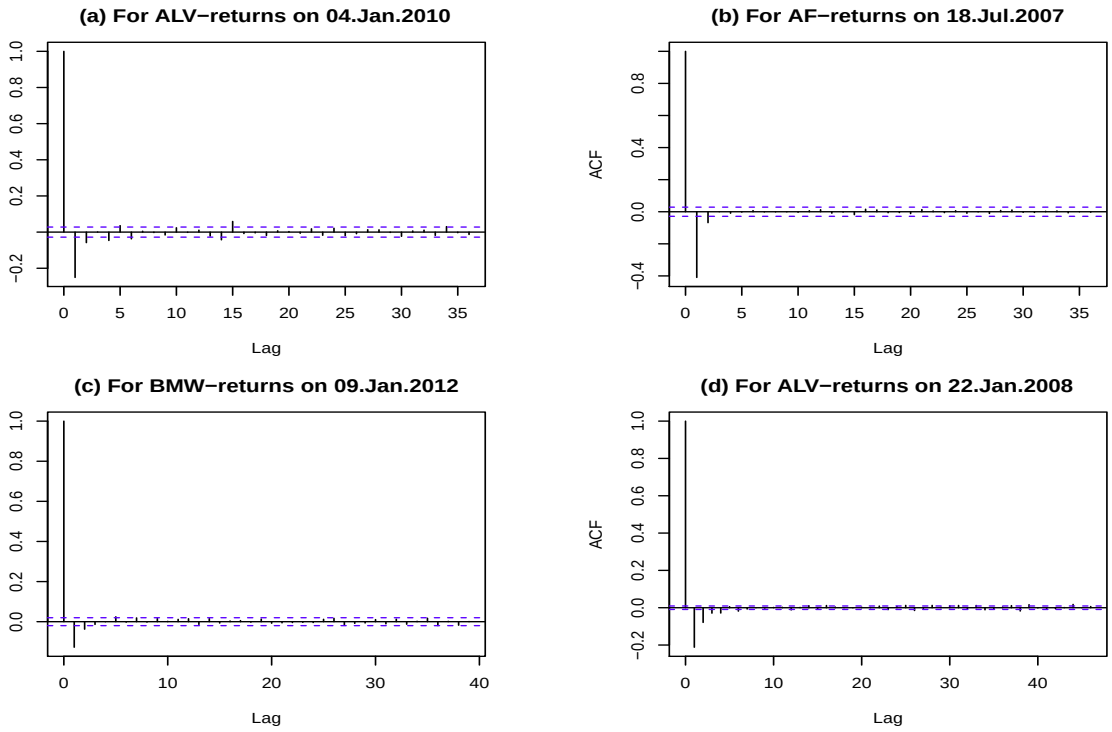

Figure 3.1: Examples of ACFs of high-frequency returns on four selected days

Empirical evidence of MS noise can be found by displaying the ACF of high-frequency returns. Figure 3.1 shows the correlograms of high-frequency returns on four selected trading days, one from each of the following German and French companies, Air France (AF), Allianz (ALV), BMW and Peugeot (PSA), respectively. From Figure 3.1 we see that $\rho_r(1)$ is always significantly negative, a clear evidence for the existence of MS noise. The independence assumption on the noise

is about true in the first three cases. Furthermore, we see the noise on the second selected day is very strong with $\hat{\rho}(1) < -0.4$ as can be seen from Figure 3.1(b). Figures 3.1(a) and (c) show that the noise on the first and third selected days is at a middle and a relatively low level, respectively. The fourth example in Figure 3.1(d) is chosen to show that strong and dependent noise could also happen. For this example not only $\hat{\rho}_r(1)$ but also those at lags 2 and 3 are significantly non-zero with $\hat{\rho}_r(2) > 0$. But the sum of $\hat{\rho}_r(1)$ to $\hat{\rho}_r(3)$ is clearly negative.

In the presence of MS noise, RV_0 can be rewritten as

$$\mathrm{RV}_0 = \sum_{i=1}^{n} (r_i^*)^2 + 2\sum_{i=1}^{n} r_i^* e_i + \sum_{i=1}^{n} e_i^2. \tag{3.2.6}$$

It is well known that the bias of RV_0 is $B(\mathrm{RV}_0) = 2n\omega^2$ and the asymptotic variance of RV_0 is $var(\mathrm{RV}_0) \approx 4nE(u_i^4)$, as $n \to \infty$. Different approaches are introduced into the literature to improve the performance of RV_0. Under the i.i.d. assumption on u_i, Zhou (1996) proposed to correct the bias in RV_0 by introducing the cross-products of lag 1 into RV_0. In this chapter his proposal is slightly modified as follows:

$$\mathrm{RV}_\mathrm{Z} = \sum_{i=2}^{n-1} (r_i^2 + r_i r_{i+1} + r_i r_{i-1}). \tag{3.2.7}$$

Under independent MS noise RV_Z is unbiased and its variance is approximately $8n\omega^4$, as $n \to \infty$. That is this estimator is still inconsistent.

3.2.2 Realized kernels

To overcome the above mentioned problems of well known estimators of the IV, Barndorff-Nielsen et al. (2008, 2009, 2011b) introduced the realized kernels, which are consistent estimators of the IV in the presence of MS noise under regularity conditions. A RK is defined by

$$\mathrm{RK} = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \quad \gamma_h = \sum_{j=|h|+1}^{n} r_j r_{j-|h|}, \tag{3.2.8}$$

where $k(u)$ is a kernel weight function, H is the bandwidth and γ_h is the h-th realized autocovariance. To ensure the non-negativity of the RK, it is assumed that $k(u)$ satisfies the Condition **K** in Barndorff-Nielsen et al. (2011b). This implies in particular that the kernel is with a non-flat top such that $k''(0)$, the second derivative of $k(u)$ at the origin, is non-zero. A variety of kernel functions in this class may be found in Table 1 of Barndorff-Nielsen et al. (2011b). The authors indicated that the use of the Parzen kernel is more preferable. For $u \geq 0$, the Parzen kernel is defined by

$$k(u) = \begin{cases} 1 - 6u^2 + 6u^3, & 0 \leq u \leq 1/2, \\ 2(1-u)^3, & 1/2 < u \leq 1, \\ 0 & u > 1. \end{cases} \tag{3.2.9}$$

This kernel will be used in the numerical part of this chapter.

Asymptotic properties of the RK are studied by Barndorff-Nielsen et al. (2008, 2011b). See also Ikeda (2015). Assume that the bandwidth H is of the order $H = O(n^\alpha)$ with $0 < \alpha < 1$, asymptotic bias and variance of an RK are given by

$$\mathrm{B(RK)} \approx [k''(0)]^2 \omega^2 \frac{n}{H^2} \tag{3.2.10}$$

and

$$var(\mathrm{RK}) \approx 4Tk_\bullet^{0.0} \int_0^T \sigma^4(\tau) d\tau \frac{H}{n} + C_1 \frac{n}{H^3} + C_2 \frac{1}{H}, \tag{3.2.11}$$

where $k_\bullet^{0.0} = \int_0^\infty k^2(u)du$, and C_1 and C_2 are two constants. The quantity $\int_0^T \sigma^4(\tau)d\tau$ is called the daily integrated quarticity. These results indicate that variance of an RK is asymptotically negligible, if $\alpha > 1/3$, and both of its asymptotic variance and bias are negligible, if $\alpha > 1/2$. The asymptotic variance is dominated by the second term on the right-hand-side of (3.2.11), if $1/3 < \alpha < 1/2$, and by the first term, if $\alpha > 1/2$. The asymptotically optimal bandwidth (Barndorff-Nielsen et al., 2009, 2011b), which minimizes the dominating part of the MSE

(mean squared error) of an RK, is given by

$$H_{\mathrm{A}} = c_0 \xi^{4/5} n^{3/5} \quad \text{with} \tag{3.2.12}$$

$$c_0 = \left\{ \frac{k''(0)^2}{k_{\bullet}^{0.0}} \right\}^{1/5} \quad \text{and} \quad \xi^2 = \frac{\omega^2}{\sqrt{T \int_0^T \sigma(\tau)^4 d\tau}}.$$

For the Parzen kernel we have $c_0 = 3.5134$. We see the optimal bandwidth for an RK with a non-flat top kernel is of the order $O(n^{3/5})$. If a bandwidth of this order is employed, the resulted RK will achieve its optimal convergence rate of the order $O(n^{-1/5})$.

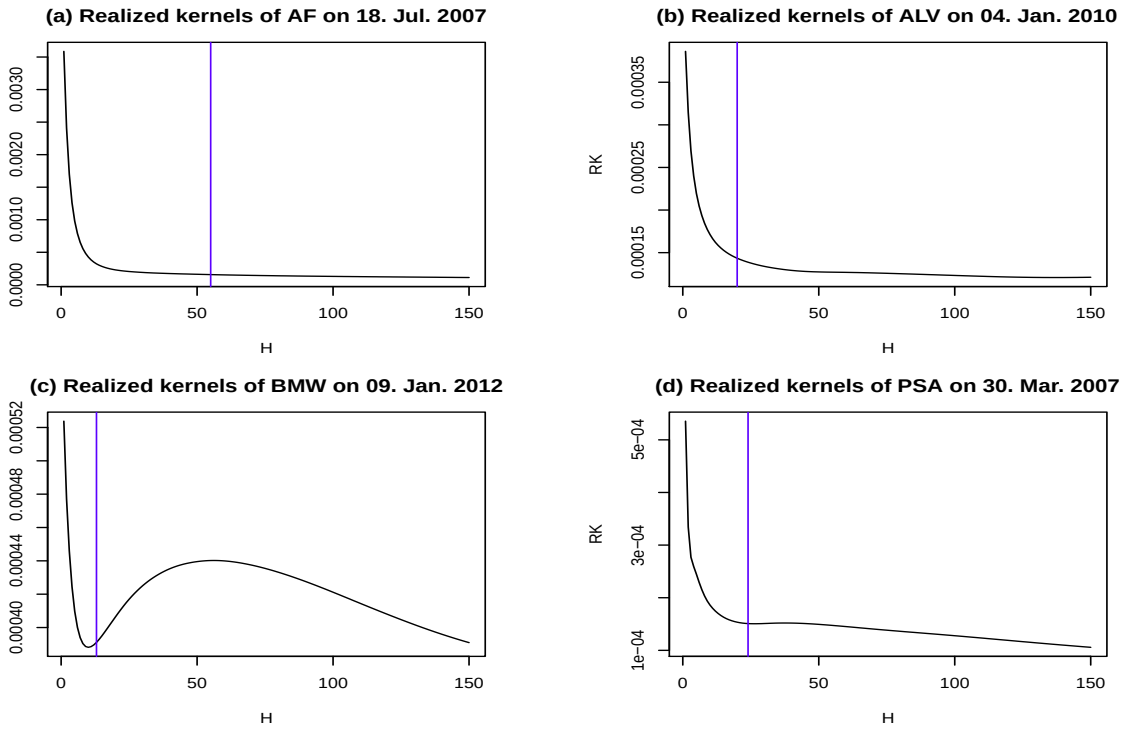

Figure 3.2: Realized kernels against H obtained on the four selected days

The above theoretical results show that realized kernels work well, only if the bandwidth is chosen properly. To show this, the dependence of the RK on the bandwidth H is displayed in Figure 3.2 for the four selected examples, where the vertical line in each panel highlights the bandwidth selected by the procedure proposed in the next section with $\hat{H} = 55$, 20, 13 and 24, respectively. Figure 3.2

shows that an RK is very sensitive to the change of the used bandwidth, if H is small. In a common case, like those in Figures 3.2(a), (b) and (d), the change in H usually does not have a clear effect on the resulted RK, if the used bandwidth is large. However, Figure 3.2(c) indicates that sometimes both of a too large or a too small bandwidth can lead to a clearly wrong estimation result. Detailed discussion on the selected bandwidths will be given in section 3.4.

3.3 Bandwidth selection for realized kernels

Examples in Figure 3.2 show that the selection of the bandwidth is a crucial problem for the application of the RK. In the current context the number of observations on a trading day is very large and one usually would also like to estimate the RK for a number of firms within a long observation period. Hence, we aim at the development of a quick bandwidth selector for the RK with nice theoretical and practical performance. An plug-in bandwidth selector can be obtained by inserting estimates of ω^2 and $\int_0^T \sigma^4(\tau)d\tau$ into H_{A}. However, the estimation of $\int_0^T \sigma^4(\tau)d\tau$ is not yet well solved in the literature. Barndorff-Nielsen et al. (2009) proposed a plug-in bandwidth selector based on the following formula:

$$H_{\mathrm{B}} = c_0 \xi_B^{4/5} n^{3/5} \tag{3.3.1}$$

with ξ^2 in H_A being replaced by $\xi_B^2 = \omega^2/\mathrm{IV}$, which is of the same order as H_{A} but with a biased factor in the constant. The reason is that $T\int_0^T \sigma^4(\tau)d\tau$ can be well approximated through IV^2, if $\sigma(\tau)$ does not vary too much. This biased version of the optimal bandwidth will also be employed in the current chapter . Now, assume that $\widehat{\mathrm{IV}}$ is an at least unbiased estimator of IV, it is easy to show that

$$\hat{\omega}^2 = \frac{\mathrm{RV}_0 - \widehat{\mathrm{IV}}}{2n} \tag{3.3.2}$$

is an consistent estimator of ω^2. Furthermore, if $\widehat{\mathrm{IV}}$ is a consistent estimator, H_{B} can be estimated consistently by replacing ξ^2_{B} with

$$\hat{\xi}^2_{\mathrm{B}} = \hat{\omega}^2/\widehat{\mathrm{IV}}. \tag{3.3.3}$$

The difference between H_{A} and H_{B} is an constant factor $H_{\mathrm{B}}/H_{\mathrm{A}} = (T\int_0^T \sigma_u^4 du/\mathrm{IV}^2)^{1/5}$, which is usually slightly bigger than one. In the following a consistent bandwidth selector of H_{B} is proposed by adapting the IPI idea of Gasser et al. (1991) to the current context with RV_0 and RV_{Z} as the initial values. The proposed algorithm reads as:

Step 1. In the first iteration let $\widehat{\mathrm{IV}}_1 = \mathrm{RV}_{\mathrm{Z}}$. Calculate $\hat{\omega}^2_1$ and $\hat{\xi}^2_1$ following (4.3.5) and (3.3.3). Insert the latter into (3.3.1) to obtain $\hat{H}_1$. Put $j = 2$.

Step 2. In the jth iteration with $j > 1$, calculate $\widehat{\mathrm{IV}}_j$ with $\hat{H}_{j-1}$. Then calculate $\hat{\omega}^2_j$ and $\hat{\xi}^2_j$, and obtain $\hat{H}_j$ similar to Step 1.

Step 3. Increase j by 1 and carry out Step 2 repeatedly. The procedure will be ended, if convergence is achieved or some stopping criterion is fulfilled, or a maximal number of iterations J is carried out. Put $\hat{H} = \hat{H}_j$.

We will see that $\hat{H}_1$ is an inconsistent bandwidth selector. But after a few iterations, $\hat{H}_j$ will become a consistent estimate of H_{B}. The detailed behavior of $\hat{H}_j$ in each iteration and the theoretical properties of the finally selected bandwidth are discussed in the following theorem and its proof.

Theorme 3.3.1. *Assume that Conditions* **K**, **SH**, **D** *and* **U** *in Theorem 2 of Barndorff-Nielsen (2011b) hold. Assume further that u_i are i.i.d. and that the end-effect as indicated in that paper is treated suitably, so that it does not affect the asymptotic performance of $\widehat{\mathrm{IV}}_j$ in the proposed procedure. Then we have*

i) $\hat{H}_j$ selected by the proposed procedure with $j \geq 3$ is a consistent estimator of H_{B} in the sense that $(\hat{H}_j - H_{\mathrm{B}})/H_B = o_p(1)$.

ii) For $j \geq 4$ the selected bandwidth is consistent with a relative convergence rate of the order $n^{-1/5}$, i.e. $(\hat{H}_j - H_{\mathrm{B}})/H_{\mathrm{B}} = O_p(n^{-1/5})$.

A sketched proof of Theorem 3.3.1. *The following proof is carried out conditioning on given number of observations n on a trading day.*

i) In the first iteration, we have $\widehat{\mathrm{IV}}_1 = \mathrm{IV}_{\mathrm{Z}} = \mathrm{IV} + O_p(n^{1/2})$. *Furthermore, it can be shown that* $\hat{\omega}_1^2$ *is* $\sqrt{n}$*-consistent such that* $\hat{\omega}_1^2 = \omega^2[1 + O_p(n^{-1/2})]$ *and* $\hat{\xi}_1^2 = O_p(n^{-1/2})$. *This results in an estimate* $\hat{H}_1 = O_p(n^{\alpha_1})$ *with* $\alpha_1 = 2/5 > 1/3$. *Following the asymptotic results summarized in (3.2.10) and (3.2.11), the use of* $\hat{H}_1$ *in the second iteration will lead to an estimate with an asymptotically negligible variance and a random bias term of the order* $O\left(\frac{n}{H_1^2}\right) = O_p(n^{1/5})$. *That is we have* $\widehat{\mathrm{IV}}_2 = \mathrm{IV} + O_p(n^{1/5}) + o_p(1)$. *Now, it can be shown that* $\hat{\omega}_2^2 = \omega^2[1 + O_p(n^{-1/2}) + O_p(n^{-4/5})]$ *with an additional term caused by the bias in* $\widehat{\mathrm{IV}}_2$, *which is still* $\sqrt{n}$*-consistent. Furthermore, we have* $\hat{\xi}_2^2 = O_p(n^{-1/5})$. *Insert these results into the proposed algorithm we obtain* $\hat{H}_2 = O_p(n^{\alpha_2})$ *with* $\alpha_2 = 13/25 > 1/2$. *The estimates* $\widehat{\mathrm{IV}}_3$, $\hat{\omega}_3^2$ *and* $\hat{\xi}_3^2$ *in the third iteration obtained with* $\hat{H}_2$ *are all consistent. Hence,* $\hat{H}_3$ *is a consistent estimate of* H_{B} *in the relative sense.*

ii) Note that the error of $\hat{H}_j$ *is dominated by that of* $\widehat{\mathrm{IV}}_j$. *Using Taylor expansion of a random function it can be shown that the rate of convergence of* H_j *is the same as that of* $\hat{\xi}_j^2$. *From the fourth iteration onwards,* $\hat{H}_j$ *achieves its optimal rate of convergence of the order* $O_p(n^{-1/5})$ *in a relative sense with* $(\hat{H}_j - H_{\mathrm{B}})/H_{\mathrm{B}} = O_p(n^{-1/5})$ *for any* $j \geq 4$. ◇

The proof above shows that, theoretically, at least three steps are required to achieve a consistent selector of H_{B}. Asymptotically, the performance of $\hat{H}_4$ might be slightly better than that of $\hat{H}_3$, because $\widehat{\mathrm{IV}}_4$ is obtained with a consistent bandwidth selector. The proposed data-driven algorithm and the above theoretical results can be easily adapted to the case, if an unbiased estimate of $\int \sigma^4(\tau)d\tau$ is used. Furthermore, note that the bandwidth for an RK is an integer. Small changes in the involved quantities often do not have any effect on the finally selected bandwidth. Hence, the proposed algorithm converges very quickly. This is also confirmed by the application in the next section.

An R code is developed for practical implementation of the proposed bandwidth selector. The procedure will be stopped, if $\hat{H}_j = \hat{H}_{j-1}$ is achieved. It is also found that sometimes the selected bandwidths take two consequent integers

alternatively. If this happens, the procedure will also be ended. Now, the larger of the two selected bandwidths will be used. Both cases will be considered as regular cases (Reg. Case). Furthermore, the following three special cases can also happen. The first special case (Sp. Case 1) is that with $\mathrm{RV_Z} < 0$. This indicates that the MS noise should be very strong (and maybe correlated). Now, $\hat{H}_1$ can not be calculated according to the proposed algorithm, because we have $\widehat{\mathrm{IV}}_1 < 0$. In this case we will manually set $\hat{\xi}_1^2 = 100/(2n)$, which will lead to a big starting bandwidth $\hat{H}_1$. Nevertheless the procedure runs very well and, after a few iterations, H_j will converge to the selected optimal bandwidth, which is independent of $\hat{H}_1$. The second special case (Sp. Case 2) is that with $\mathrm{RV_Z} > \mathrm{RV}_0$, which indicates probably that there is no MS noise on that day. Now, the procedure cannot be carried out, because we have $\hat{\omega}_1^2 < 0$. In this case, we will set $\hat{H} = 0$ and simply use RV_0. The last special case (Sp. Case 3) will happen, when RK becomes bigger than RV_0 in some iteration with $j > 1$. This means again that there is no strong MS noise in the observed prices on that day. And now the proposed algorithm cannot be carried out further. Hence, we will put $\hat{H} = \hat{H}_j$ as the selected optimal bandwidth.

Note that the optimal bandwidth for an RK under independent and dependent MS noise is of the same order of magnitude. The proposal bandwidth selector can hence be applied to the case with dependent MS noise. The example in Figure 3.2 (d) also indicates that the proposal works in this case. But now, the selected bandwidth is only sub-optimal, because it is only of a correct order but with a clearly biased constant.

3.4 Application

The proposed algorithm is applied to the datasets of AF, ALV, BMW and PSA from 2. Jan. 2006 to 30. Jun. 2012, downloaded from the "Thomson Reuters" Corporation. The total number of trading days for the two German Stocks is 1655 and that for the two French Stocks is 1664. The numbers of days in the four cases with different behavior of the algorithm as described in the last section are listed

in Table 3.1. We see the proposed algorithm converged for more than 99% of

Table 3.1: Numbers of days in different cases for the four companies

Cases / Firms	**Reg. Case**	**Sp. Case 1**	**Sp. Case 2**	**Sp. Case 3**
AF	1662	0	1	1
ALV	1652	0	1	2
BMW	1643	0	1	11
PSA	1653	1	3	8

those datasets. The three special cases only happened with a very small chance. Sp. Case 1 occurred only once by PSA. Sp. Case 2 and Sp. Case 3 for AF and ALV also occurred rarely. For BMW and PSA, Sp. Case 3 occurred on 11 and 8 days, respectively. Now, the ratio of Sp. Case 3 is still clearly smaller than 1%. Therefore, the three special cases can be considered as some rare extreme events. Trading days on which Sp. Cases 2 or 3 happened will be called challenging days. Now, the proposed algorithm does not work well. This will be discussed in section 4.2 in detail.

3.4.1 Summary of the general findings

Figure 3.3 shows the histograms of the selected optimal bandwidths for the four companies. It seems that the proposed bandwidth selector is nearly asymptotically normally distributed. For finite samples the distribution is sometimes slightly skewed to the right with very fewer extremely large selected bandwidths. The selected bandwidths are usually between 5 and 25. The largest selected bandwidth is 55 by AF on 18. Jul. 2007 as indicated in Figure 3.3 (a). As defined before, the selected bandwidth is 0, if Sp. Case 2 happened. From Figure 3.2 we can see that the use of the selected bandwidth leads to an estimate, which is at a very low level, but not the lowest value of all possible RK. This feature is as expected and shows that the proposed bandwidth selector works very well in practice. This nice property is particularly highlighted by Figure 3.2 (c). Note that the bandwidth

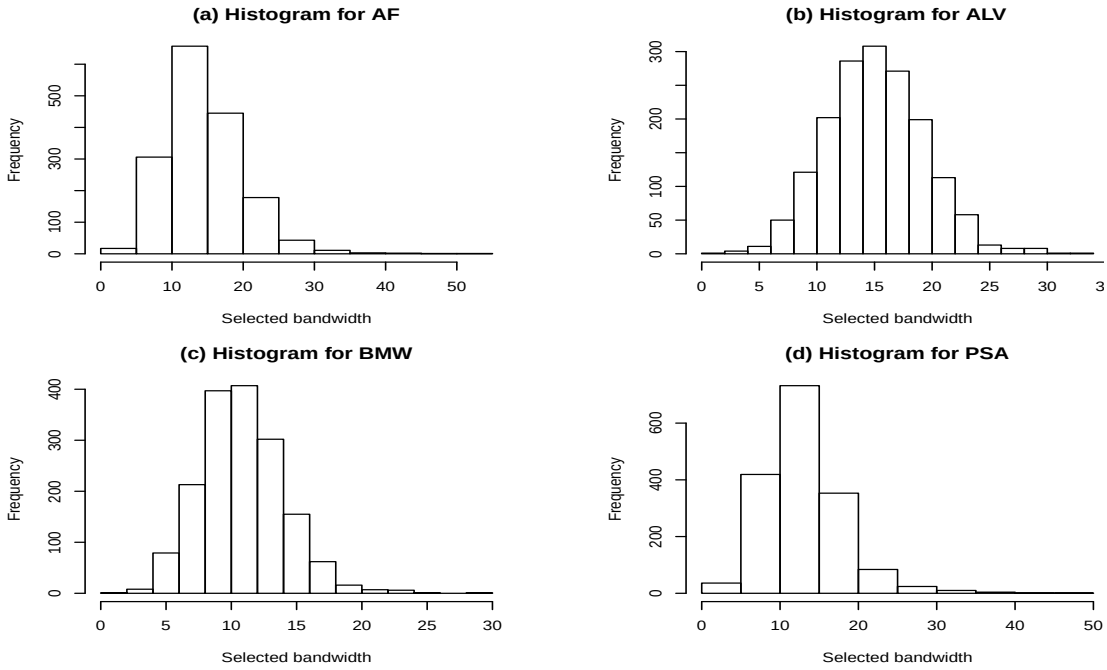

Figure 3.3: Histograms of selected bandwidth for all examples.

selected by the procedure of Barndorff-Nielsen et al. (2009) is usually very large. This is not only caused by the use of different algorithms, but also by the different features of the used datasets. It is of great interest to carry out a comparative study between the two proposals theoretically and through simulation. This is however beyond the aim of the current chapter and will be discussed elsewhere. The histograms of the numbers of iterations for all companies are displayed in Figure 3.4. In the R code a maximal number of iteration $J = 15$ is used. For the datasets under consideration this limit is never achieved. The maximal number of iterations occurred is 11 by PSA on one day. The maximal numbers of iterations by AF, ALV and BMW are 6, 5 and 6, respectively. And the most possible number of iterations for all of the four companies is 3. Furthermore, in most of the cases the proposed algorithm converges within four iterations. This confirmed the results of Theorem 3.3.1.

The estimated RV_0, RV_Z and RK are summarized in Table 3.2, where the t statistic for the differences between RV_0 and RV_Z, and those between RV_Z and RK are also given in the second and third rows, respectively. These t values are calcu-

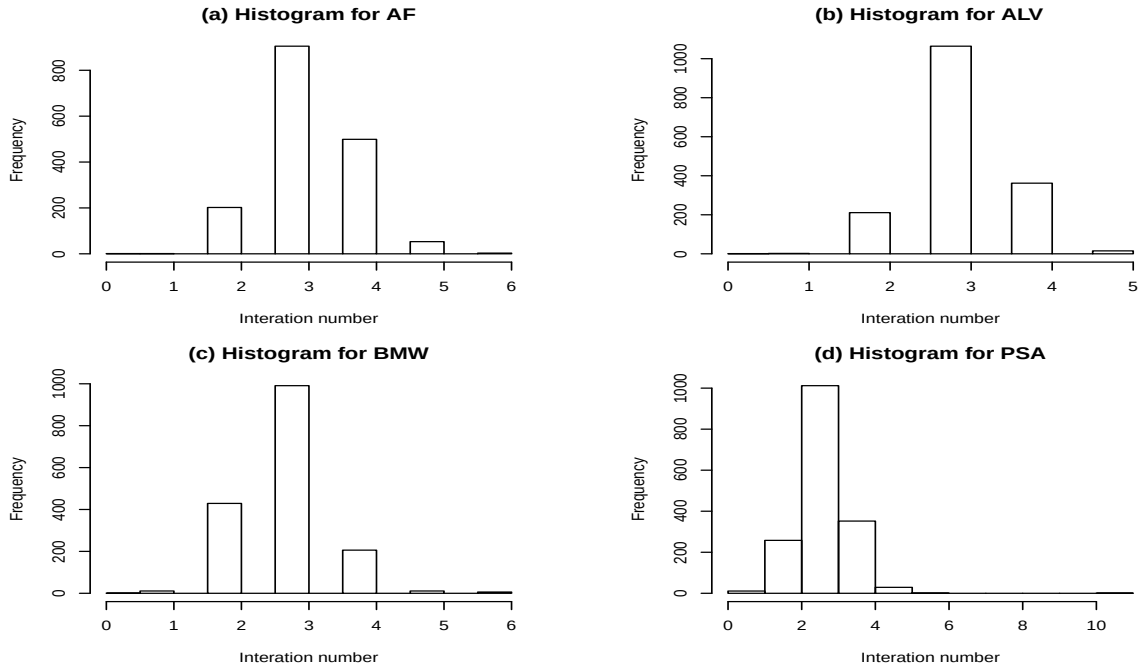

Figure 3.4: Histograms of the number of iterations for all examples.

Table 3.2: Statistics of RV_0, RV_Z and RK; t between RV_0 & RV_Z, and RV_Z & RK

	AF			ALV			BMW			PSA		
	mean	s.d.	t	mean	s.d.	t	mean	s.d.	t	mean	s.d.	t
RV_0	12.74	14.60	—	10.18	24.03	—	7.71	11.35	—	12.11	14.02	—
RV_Z	8.05	8.84	26.79	6.82	16.25	17.27	5.54	8.26	26.09	8.28	9.39	27.20
RK	6.29	6.62	23.48	5.28	10.02	9.41	5.16	6.55	6.90	7.07	7.57	19.03

lated under the assumption that those differences in a given case are i.i.d. We see that for each of the four companies the mean of RV_0 is much larger than those of RV_Z and RK. The mean of RV_Z is also bigger than that of RK. The differences between those mean values are always very highly significant. The difference between the means of RV_0 and RV_Z indicates the part of the bias caused by the MS noise, which can be discovered by $\hat{\rho}(1)$ of the returns. And the difference between the means of RV_Z and RK indicates additional bias caused by possible dependent MS noise, which can not be reflected by $\hat{\rho}(1)$ of the returns. The differences between

the standard deviations of those estimators are similar to that mentioned above. It can also be shown that those differences are always significant. Details to this end are omitted to save space. In summary, the use of the proposed data-driven RK will lead to a clear reduction of the bias and the variation, comparing with the two well known estimators RV_0 and RV_Z. These empirical findings show that the proposed approach works well in practice. However, the practical performance of the proposed data-driven algorithm for RK still need to be confirmed by means of a simulation study.

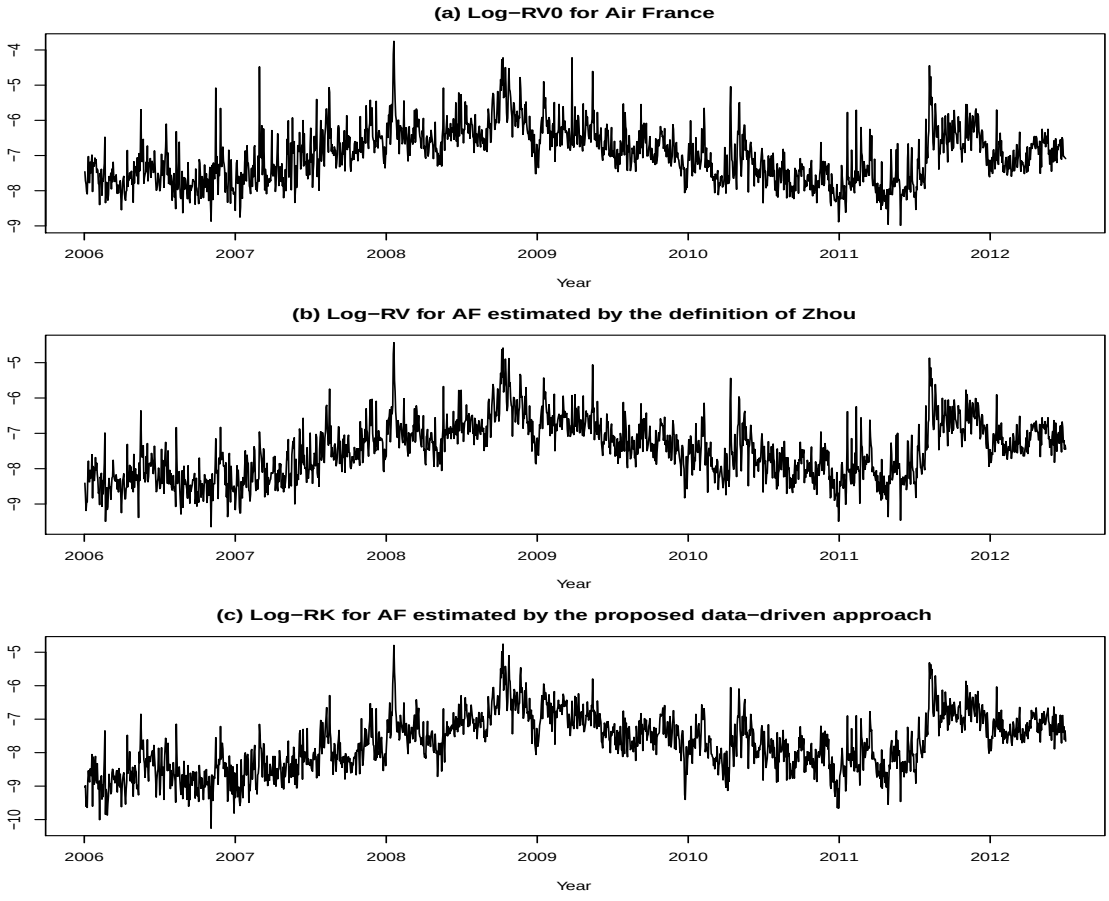

Figure 3.5: Logarithmic transformation of all realized volatility estimators for Air France

The results of RV_0, RV_Z and RK for AF after logarithmic transformation are shown in Figure 3.5. From this figure we can see that, in addition to the differences among the estimates obtained by these different approaches, they also

exhibit quite similar common patterns. In particular all of these series seem to have a non-stationary trend component and possible structural breaks caused by two financial crises, i.e. the global financial crisis in 2008 and the European debt crisis in 2011, respectively. The results of the data-driven RK for ALV, BMW and PSA, again in log-scale, are displayed in Figure 3.6. We see, these series also share similar patterns as those displayed in 3.5(c). Moreover, an interesting empirical finding is that, in addition to the common general tendency of those RK series, they seem also correlated to each other strongly. This feature is helpful for further modeling and forecasting of RK.

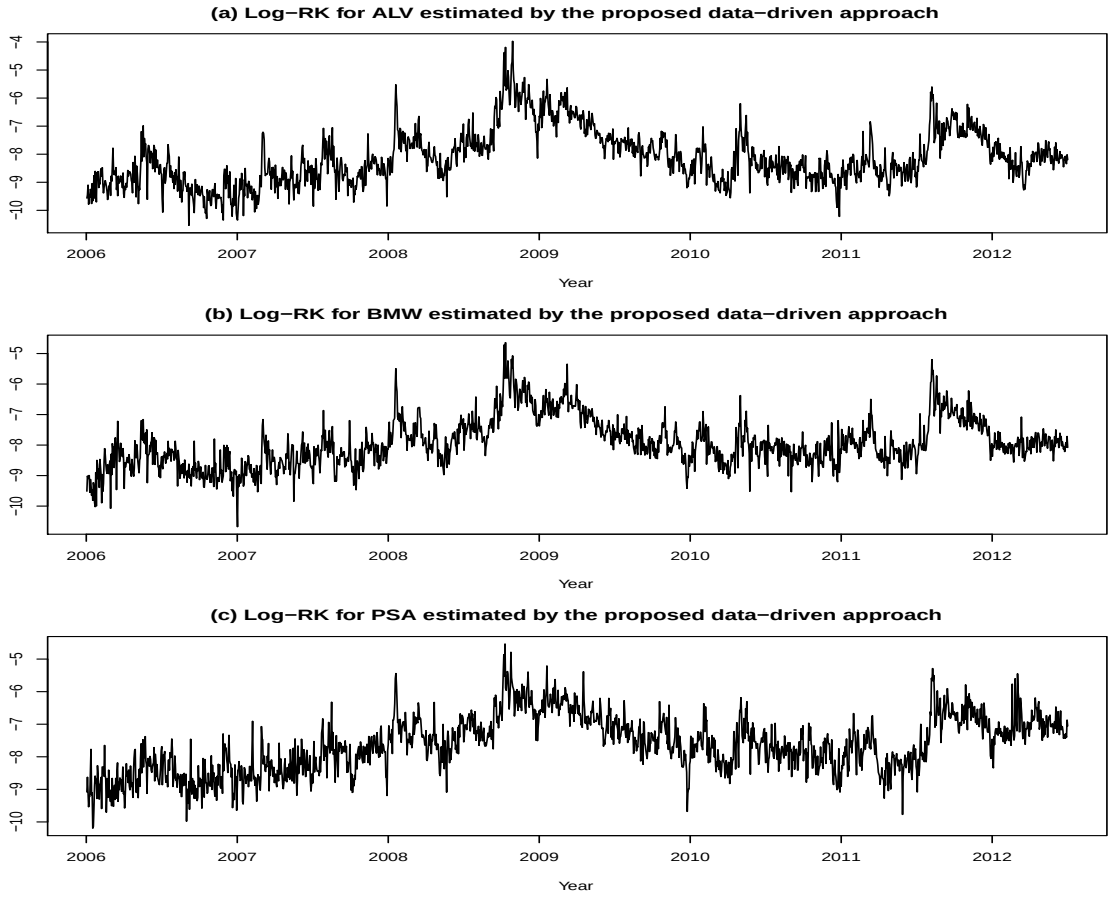

Figure 3.6: Logarithmic transformation of realized kernels for the other three companies

3.4.2 Detailed analysis of two challenging cases

Now, we will discuss briefly, why the proposed data-driven algorithm does not work well in Sp. Cases 2 and 3. The dataset of AF on the 15. Sept. 2011 was chosen as an example of Sp. Case 2 and that of BMW on the 21. May 2009 was chosen as an example of Sp. Case 3. Figure 3.7 shows the ACF of the intraday returns on those two challenging days. As shown in Figure 3.1, usually, MS noise will cause a clearly significant negative ACF at lag 1. From Figure 3.7 we can see

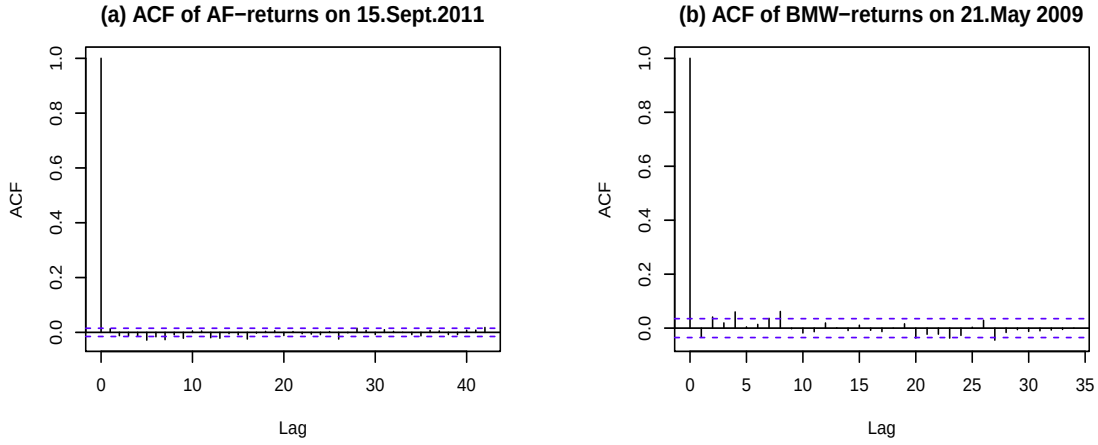

Figure 3.7: ACF of the high-frequency returns on the two challenging days

that this is not true in both examples selected here. Figure 3.7 (a) shows that almost all of the estimated ACF in this case are insignificant. But the ACF at lag 1 happens to be slightly positive. This results in turn in the fact that RV_Z is slightly larger than RV_0. In this case we proposed the use of $\hat{H} = 0$, because now the effect of the MS noise seems to be unclear. One problem can arise in the presence of dependent MS noise. Now, it can happen that although the ACF at lag 1 is positive, but some ACF at higher lags can be negative so that RV_0 is still biased. This kind of effect of MS noise can however not be corrected by the proposed data-driven RK. The problem in Sp. Case 3 is different. As we can see, now the ACF at lag 1 is negative and hence the proposed bandwidth selection algorithm can be started. However, some other ACF are clearly positive so that the sum of the ACF is now positive. This indicates again the existence of possible dependent MS noise. This kind of noise could however cause a negative

bias in RV_0. Following our proposal, the resulted RK in this special cease is always slightly bigger than RV_0. The effect of this kind of possibly dependent MS noise can also not be captured by the proposed algorithm. Both examples indicate that the proposed algorithm should still be improved and it is worthy to development a data-driven RK by taking possibly dependent microstructure noise into account.

3.5 Further analysis using the Semi-FI-Log-ACD

Further analysis of the obtained RK is of great interest. Ebens (1999) showed that the distribution of the logarithmic volatility is approximately normal. Anderson et al. (2003), Corsi (2009) and Koopman et al. (2005) showed that logarithmic realized volatility may exhibit high persistence. From Figures 3.5 and 3.6 we can see that the logarithmic RK may also exhibit a deterministic nonparametric trend. The well known SEMIFAR (Beran and Feng, 2002a) is a nonparametric regression model with long-range dependence. Most recently, Beran et al. (2015) proposed to apply the SEMIFAR model to logarithmic transformation of nonnegative financial time series. Their proposal is hence called an ESEMIFAR model, which can be applied to RK. See also Feng and Zhou (2015a) for discussion on forecasting based on this approach. Assume that Z_t, the log-transformed RK, follow a SEMIFAR model

$$(1-B)^d \phi(B)[Z_t - \mu(\tau_t)] = \varepsilon_t, \tag{3.5.1}$$

where B denotes the backshift operator, $\phi(B)$ is the AR-characteristic polynomial, ε_t are i.i.d. normally distributed random variables with $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = \sigma_\varepsilon^2$; $d \in (-0.5, 0.5)$ and $\tau = t/n$ denotes the rescaled time. The existing data-driven algorithms of the SEMIFAR can be used to fit (3.5.1), where the AR model is selected by the BIC. A very nice property of this proposal is that, if $d > 0$, the long memory parameter in the original and the log-data is the same. See Beran et al. (2015) for more details.

In the following, the RK series of Air France is used as an example. Like the nonparametric trend, a financial crisis will also cause spurious long memory, if

long memory is estimated without taking possible structural breaks into account. Hence, we will apply the ESEMIFAR model to the whole series as well as to the two sub-series from 2. Jan. 2006 to 30. Sept. 2008, and from 1. Oct. 2008 to 30. Apr. 2011. These sub-periods are defined manually. Discussion on the detection of structural breaks under the SEMIFAR model is beyond the purpose of this chapter. The sub-series after May 2011 is very short and is hence not considered.

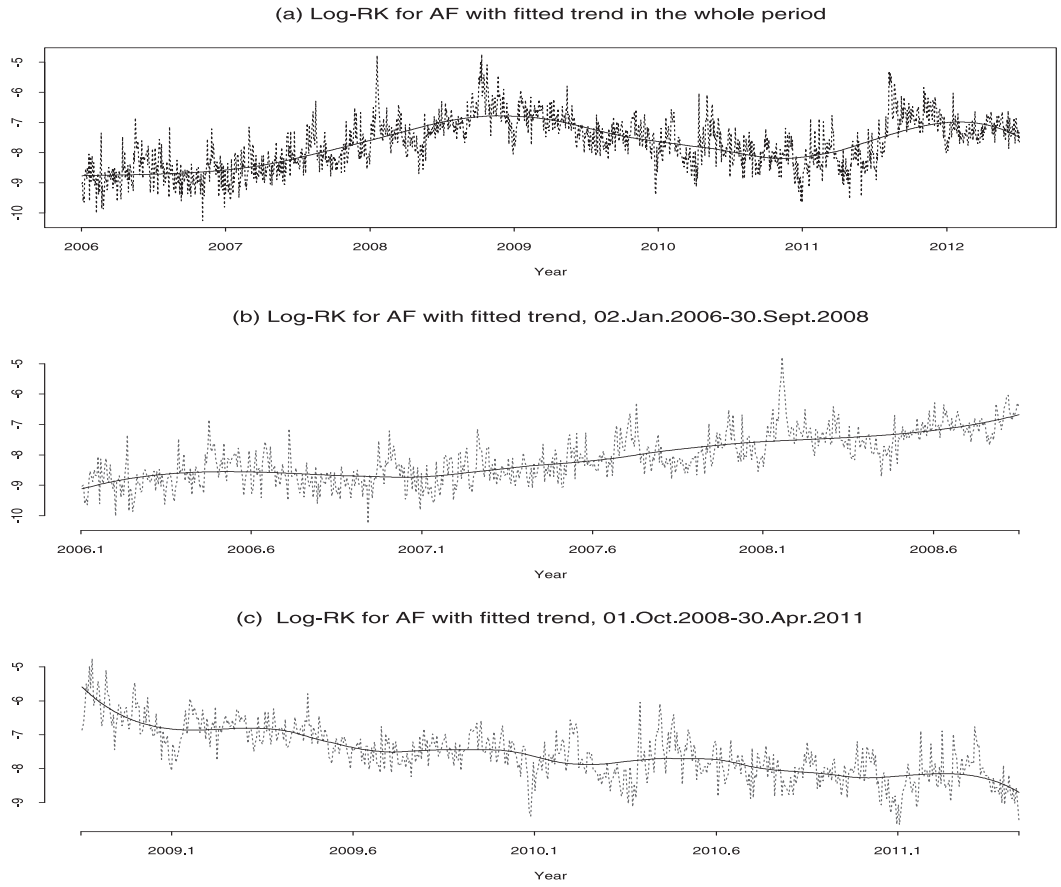

Figure 3.8: Estimated trend by ESEMIFAR together with the log-data

An ESEMIFAR model with a third order local polynomial is fitted to the whole series and to each of the two sub-series mentioned above. The fitted trends together with the data are displayed in Figure 3.8(a) to (c), respectively. The trend in Figure 3.8(a) indicates clear effect of the two financial crises on the market volatility. However, it seems that there is no more structural breaks in the two sub-series. And now the ESEMIFAR fits the data well. The selected bandwidths

Table 3.3: Results of ESEMIFAR for realized kernels of Air France

Series	$\hat{h}$	$\hat{d}$ **& %-CI**	$\hat{p}$	$\hat{\phi}_1$ **& 95%-CI**	**trend**
Jan. 2006 - Jun. 2012	0.191	0.437 [0.399, 0.475]	0	—	insign.
Jan. 2006 - Sep. 2008	0.269	0.380 [0.322, 0.438]	0	—	sign.
Oct. 2008 - Apr. 2011	0.149	0.285 [0.175, 0.395]	1	0.152 [0.014, 0.290]	sign.

($\hat{b}$), the estimated long memory parameters ($\hat{d}$) and the selected AR model, if applicable, are listed in Table 3.3, where the 95%-confidence intervals and the results of the significant test of the fitted trend are also given. From this table we can see that in both sub-periods realized kernels exhibit significant long memory and a significant non-parametric trend simultaneously. In the second sub-period, the short memory part of this model is also significant. Comparing the results for the whole series with those for the two sub-periods, we can see that possible structure breaks cause by the two financial crises exhibit at least the following effects on the estimated ESEMIFAR model: The possible structure breaks resulted in clear overestimation of the long memory parameter, which in turn caused the wrong conclusion, that the estimated trend were insignificant.

3.6 Final remarks

An IPI algorithm for realized kernels under independent MS noise was proposed. To our knowledge this is the first IPI algorithm in the current context. It is shown that this proposal has some nice theoretical properties, runs very quickly and works usually very well in practice. Possible problems which can happen on some challenging days are discussed in detail. It is also proposed to analyze the resulted RK using the most recently proposed ESEMIFAR model. We also tried to apply this model to different pieces of the whole series. There are still some open questions in this context. Firstly, it is better, if one can find more reasonable solutions to the problems on the challenging days. Secondly, it is worthy to extend the current proposal to cases with dependent MS noise. Thirdly, the

proposed bandwidth selector can also be improved, if an unbiased estimator of the daily integrated quarticity can be developed. Furthermore, to apply the idea of the piecewise ESEMIFAR model properly, a suitable approach for detecting structural breaks under the SEMIFAR model should also be developed. Finally, the development of a multivariate semiparametric long memory time series approach for jointly modeling of different RK series is also of great interest.

Chapter 4

A comparison study of realized kernels using different sampling frequencies

This chapter has been published with slight differences as Working Paper (2018-02) in the Working Paper Series of the Faculty of Business Administration and Economics at Paderborn University.

4.1 Introduction

In recent years, there has been a large and rapidly expanding literature on estimation of the daily integrated volatility (IV). With the improvement of availability of high-frequency financial data Andersen et al. (2001a, b) proposed a model-free volatility approach, called realized volatility (RV_0), which exploits the information in high frequency returns and is constructed as the sums of intraday squared returns. However, observed prices are contaminated by market MS noise, which leads to the bias problem at high sampling frequencies (Hanse and Lunde, 2006a). To solve this bias problem, different approaches are introduced into the literature. Bandi and Rusell (2006, 2008, 2011b) proposed a method of selecting the optimal sampling frequency based on a trade-off between the variance and bias. Zhang et

al. (2005) proposed a subsampling method, called two time scales, to deal with the bias problem under i.i.d. MS noise. Their proposal was developed to the case under dependent MS noise by Zhang (2006) and Aït-Sahalia et al. (2008). This is also the first consistent estimator of IV. Furthermore, a pre-filter was used to weaken the effects of MS noise (e.g. Bollen and Inder, 2002). Large (2007) introduced an alternative estimator to control the MS effects. In addition, Zhou (1996) proposed to include the cross-products between two consequent observations. The author showed that his estimator, called $\mathrm{RV_Z}$ is unbiased under i.i.d. MS noise. This is also the first kernel method to deal with the problem of MS noise. Hansen and Lunde (2004, 2006b) studied extensively this estimator and proposed a simple kernel-based estimator. Barndorff-Nielsen et al. (2008, 2009, 2011b) proposed realized kernels, which are a generalization of $\mathrm{RV_Z}$ and also are consistent estimators of the IV under given conditions.

A crucial problem by applying RK is the selection of the bandwidth. This problem is investigated first by Barndorff-Nielsen et al. (2009) for non-negative RK with an asymptotically optimal bandwidth of the order $O(n^{3/5})$ and an optimal rate of convergence of the order $O(n^{-1/5})$, where n is the number of observations on a trading day. Feng and Zhou (2015b) introduced a fully automatic iterative plug-in (IPI) algorithm to select bandwidth for RK by adapting the idea of Gasser et al. (1991). However, their algorithm is only applicable for the case under the assumption of i.i.d. MS noise. And the algorithm does not work, if three special cases occur. In this chapter we improve this algorithm slightly so that it works for all cases. This improved algorithm is called the IN algorithm in the context of this chapter. Ikeda (2015) proposed two-scale RK, which is a convex combination of two realized kernels with different bandwidths. He showed that his estimator converges to the IV in the presence of the dependent MS noise. Based on Ikeda (2015)'s conclusion, Wang (2014) proposed two IPI algorithms for RK under dependent noise assumption. In his work algorithm B is a fully automatic data-driven algorithm and after slight adjustment we call this algorithm in this chapter the DN algorithm. End effects in the computation of the RK are considered. The non-flat-top Parzen kernel is utilized, because it can guarantee

the nonnegativity of RK. Theoretically, in the fourth iteration the two bandwidths G and H have achieved the optimal rate of convergence. After several iterations, the resulted RK achieves its optimal rate of convergence of the order $O(n^{-1/5})$. The processes of the IN and DN algorithms for RK based on different sampling frequency starting with different bandwidths are investigated. It is shown, that no matter the starting bandwidths, the selected bandwidths as of the second iteration are very close and that the final selected bandwidths are indeed the same. Most recently, Liu et al. (2015) has studied the accuracy of a wide variety of volatility measures constructed from high-frequency data. They concluded that when RV_0 calculated by 5-minute returns is taken as the benchmark measure, it is very hard to be beaten by any measure. Furthermore, if no benchmark is specified, the best estimators appear to e.g. RV_0 based on 1-minute data and RK. Barndorff-Nielsen et al. (2009) compared tick-by-tick and 1-minute RK with RV_0 computed by several sampling frequencies. Their conclusions were obtained by measuring the disagreement between these estimates based on transaction prices and mid-quote prices over 123 days. It was found that both RK estimates are better than any of the RV_0 and the statistical results of both RK estimates are similar. Motivated by all of these we study a comparison of RK using different sampling frequencies (tick-by-tick, 1-minute, 5-minute and 15-minute) calculated by the IN algorithm as well as RK using tick-by-tick returns calculated by the DN algorithm within a long period over 2000 trading days. Meanwhile, RV_0 computed from the different frequency returns are compared with these RK estimators mentioned above. In total, we have 4 sampling frequencies, 2 types of realized measures and two algorithms of a given transaction price series. The detailed comparison of these realized estimators are investigated by comparing their performances in the computation of Value-at-Risk based on the Semi-FI-Log-ACD model (also called the exponential SEMIFAR model, Feng and Zhou, 2015a). Value-at-Risk is considered as an extremely important measure in financial risk management to determine the amount of assets needed to cover possible losses. It is found that the performances of two RK estimators based on the tick-by-tick returns calculated by the IN and DN algorithms are better than any other realized estimator and

hence recommended using in practice.

The rest of this chapter is organized as follows. Different sampling schemes and several volatility estimators used in this chapter are introduced in section 4.2. The improved bandwidth selectors under independent and dependent MS noise assumptions are proposed in section 4.3. The implementation of algorithms as well as the comparison of volatility estimators based on different frequencies of data are reported in section 4.4. Final remarks in section 4.5 conclude.

4.2 Realized measures

4.2.1 Different sampling schemes

Prices are practically observed at discrete and irregularly time intervals. Sampling schemes are rules of data recording. Different sampling schemes can be used for calculating realized measures. Let $p^*_{t,i}$, $i = 1, \ldots, n$, be the logarithmic efficient asset prices at time points $0 = \tau_0 < \tau_1 < \ldots \tau_n < \tau_{n+1} = T$ on trading day t, where n is the total number of observations at day t. Support that, $p^*_{t,i}$ follow a continuous time diffusion process $dp^*_t = \sigma_t dW_t$, where W is a standard Brownian motion and σ_t is the spot volatility process. The daily integrated volatility is given by

$$\mathrm{IV}_t = \int_0^T \sigma_t^2 dt. \tag{4.2.1}$$

We divide the interval $[0, T]$ in n subintervals. The length of the ith subinterval is defined by $\delta_{i,n} = \tau_i - \tau_{i-1}$. The integrated volatility for each of the subintervals is

$$\mathrm{IV}_{i,t} = \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt.$$

The choice of sampling schemes and sampling frequencies can have a strong influence on the estimation of the IV. Sampling schemes are mainly classified according to the concept of time.

Calendar Time Sampling (CTS). CTS is sampled by regularly spaced calendar time and is defined by $\delta_{i,n} = \frac{1}{n}$ for all i (see e.g. French et al. (1987);

Hsieh (1991); Andersen and Bollerslev (1998); Andersen et al. (2001a)). Because of irregularity of the intraday data, calendar time sampled data must be builded artificially (refer to Wasserfallen and Zimmermann (1985); Andersen and Bollerslev (1997); and Dacorogna et al. (2001)). In this chapter we use this sampling scheme. The original tick-by-tick price is sampled by different sampling frequencies, namely 1-minute, 5-minute or 15-minute. We utilize the data for the period between 9:00 am and 17:30 pm and use the previous tick method that takes the first observation as the sampled price (Hansen and Lunde, 2006a). For instance, when the price process is sampled at 1-minute interval, this will yield 511 price observations on one day.

Tick Time Sampling (TkTS) and Transaction Time Sampling (TrTS). Prices are recorded at every price change in TkTS (see e.g. Corsi et al., 2001; Zhou, 1996). Prices are sampled every kth transaction in TrTS (see e.g. Hansen and Lunde, 2006a). Grifin and Oomen (2008) investigated the difference between transaction time and tick time sampling. They found that the MSE of RV in tick time is lower than that in transaction time, especially when the level of noise, number of ticks, or the arrival frequency of efficient price moves is low.

Business Time Sampling (BTS). The sampling time are chosen such that $\mathrm{IV}_{i,t} = \frac{\mathrm{IV}_t}{n}$. The observation times for BTS are unobserved. The BTS transactions are often sampled in order to ensure approximately equal volatility of the returns over each interval. Peters and Vilder (2006), Andersen et al. (2007) and Andersen et al. (2010) selected time points to sample BTS returns with a target volatility.

The CTS, TrTS and TkTS schemes are constructed based on explicit criteria such as the regular calendar-time length or the number of ticks/price change. In addition, the observation times are observed. In contrast, the BTS scheme depends on the unobserved latent volatility. As a result, the most widely used sampling scheme is CTS. In many cases, the BTS scheme has the minimal MSE of realized variance among all sampling schemes, however, it is used less frequently in practice.

4.2.2 The effect of microstructure noise

Let $p_{t,i}$ be the ith logarithmic observed prices during day t. Assuming that the prices are observed with noise, we have

$$p_{t,i} = p^*_{t,i} + u_{t,i}, \tag{4.2.2}$$

where $u_{t,i}$ represents a stationary logarithmic microstructure noise process with mean zero. It is assumed that $u_{t,i}$ is independent of $p^*_{t,i}$. The noise process itself can be a white noise or a dependent process. The corresponding noise contaminated observed returns are obtained by

$$r_{t,i} = p_{t,i} - p_{t,i-1} = r^*_{t,i} + e_{t,i}, \tag{4.2.3}$$

where $e_{t,i} = u_{t,i} - u_{t,i-1}$ is the noise in the observed returns. In the presence of MS noise, realized variance $\mathrm{RV}_0 = \sum_{i=1}^{n} r_{t,i}$ is no more consistent of the integrated variance, and can be rewritten as

$$\mathrm{RV}_0 = \sum_{i=1}^{n} (r^*_{t,i})^2 + 2\sum_{i=1}^{n} r^*_{t,i} e_{t,i} + \sum_{i=1}^{n} e^2_{t,i}.$$

Under independent noise assumptions, the variance of noise is $var(u_{t,i}) = \omega^2$. The bias of RV_0 is $\mathrm{B}(\mathrm{RV}_0) = 2n\omega^2$ and the asymptotic variance of RV_0 is $var(\mathrm{RV}_0) \approx 4nE(u_i^4)$, as $n \to \infty$. For the dependent noise structure, Zhang (2006) and Aït-Sahalia et al. (2008) showed that the bias is still $2n\omega^2$ and the variance tents to $4n\Theta$, where $\Theta = V[(u_{t,1} - u^2_{t,0})] + 2\sum_{i=1}^{\infty} Cov[(u_{t,1} - u^2_{t,0}), (u_{t,i+1} - u^2_{t,i})]$. However, for large values of n, the bias and variance of RV_0 are infinite. Different approaches have been introduced into the literature to improve the performance of RV_0 in the presence of MS noise. Andersen et al. (2001a, 2003) proposed to select arbitrarily lower frequencies returns to treat the MS bias, such as every 5 or 15 minutes, instead of at every tick. RV_0 using different sampling frequencies is defined by

$$\mathrm{RV}^s_0 = \sum_{i=0}^{n_s} r^2_{t,i}, \tag{4.2.4}$$

where n_s denotes the number of observations. It is showed that the bias due to noise for independent and dependent is given by $2n_sE(u_{t,i}^2)$. The bias is reduced when $n_s < n$, however, the variance is increased due to discretization. This leads to the well-known bias-variance trade-off.

Under the i.i.d. assumption on $u_{t,i}$, Zhou (1996) proposed to use the first order correlation to correct the bias in RV_0. This estimator RV_Z is unbiased but inconsistent. The version of RV_Z with different sampling is given by

$$\mathrm{RV}_\mathrm{Z}^s = \sum_{i=2}^{n_s-1}(r_{t,i}^2 + r_{t,i}r_{t,i+1} + r_{t,i}r_{t,i-1}). \tag{4.2.5}$$

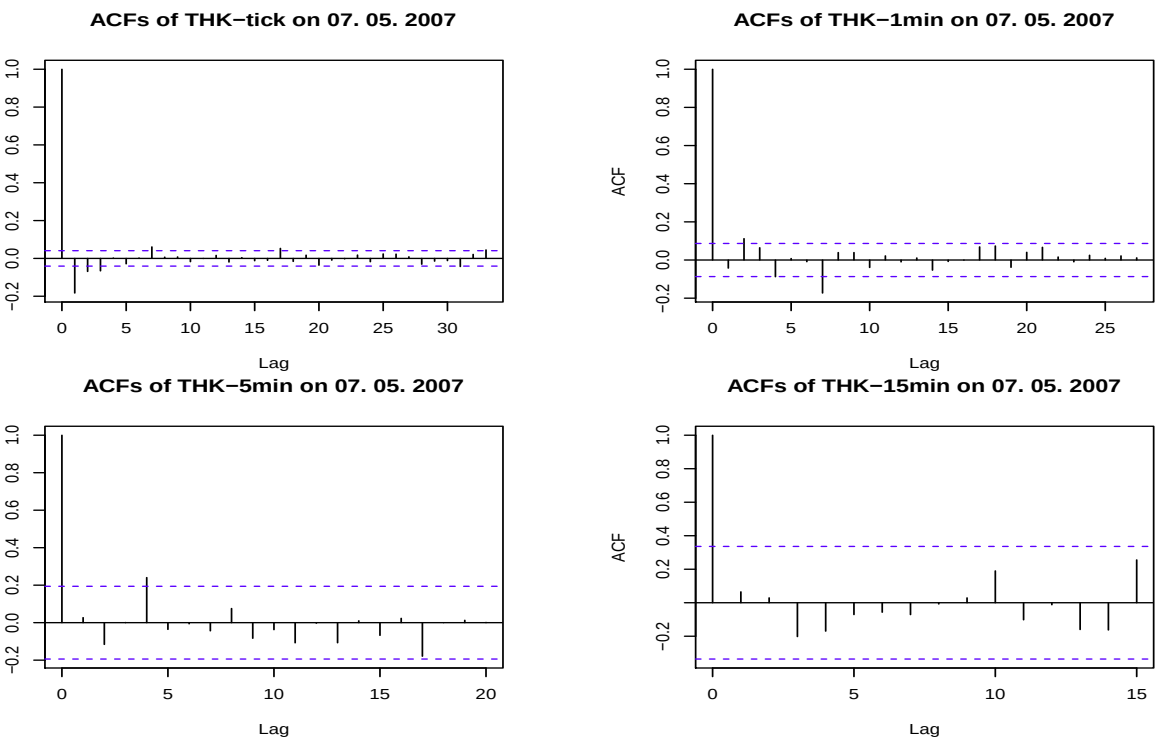

Figure 4.1: ACFs of Thyssenkrupp on 20. Jan. 2012.

Empirical evidence of MS noise can be found by displaying the ACFs of returns. Figure 4.1 to Figure 4.4 show the ACFs of different sampling-frequency returns on four selected trading days, respectively. Generally speaking, the ACFs on every selected trading day vary a lot by changing the sampling-frequency. ACFs of tick-by-tick returns for Thyssenkrupp on 07. May 2007 are displayed in Figure 4.1 (a).

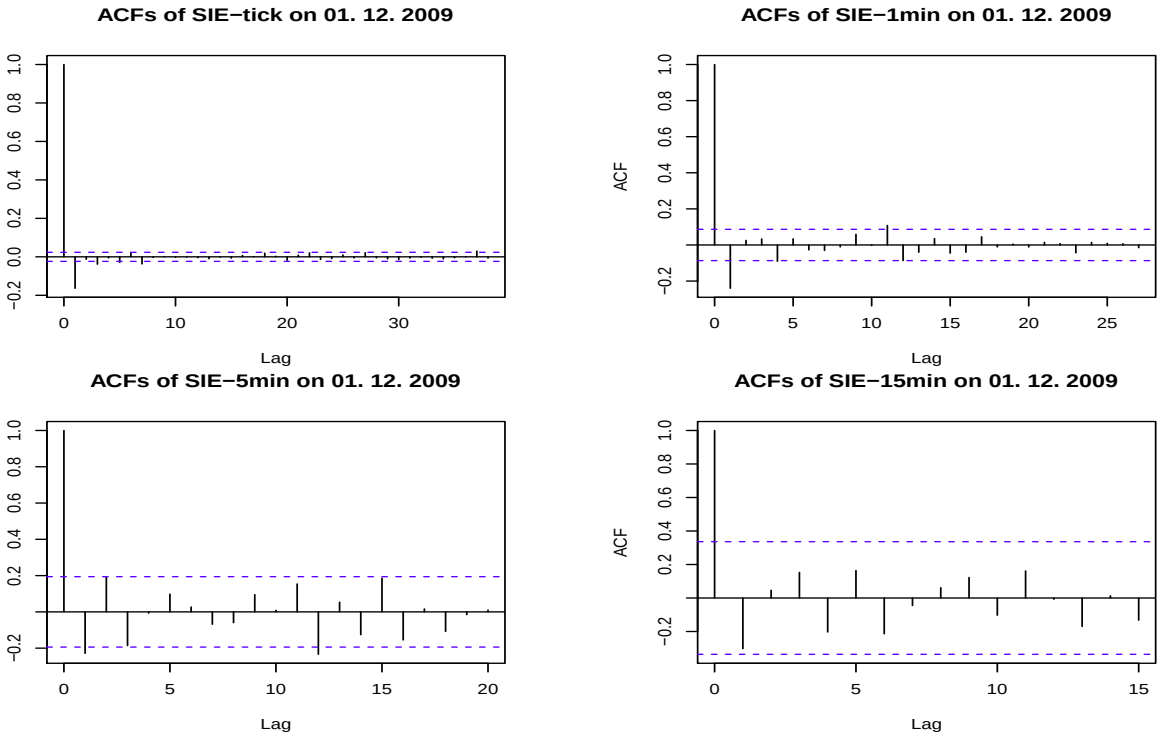

Figure 4.2: ACFs of Siemens on 13. Jun. 2008.

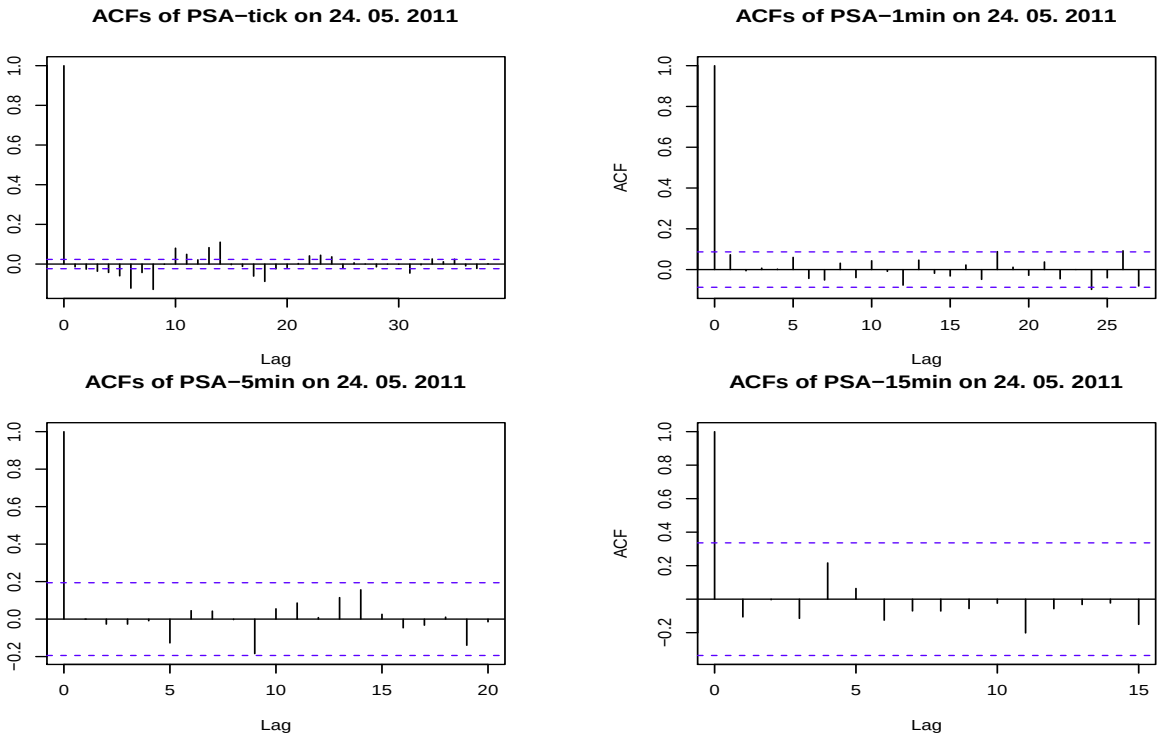

Figure 4.3: ACFs of Peugeot on 24. May 2011.

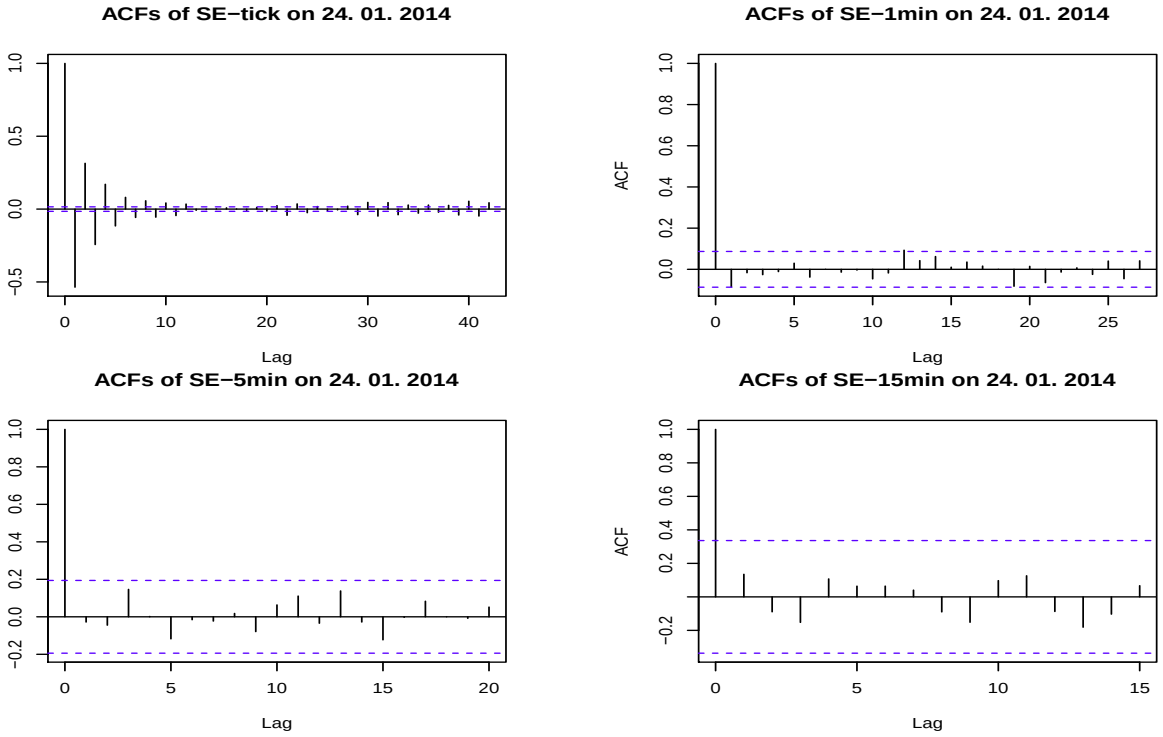

Figure 4.4: ACFs of Schneider Electric on 24. Jan. 2010.

Table 4.1: Detailed results of different realized estimators ($*10^4$) for selected examples

	RV_0				RK			
	tick	1-min	5-min	15-min	IN-tick	IN-1min	IN-5min	DN-tick
THK 07. 05. 2007	3.867	2.413	2.393	2.362	1.558	2.425	2.439	1.611
SIE 01. 12. 2009	3.761	2.518	1.635	0.906	2.258	1.941	1.278	1.744
PSA 24. 05. 2011	5.064	2.699	2.551	1.765	2.657	2.924	2.864	2.448
SE 24. 01. 2014	7.660	2.472	2.044	2.028	2.150	2.114	2.016	2.147

It shows that $\hat{\rho}_r(1) < 0$ and the ACFs at lags 2 and 3 are slightly not-zero. This indicates the existence of possible dependent MS noise. From Figure 4.1 (b) and

(c) we can see that the correlations in 1-minute and 5- minute returns are reduced, but several lags are still outside two bounds. There are almost no correlations in 15-minute returns, which can be seen in Figure 4.1 (d). Figure 4.2 displays the ACFs of Siemens on 01. Dec. 2009. In Figures 4.2 (a), (b) and (c) not only $\hat{\rho}_r(1)$ but also the following several lags are slightly not-zero. The correlations are reduced with the decrease of sample frequencies and there are also almost no correlations in 15-minute returns. ACFs of Peugeot on 24. May 2011 are shown in Figure 4.3. In Figure 4.3 (a) $\hat{\rho}_r(1)$ is not significant, but some ACFs at higher lags are clearly significant. This indicates the presence of dependent MS noise and that the simple independence assumption on the noise may not be sufficient. Using the DN algorithm to calculate RK may be suitable in such a case. The correlations are clearly reduced by using 1-minute returns in Figure 4.3 (b). As has been demonstrated before, there are also almost no correlations in 5- and 15-minutes returns. Figure 4.4 shows the ACFs for Schneider Electric on 24. Jan. 2014. In Figure 4.4 (a) the noise is very strong with $\hat{\rho}_r(1) < -0.5$ and the several lags, which follow, are also clearly significant. On this day $\mathrm{RV_Z} < 0$, which is one of the special cases for the IN algorithm. The correlations diminish clearly by using 1-minute returns in Figure 4.4 (b). There are almost no correlations in 5-minute and 15-minute returns. In conclusion, from Figure 4.1 to Figure 4.4 it can be seen that in most cases $\rho_r(1)$ is significantly negative and there are possible dependent MS noise in tick-by-tick returns and for some cases there may be still MS noise in 1-minute and 5-minute returns. Furthermore, the correlations reduce with the decrease of sample frequencies. The reason is that the diminished number of observations reduces the bias and extends the two ACF bounds, which makes the daily returns uncorrelated. However, decreasing the sampling frequency toward to increase the variance and reduce the accuracy of volatility estimators. Hence, a realized estimator with an optimal frequencies is necessary to investigate, which can remove the effect of MS noise and grantee the accuracy of the estimation at the same time. The corresponding numerical results of $\mathrm{RV_0}$ and RK computed from different sample frequencies for these four selected days are provided in Table 4.1. From the first four columns we see that the values of $\mathrm{RV_0}$ for all examples

reduce by the decease of the sample frequencies. For instance, for SE on 24. Jan. 2014 the tick-by-tick RV_0 is even three times bigger than the 1-minute RV_0. The last four columns show us that the results of RK computed from different frequencies returns by the IN algorithm are clearly different form the results of tick-by-tick RK using the DN algorithm. These differences are caused by the calculated dependent MS noise with the DN algorithm. One of our purpose in this chapter is to investigate which of these realized estimators can estimate the financial market volatility more accurately.

4.3 Bandwidth selection for realized kernels

4.3.1 Realized kernels

Realized kernels are introduced by Barndorff-Nielsen et al. (2008, 2009, 2011b) and can be thought of as an extension of RV_{Z}.

$$\mathrm{RK} = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \quad \gamma_h = \sum_{i=|h|+1}^{n} r_i r_{i-|h|}, \tag{4.3.1}$$

where $k(x)$ is a kernel weight function, H is the selected bandwidth and γ_h is the h-th realized autocovariance. In this chapter we use the Parzen kernel, which can guarantee the nonnegative realized kernels estimate.

Under $H = O(n^{\alpha})$ with $1/2 < \alpha < 1$ and further regularity conditions, RK is a consistent estimator of IV with the AMSE (asymptotic mean squared error):

$$\mathrm{AMSE}(H) = [K''(0)]^2\Omega^2 n^2 H^{-4} + 4K_{\bullet}^{0,0}\,(T\mathrm{IQ})\,Hn^{-1}, \tag{4.3.2}$$

where $K_{\bullet}^{0.0} = \int_0^{\infty} k(x)^2 dx$ is a constant, IQ$= \int_0^T \sigma_t^4 dt$ is called the integrated quarticity, $\Omega = \Sigma_{h>0}\Omega(h)$ is the long-run variance of $u_{t,i}$ and $\Omega(h) = \Sigma_{i=h+1}^{n} u_{t,i}u_{t,i-h}$. The first part and second part in Eq.(4.3.2) are the asymptotic squared bias and variance of RK, respectively.

The asymptotically optimal bandwidth which minimizes the AMSE is provided

by

$$H = c_0 \xi^{4/5} n^{3/5} \quad \text{with} \quad c_0 = \left\{ \frac{k''(0)^2}{k_{\cdot}^{0.0}} \right\}^{1/5} \quad \text{and} \quad \xi^2 = \frac{\Omega}{\sqrt{T\text{IQ}}}, \tag{4.3.3}$$

where $c_0 = 3.5134$ for the Parzen kernel. The bandwidth H depends on the unknown quantities Ω and IQ.

4.3.2 Bandwidth selection under i.i.d. noise

The crucial problem when applying RK is the selection of the bandwidth. The asymptotically optimal bandwidth given in Eq. (4.3.3) provides a base for selecting the bandwidth. To do this, we need to estimate Ω and IQ. When $u_{t,i}$ is under the i.i.d. noise assumption, Ω reduces to ω^2. Following Barndorff-Nielsen et al. (2009), IQ in Eq. (4.3.3) is replaced with IV^2, because the former is not too far from the latter, under conditions, when σ_t^2 does not vary significantly. It can be easy shown that the conclusions for RK in Barndorff-Nielsen et al. (2008, 2009, 2011b) are also suitable for RK based on different sampling frequencies (RK_s). Under the i.i.d assumption a biased version of H, called H^{B} constructed at different sampling frequencies is given by

$$H_s^{\text{B}} = c_0 \xi_{\text{B}}^{4/5} n_s^{3/5} \quad \text{with} \qquad \xi_{\text{B}}^2 = \frac{\omega_s^2}{\text{IV}_{\text{s}}}. \tag{4.3.4}$$

Now, assuming that $\widehat{\text{IV}}_s$ is at least an unbiased estimator of IV_s, it can be shown that

$$\hat{\omega}_s^2 = \frac{\text{RV}_0^{\text{s}} - \widehat{\text{IV}_{\text{s}}}}{2n_s} \tag{4.3.5}$$

is a consistent estimator of ω_s^2. Barndorff-Nielsen et al. (2009) proposed to select the bandwidth by plugging two complex estimates of IV and ω^2. Their proposal is not fully data-driven and the selected bandwidth does not converge to H^{B}. Instead of these two complex estimators, Feng and Zhou (2015b) utilized RV_{Z} as an initial value of IV for estimating the starting bandwidth H_1. Then inserting $\hat{H}_1$ into Eqs. (4.3.4) and (4.3.5), and using the IPI idea to obtain optimal bandwidth

$\hat{H}^{\mathrm{B}}$ automatically. This is the first IPI algorithm and it is shown the selected bandwidth is consistent to H^{B}. However, their algorithm does not work, if three following special cases occur. They are $\mathrm{RV_Z}$ is smaller than zero (Case 1), $\mathrm{RV_Z}$ is larger than RV_0 (Case 2), and RK is larger than RV_0 (Case 3). For Case 1 a large starting bandwidth should be manually set, since $\widehat{\mathrm{IV}}_1 < 0$. For Case 2 they set $\hat{H}^{\mathrm{B}} = 0$ and simply use RV_0, because $\hat{\omega}_1^2 < 0$. And for Case 3 they put $\hat{H}^{\mathrm{B}} = \hat{H}_j$ as the selected optimal bandwidth, because $\hat{\omega}_j^2 < 0$. The last two cases imply that there is possibly no strong MS noise on those days. To avoid these three cases, in this chapter we let $\hat{\omega}_{s,1}^2 = \frac{\mathrm{RV}_0^{\mathrm{s}}}{2n_s}$ replace $\hat{\omega}_{s,1}^2 = \frac{\mathrm{RV}_0^{\mathrm{s}} - \mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}}}{2n_s}$. It aims to avoid $\mathrm{RV}_0^s - \mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}} < 0$. It can be observed that $\frac{\mathrm{RV}_0^{\mathrm{s}}}{2n_s}$ is not far from $\frac{\mathrm{RV}_0^s - \mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}}}{2n_s}$, if $n \to \infty$. Moreover, end effects in the computation of the RK is also considered with $m = 2$. For more details please refer to section 2.2 in Barndorff-Nielsen et al. (2009).

Let j denote the number of iterations. The IN algorithm unfolds as:

Step 1. To resolve end-effects problem. Let $p_0 = \frac{1}{m}(p_{t,1} + \ldots p_{t,m})$, $p_j = p_{t,j+1}$, $j = m \ldots, n - m - 1$ and $p_n = \frac{1}{m}(p_{t,n-m+1} + \ldots p_{t,n})$, where we put $m = 2$.

Step 2. In the first iteration $j = 1$, let $\widehat{\mathrm{IV}}_{s,1} = \mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}}$, calculate $\hat{\xi}_{s,1}^4$ with

$$\hat{\xi}_{s,1}^4 = \left(\frac{\frac{\mathrm{RV}_0^{\mathrm{s}}}{2n_s}}{\mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}}} \right)^2 . \tag{4.3.6}$$

Insert $\hat{\xi}_{s,1}^4$ into Eq. (4.3.4) to obtain $\hat{H}_{s,1}$. Let $j = 2$.

Step 3. In the jth iteration with $j > 1$, calculate $\widehat{\mathrm{IV}}_{s,j}$ with $\hat{H}_{s,j-1}$. Then insert

$$\hat{\xi}_{s,j}^4 = \left(\frac{\frac{\mathrm{RV}_0^{\mathrm{s}}}{2n_s}}{\widehat{\mathrm{IV}}_{s,j}} \right)^2 \tag{4.3.7}$$

into Eq. (4.3.4) to obtain $\hat{H}_{s,j}$.

Step 4. Iteratively carry out this procedure until convergence has been achieved.

The algorithm will be stopped, if $\hat{H}_{s,j}$ is equal to $\hat{H}_{s,j-1}$. As showed in Theorem 1

in Feng and Zhou (2015b), theoretically, both of $\widehat{\mathrm{RK}}_s$ and $\hat{H}_s^{\mathrm{B}}$ become consistent form in the third iteration, while their rate of convergence can still be improved in the fourth iteration. Thereafter, $\widehat{\mathrm{RK}}_s$ achieves its optimal rate of convergence of the order $O(n^{-1/5})$ and this rate of convergence is also shared by $(\hat{H}_s^{\mathrm{B}} - H_s^{\mathrm{B}})/H_s^{\mathrm{B}}$. An R code is developed for practical implementation of the proposed bandwidth selector.

4.3.3 Bandwidth selection under dependent noise

From Eq. (4.3.2) it can be seen that for dependent noise the estimation of the long-run variance Ω may enable the correction of the leading bias of RK. Following Ikeda (2015) we utilize $MK(G)$ to estimate Ω. Define

$$MK(G) = (|k''(0)|nG^{-2})^{-1}\mathrm{RK}(G), \tag{4.3.8}$$

where RK(G) are the realized kernels for bandwidth G. Under the assumption G=$O(n^{\beta})$, $\beta \in (0, 1/2]$ and if $n \to \infty$

$$MK(G) \longrightarrow \Omega + \lim_{n\to\infty} n^{-1}G^2(|k''(0)|)^{-1}\mathrm{IV}_t. \tag{4.3.9}$$

In addition, given G=$O(n^{\beta})$ for $1/(2q+1) \leq \beta < 1/3$ or $\beta = 1/3$, where q is the characteristic exponent of $k''(0)$. If $q = 1$ and $\beta = 1/3$ (for Parzen kernel) are allowed for the asymptotic normality of $MK(G)$. According to the related results in Ikeda (2015), the AMSE_G for the Parzen kernel is obtained by

$$\mathrm{AMSE}_G = \left(\frac{\mathrm{IV}_t}{k''(0)}\right)^2 \frac{G^4}{n^2} + O\left(\frac{G}{n}\right) + O\left(\frac{1}{G^2}\right). \tag{4.3.10}$$

The asymptotically optimal bandwidth $G = O(n^{1/3})$ for MK(G) can be obtained by minimizing the AMSE_G. Two-scale realized kernels can be used to estimate Ω. However, the estimator can be negative. Wang (2014) proposed two IPI algorithms A and B under dependent noise assumption. Algorithm A starts with $H_1 = n^{\alpha}$, $\alpha \in (1/2, 1)$. In the second iteration $\hat{H}_2$ has already achieved the rate of

convergence of the order $O(n^{3/5})$ and after only a few iterations $\hat{H}_j$ converges to H^{B}. He also demonstrated that $\hat{H}_j$ is consistent even when α is outside $(1/2, 1)$. Algorithm B is a fully automatic data-driven algorithm and with slight adjustments it is called the DN algorithm in this chapter. Let j denote the number of iterations. The DN algorithm unfolds as:

Step 1. To resolve end-effects problem. Let $p_0 = \frac{1}{m}(p_{t,1} + \ldots p_{t,m})$, $p_j = p_{t,j+1}$, $j = m \ldots, n - m - 1$ and $p_n = \frac{1}{m}(p_{t,n-m+1} + \ldots p_{t,n})$, where we put $m = 2$.

Step 2. In the first iteration, let $\widehat{\mathrm{IV}}_1 = \mathrm{RV_Z}$, calculate $\hat{\xi}_1^4$ with

$$\hat{\xi}_1^4 = \left(\frac{\frac{\mathrm{RV_0}}{2n}}{\mathrm{RV_Z}} \right)^2 . \tag{4.3.11}$$

Insert $\hat{\xi}_1^4$ into Eq. (4.3.4) to obtain $\hat{H}_1$. Let $j = 2$.

Step 3. In the jth iteration with $j > 1$, calculate $\widehat{\mathrm{IV}}_j$ with $\hat{H}_{j-1}$. Meanwhile, similar to step 2, let $G = H_{j-1}^{5/9}$ to obtain $\widehat{\mathrm{RK}}(G)$. The unknown value $\hat{\xi}_j^4$ is calculated by

$$\hat{\xi}_j^4 = \left(12^{-1} n^{-1} G_j^2 \frac{\mathrm{RK(G)}}{\mathrm{IV}_j} \right)^2 \tag{4.3.12}$$

and then $\hat{H}_j$ can be obtained.

Step 4. Iteratively carry out this procedure until convergence has been achieved.

Like in the IN algorithm IV^2 is also used to estimate IQ. Let $\Omega_1 = \frac{\mathrm{RV_0}}{2n}$, so that a negative Ω_1 is avoided. Please note that for both IN and DN algorithms $\hat{H}_j$ is obtained by truncating the integer part of selected optimal bandwidth and subsequently adding 1. The proof for the convergence of this algorithm was given in Wang (2014). Theoretically, in the fourth iteration G and H have achieved to their optimal rate of convergence, respectively. After several iterations the resulted RK achieves its optimal rate of convergence of the order $O(n^{-1/5})$.

4.4 Application

The datasets of ten European companies (Air France (AF), Allianz (ALV), BMW, Deutsche Bank (DBK), Michelin (MIC), Peugeot (PSA), RWE, Schneider Electric (SE), Siemens (SIE) and Thyssenkrupp (THK)) from 02. Jan. 2006 to 30. Sept. 2014 are used as data examples, which were downloaded from "Thomson Reuters" Corporation. The total number of trading days for the six German companies is 2226 and for the four French companies it is 2239. The trading times for all ten companies are from 9:00 to 17:30. We eliminate the data outside this time interval and also the data containing clerical errors. In addition to tick-by-tick data, we also employ 1-minute, 5-minute and 15-minute data, which are computed based on tick-by-tick data. Two realized measures, RV_0 and RK are considered. The IN algorithm is utilized to calculate RK obtained based on tick-by-tick, 1-minute and 5-minute returns. The DN algorithm is used to calculate tick-by-tick RK. We consider end effects in the computation of the RK with $m = 2$. The absolute difference for the ten European companies are less than 0.878% on average, which confirms the conclusion in Barndorff-Nielsen et al. (2009) that the end effect can be ignored in practice.

4.4.1 Implementation of algorithms

Figure 4.5 shows the processes of the IN and DN algorithms starting with different bandwidths for four selected days. In Figure 4.5, "u" and "l" represent the selected bandwidths in each iteration when a fixed upper and a lower starting bandwidths are used, respectively. And "z" stands for selected bandwidths in each iteration, when algorithms start with $\hat{H}_1$ estimated by Eq. (4.3.4). Figure 4.5 (a) displays the process of the IN algorithm based on tick-by-tick returns for THK on 07. May 2007. We choose 50 as the upper starting bandwidth and the selected bandwidths $\hat{H}_{s,j}$ in 3 iterations are 50, 14, 14, respectively. When 5 is chosen as the lower starting bandwidth, the selected bandwidths in 4 iterations are 5, 12, 14, 14, respectively. The calculated starting bandwidth $\hat{H}_1$ is 16 by using $\mathrm{RV}^s_{\mathrm{Z}}$ as the estimator of $\mathrm{IV}_{s,1}$, after 3 iterations ($\hat{H}_{s,j}$=16, 14, 14), the optimal bandwidth

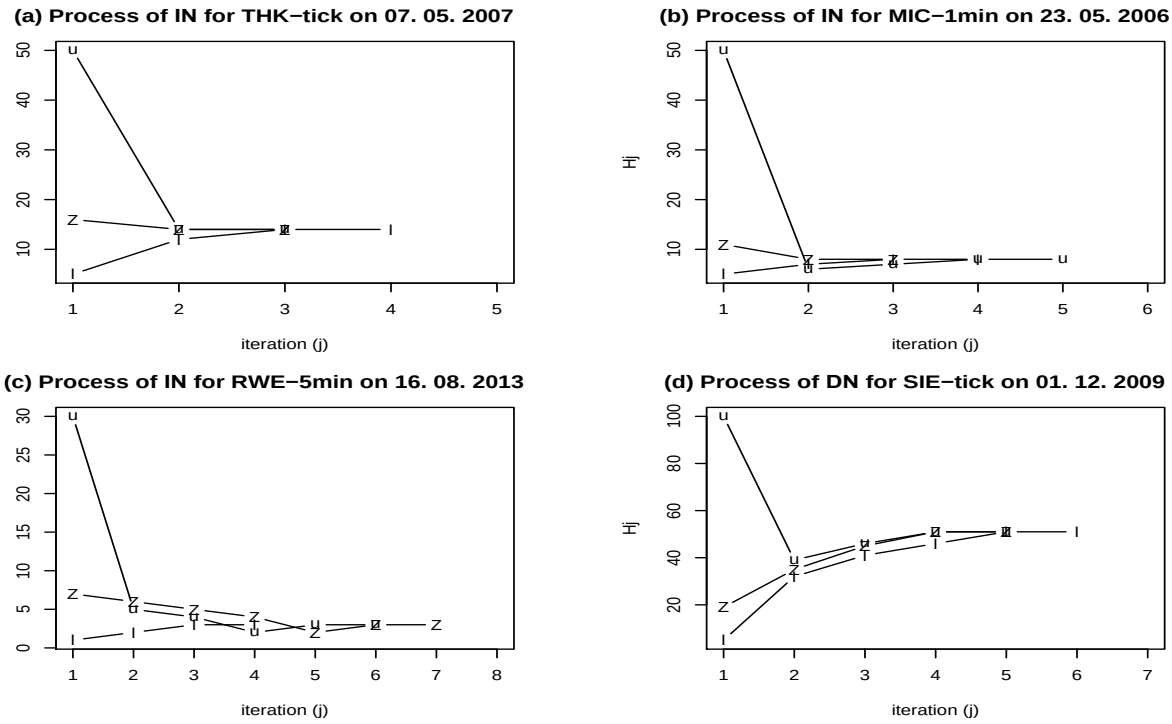

Figure 4.5: Process of algorithms for four selected days.

is also 14. Likewise, the processes of the IN algorithm based on 1-minute and 5-minute returns for MIC and RWE are plotted in Figure 4.5 (b) and (c), respectively. Figure 4.5 (d) plot the process of the DN algorithm based on tick-by-tick returns for SIE on 01. Dec. 2009. In Figure 4.5 (d) the upper starting bandwidth is 100, the selected bandwidths in 5 iterations are 100, 39, 46, 51, 51, respectively. We choose 5 as the lower starting bandwidth, the selected bandwidths in 6 iterations are 5, 32, 41, 46, 51, 51, respectively. The calculated starting bandwidth is 19 by using $\mathrm{RV}^s_{\mathrm{Z}}$ as the estimator of $\mathrm{IV}_{s,1}$, after 5 iterations ($\hat{H}_{s,j}$=19, 35, 45, 51, 51), $\hat{H}^{\mathrm{B}}$ is also equal to 51. From Figure 4.5 we can see that both algorithms work very well and only a few iterations are required. In addition, no matter what the starting bandwidths are, the selected bandwidths as of the second iteration are very close and that the final selected bandwidths are indeed the same.

Figure 4.6 displays the histograms of selected bandwidths and iteration numbers of RK for DBK. The histograms of $\hat{H}^{\mathrm{B}}$ calculated by the IN algorithm for tick-by-tick, 1-minute and 5-minute RK are plotted in Figures 4.6 (a), (b) and (c),

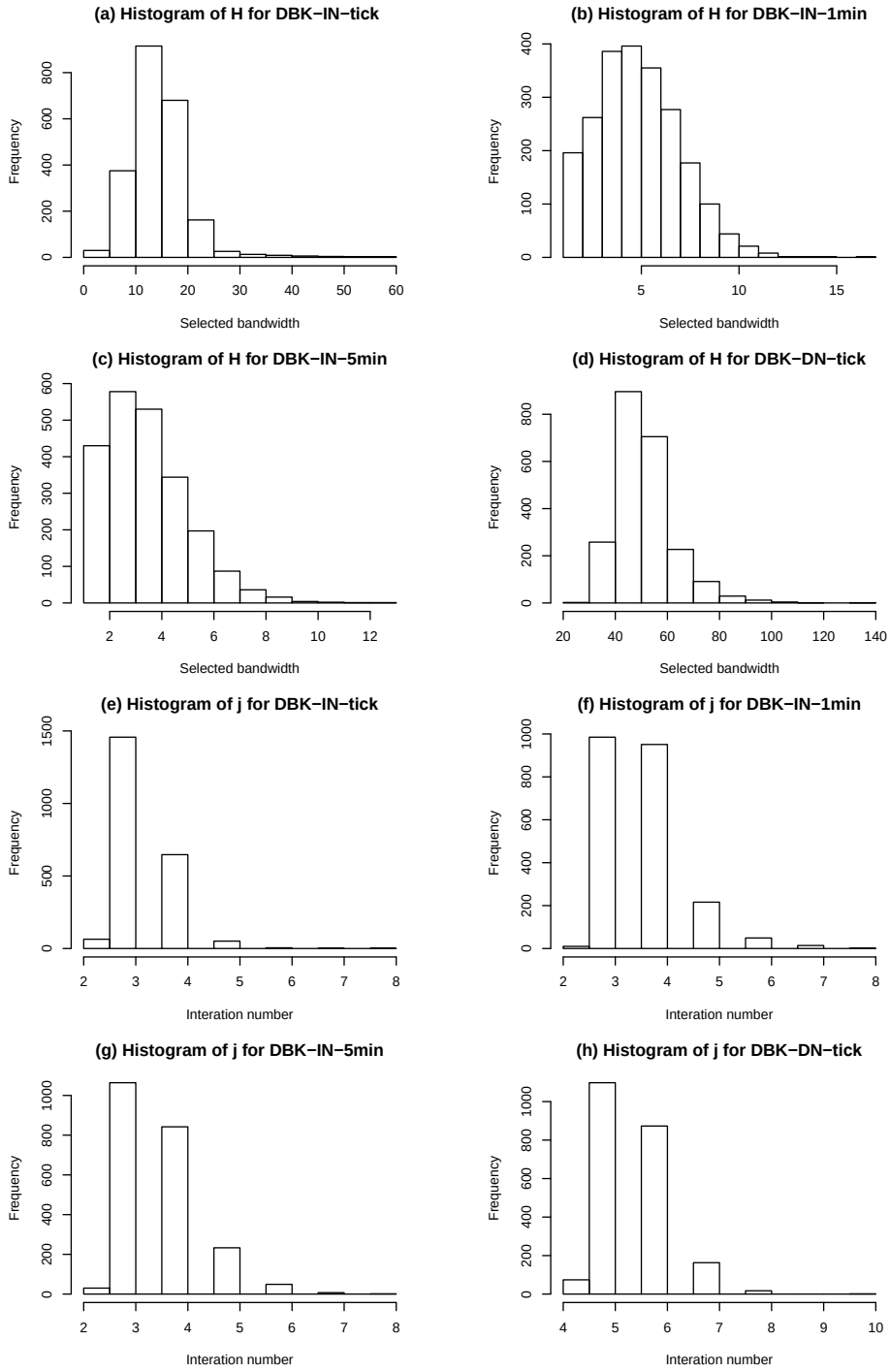

Figure 4.6: Histograms of selected bandwidth and interation number for Deutsche Bank.

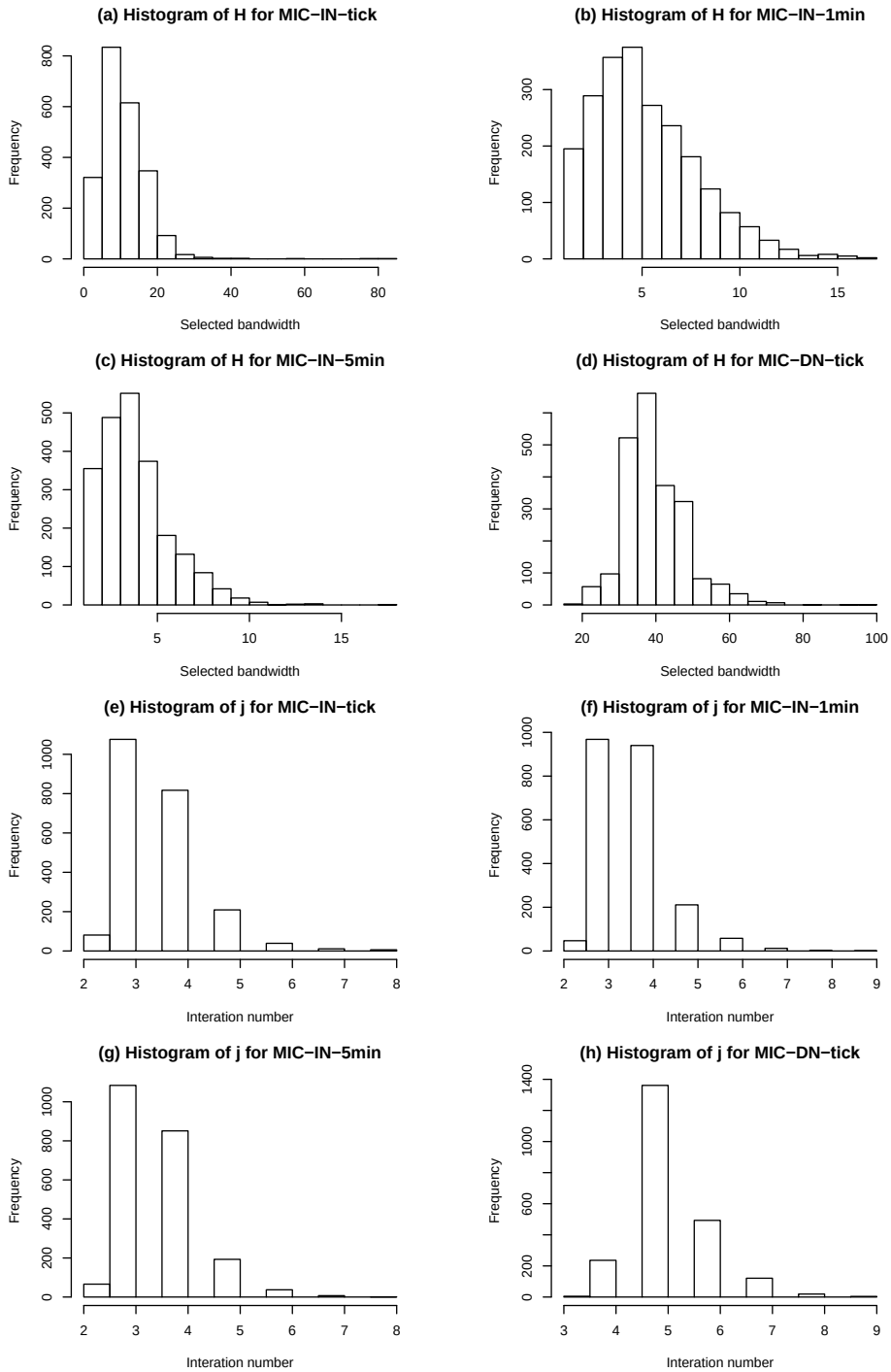

Figure 4.7: Histograms of selected bandwidth and interation number for Michelin.

respectively. Figure 4.6 (d) provides the histogram of $\hat{H}^{\mathrm{B}}$ for the tick-by-tick RK calculated by the DN algorithm (RK-DN-tick). The corresponding histograms of iteration numbers are displayed in Figures 4.6 (e), (f), (g) and (h), respectively. The selected bandwidths in Figure 4.6 (d) are bigger than those in Figure 4.6 (a), the reason is that the dependent MS noise is taken into account. The selected bandwidths for 1-minute RK in Figure 4.6 (b) are smaller than those for tick-by-tick RK, but larger than those for 5-minute RK. This corresponds to their ACFs results, that the correlations in 5-minute returns are smaller than the correlations in 1-minute and tick-by tick returns. In R code a maximal number of iterations $J = 20$. For the datasets under consideration this limit is never achieved. The commonly required number of iterations for DBK-IN-tick, DBK-IN-1min, DBK-IN-5min and DBK-DN-tick are 3, 3, 3, 5, respectively, which are also true for the other nine companies. These confirm that both IN and DN algorithms work very well in practice and only a few iterations are required. The illustrative results for MIC can be found in Figure 4.7. The histograms for the other eight companies show the same conclusions as in Figures 4.6 and 4.7, and are given in the appendix to chapter 4. Please note that the IN algorithm does not work for 15-minute RK of all the ten stocks, because on some days the number of observations is smaller than the selected bandwidths ($n < \hat{H}^{\mathrm{B}}$). For the same reason, 5-minute is not recommended using for calculating RK.

Table 4.2 lists the number of the three special cases ($\mathrm{RV_Z} < 0$, $\mathrm{RV_Z} > \mathrm{RV_0}$ and $\mathrm{RK} > \mathrm{RV_0}$) mentioned in Feng and Zhou (2015b) for RK-IN-tick, RK-IN-1min, RK-IN-5min and RK-DN-tick of the ten companies, respectively. We see that for all companies Case 1 occur rarely, however, Case 2 and Case 3 happen with a very big chance. For instance, Case 2 occur on 116 days for MIC-IN-tick and Case 3 occur on 226 days for SE-DN-tick. This shows us the necessity of adjusting Feng and Zhou (2015b)'s algorithm by replacing $\hat{\omega}^2_{s,1} = \frac{\mathrm{RV}^s_0 - \mathrm{RV}^s_{\mathrm{Z}}}{2n_s}$ with $\hat{\omega}^2_{s,1} = \frac{\mathrm{RV}^s_0}{2n_s}$. After this adjustment, the proposed IN and DN algorithms work automatically very well for all days. Meanwhile, we find that for all the companies Case 3 occur much more often for DN-tick than that for IN-tick. For instance, Case 3 in THK occur on 100 days for DN-tick and only on 26 days for IN-tick. This difference shows

that the obtained RK results change obviously if the dependent noise takes into account. In addition, we see that Case 2 and Case 3 occur much more often when the sampling frequencies decline.

Table 4.3 lists the mean and standard deviations of RV_0 and RK obtained based on tick-by-tick, 1-minute, 5-minute and 15-minute returns for the ten companies over the whole period, respectively. The mean values and standard deviations for tick-by-tick RV_0 are larger than any other realized estimators in Table 4.3. The mean values of RV_0 diminish with the decline of sampling frequency. The standard deviations of 1-minute, 5-minute and 15-minute RV_0 for all companies are much smaller than those of tick-by-tick RV_0. As mentioned before, the algorithm IN does not work for 15-minute RK of all companies, because on some days $n < \hat{H}^{\mathrm{B}}$. Therefore, the results are not listed in Table 4.3. For the same reason, 5-minute returns are not recommended using to calculate RK. In most cases, the standard deviations of RK-DN-tick are smaller than those of RK-IN-tick and of RK-IN-1min. The mean values of RK-DN-tick lie between the mean values of RK-IN-tick and RK-IN-1min. The differences between the means of RK-DN-tick and RK-IN-tick indicate the bias cased by estimated dependent MS noise.

Table 4.2: The number of different cases for the ten companies

Firm \ Cases		**IN-tick**	**IN-1min**	**IN-5min**	**DN-tick**
AF	$RV_Z < 0$	0	2	1	0
	$RV_Z > RV_0$	2	308	698	2
	$RK > RV_0$	2	332	748	65
ALV	$RV_Z < 0$	3	0	1	3
	$RV_Z > RV_0$	1	234	683	1
	$RK > RV_0$	0	244	717	10
BMW	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	43	463	748	43
	$RK > RV_0$	52	477	810	156
DBK	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	2	436	818	2
	$RK > RV_0$	5	467	877	21
MIC	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	116	499	748	116
	$RK > RV_0$	133	505	805	356
PSA	$RV_Z < 0$	1	0	1	1
	$RV_Z > RV_0$	6	431	725	6
	$RK > RV_0$	11	437	786	110
RWE	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	25	334	662	25
	$RK > RV_0$	33	336	729	74
SE	$RV_Z < 0$	1	1	1	1
	$RV_Z > RV_0$	68	375	697	68
	$RK > RV_0$	77	391	753	226
SIE	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	20	345	750	20
	$RK > RV_0$	21	344	802	29
THK	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	22	393	720	22
	$RK > RV_0$	26	403	779	100

Table 4.3: Mean ($*10^4$) and standard deviation ($*10^4$) of RV_0 and RK based on the different sampling frequencies for the ten companies

		AF	AIV	BMW	DBK	MIC	PSA	RWE	SE	SIE	THK
RV_0-tick	mean	11.253	8.289	6.323	17.031	8.059	11.861	5.051	7.491	6.092	7.848
	s.d.	13.077	20.982	10.094	143.075	13.343	13.382	7.559	13.962	14.040	12.776
RV_0-1min	mean	7.042	4.744	4.653	6.219	5.534	8.135	3.396	4.693	3.694	5.592
	s.d.	6.274	9.902	5.951	11.649	6.239	7.646	4.757	6.561	6.996	7.086
RV_0-5min	mean	5.949	4.188	4.303	5.679	4.804	7.171	2.976	3.839	3.324	4.918
	s.d.	5.710	9.479	6.221	10.227	5.651	7.302	4.530	4.957	6.910	6.279
RV_0-15min	mean	5.458	3.988	4.105	5.544	4.565	6.748	2.805	3.526	3.169	4.645
	s.d.	5.663	9.558	6.172	11.019	5.608	7.695	4.884	4.400	7.132	5.869
RK-IN-tick	mean	5.858	4.290	4.293	5.811	4.751	7.119	3.042	4.119	3.468	5.203
	s.d.	5.748	8.804	5.833	9.533	5.695	7.074	4.137	6.357	6.454	7.492
RK-IN-1min	mean	5.518	3.995	4.091	5.472	4.562	6.843	2.783	3.573	3.211	4.688
	s.d.	5.169	8.999	5.786	10.065	5.371	6.895	3.859	4.796	7.130	5.893
RK-IN-5min	mean	4.836	3.679	3.733	4.977	4.114	6.063	2.472	3.177	2.901	4.243
	s.d.	4.914	9.153	5.847	9.191	5.185	6.833	3.750	4.214	6.895	5.705
RK-DN-tick	mean	5.579	4.106	4.194	5.533	4.701	6.980	2.864	3.876	3.328	4.830
	s.d.	5.165	8.477	5.579	9.586	5.508	6.916	3.984	5.608	6.377	6.074

4.4.2 Comparison of realized estimators based on Value-at-Risk

Further comparison of these realized estimators is carried out by assessing their performances in the computation of Value-at-Risk (VaR). Financial risk managers often report the risk of investments using the concept of VaR, which estimates the maximum loss at given confidence interval in a certain period. VaR is widely used by investors and regulators in the financial industry to measure the amount of assets needed to cover possible losses and has been considered as a expectable banking risk measure. The Basel Committee demands banks use the VaR in establishing the minimum capital necessary at investments in order to reduce the fragility of international active banks. After its proposal it becomes a popular risk assessment tool in financial service firms. In the literature, Beltratti and Morana (2005) as well as Giot and Laurent (2004) investigated the performances of VaR measures based on the GARCH and lnRV-FARIMA models. Their results show that the lnRV-FARIMA class provide a superior performance by computing VaR. Feng and Zhou (2015a) showed the logarithmic RK can be well described by a long-memory process and may also exhibit a deterministic nonparametric trend. Hence, in our framework we compute the one step ahead VaR by means of the semiparametric FARIMA model (SEMIFAR also called the Semi-FI-Log-ACD model in Feng and Zhou, 2015a) based on the different logarithmic realized estimators. The VaR calculated based on the the SEMIFAR model at a given confidence interval $(1-\alpha)$ is obtained by

$$\mathrm{VaR}_{1-\alpha}^{t} = s(\tau_t)\sigma_t Z_{1-\alpha}, \tag{4.4.1}$$

where $s(\tau_t)$ is the local variance, σ_t is the standard deviation (volatility) of investments at time t and $Z_{1-\alpha}$ is loss distribution for the corresponding N(0,1) quantile. In the first step we use the intraday returns of different sampling frequencies to compute the realized estimators mentioned in section 4.1. In the second step we estimate the conditional mean using the SEMIFAR model based on the logarithmic realized estimators. The estimated total means in the original data $s(\tau_t)\sigma_t$ is

obtained through the exponential transformation of these estimated deterministic trend and the estimated conditional mean.

A wealth of literature focuses on measuring the quality of the VaR calculations. Backtesting is a statistical procedure, where the loss forecast calculated by VaR is compared with the actual losses. Kupiec (1995) proposed a basic tests to examine the frequency of losses in excess of VaR. By means of a simple backtesting Peitz (2016) compared VaR calculations based on the parametric with semiparametric models. In his proposal the observed amount of exceptions (Points over VaR) is compared to the expected amount/benchmark $n * \alpha$. We utilize the simple backtesting in Peitz's work. Let $\alpha = 5\%$, n is 2226 for the six German companies and 2239 for the four French companies. The corresponding benchmarks for all ten companies are 111. Figure 4.8 to Figure 4.17 show the one day dynamic ahead 95% VaR based on the lnRV-SEMIFAR model for the ten example companies from 02. Jan. 2006 to 30. Sept. 2014. The loss distribution is assumed normal. In each figure the black line shows the loss/negative returns and the red line depicts the 95% VaR values by means of the SEMIFAR model based on the different logarithmic realized estimators. Table 4.4 lists the corresponding numerical results of points over VaR and the deviations from the benchmark for Figure 4.8 to Figure 4.17. Table 4.5 summarizes the total absolute deviation and total deviation from the benchmark of 95% VaR based on the lnRV-SEMIFAR model. In the empirical implementation, RK-IN-tick with the smallest total absolute deviation 121 is the best estimator in the computation of VaR. RK-DN-tick is the second best estimator with the total absolute deviation 145. This confirms that the independent MS noise assumption is more suitable on the most trading days in our data examples. From Figure 4.1 to 4.4 we can also see that in most cases the ACFs at lag 1 are clearly significant and the results imply that some at higher lags slightly significant ACFs occur only occasionally. Meanwhile, the difference between the total absolute deviation of RK-IN-tick and that of RK-DN-tick is not big. Both can be considered as the good volatility estimators. In addition, the worst realized estimator is RV_0-tick. Its deviations for all ten companies are negative, the corresponding total deviation is -568. From the figures (a) of

all ten companies we can see that the red lines are far over the daily losses, which confirms the numerical result that the 95% VaR based on the ln(RV_0-tick)-SEMIFAR seriously overestimate the financial risk. Next overestimated realized estimator is RV_0-1min with the total deviation -167. The deviations of RK-IN-5min for all ten companies are however positive, which means the underestimation of the financial risk. This fact can be confirmed in the figures (g) of all ten companies, where a few black lines are outside the red lines. The same results are also obtained for RK-IN-1min, RV_0-5min, RV_0-15min, based on them the financial risk can be seriously underestimated. Moreover, the performance of RV_0-15min is worse than that of RV_0-1min, while the performance of RK-IN-5min is worse than that of RK-IN-1min and RK-IN-tick. This shows again the necessity of choosing a realized estimator together with a suitable frequency.

In summary, after comparing the performances of different realized estimators in the computation of VaR based on the SEMIFAR model we find that the performances of RK-IN-tick and RK-DN-tick are better than any other realized estimator and are hence recommended using as the estimators of IV in practice.

4.5 Final remarks

In this chapter we improved the IPI algorithms for RK under independent and dependent noise assumptions. End effects are considered for both algorithms. The data from ten European companies of 9 years are used as data examples. We apply the IN algorithm to tick-by-tick, 1-minute, 5-minute and 15-minute returns and apply the DN algorithm to tick-by-tick returns. Both algorithms are fully automatic and now work very well for all data examples. Furthermore, we compare these RK estimators with RV_0 calculated by tick-by-tick, 1-minute, 5-minute and 15-minute returns. The IN algorithm does not work well for the data-frequencies over 5-minute, because on same days the number of observations is smaller than the selected bandwidths. Further comparison of these realized measures is carried out by assessing their performances in the computation of VaR based on the SEMIFAR model. It is found that RK-IN-tick and RK-DN-

tick estimators have good performances and are hence recommended using as the estimators of IV in practice. However, a modified approach to rate these realized estimators should be proposed by means of comparing their mean squared error (MSE) values. This is also one of our future research directions.

Acknowledgments: I would like to express my deepest gratitude to Prof. Dr. Yuanhua Feng, who gives me some useful suggestions to improve the quality of this chapter.

Table 4.4: "Points over 95% VaR" (benchmark: 111) by means of the SEMIFAR model based on the different logarithmic realized measures for the ten companies

		AF	ALV	BMW	DBK	MIC	PSA	RWE	SE	SIE	THK
RV_0-tick	PoV	53	41	53	52	62	61	47	57	38	78
	deviation	-58	-70	-58	-59	-49	-50	-64	-54	-73	-33
RV_0-1min	PoV	92	96	80	102	89	111	88	102	81	102
	deviation	-19	-15	-31	-9	-22	0	-23	-9	-30	-9
RV_0-5min	PoV	214	134	107	136	122	147	115	144	111	138
	deviation	+103	+23	-4	+25	+11	+36	+4	+33	0	+27
RV_0-15min	PoV	157	138	107	131	126	153	123	146	116	153
	deviation	+46	+27	-4	+20	+15	+42	+12	+35	+5	+42
RK-IN-tick	PoV	138	109	86	110	112	131	101	117	84	113
	deviation	+27	-2	-25	-1	+1	+20	-10	+6	-27	+2
RK-IN-1min	PoV	148	137	103	130	116	144	113	139	110	141
	deviation	+37	+26	-8	+19	+5	+33	+2	+28	-1	+30
RK-IN-5min	PoV	187	158	127	154	152	176	147	176	139	169
	deviation	+76	+47	+16	+43	+41	+65	+36	+65	+28	+58
RK-DN-tick	PoV	143	122	87	119	112	131	110	124	92	127
	deviation	+32	+11	-24	+8	+1	+20	-1	+13	-19	+16

Table 4.5: Summary of "Points over 95% VaR" based on the lnRV-SEMIFAR

Realized measure	**Total absolute deviation**	**Total deviation**
RK-IN-tick	121	-9
RK-DN-tick	145	57
RV_0-1min	167	-167
RK-IN-1min	189	171
RV_0-15min	248	240
RV_0-5min	266	258
RK-IN-5min	475	475
RV_0-tick	568	-568

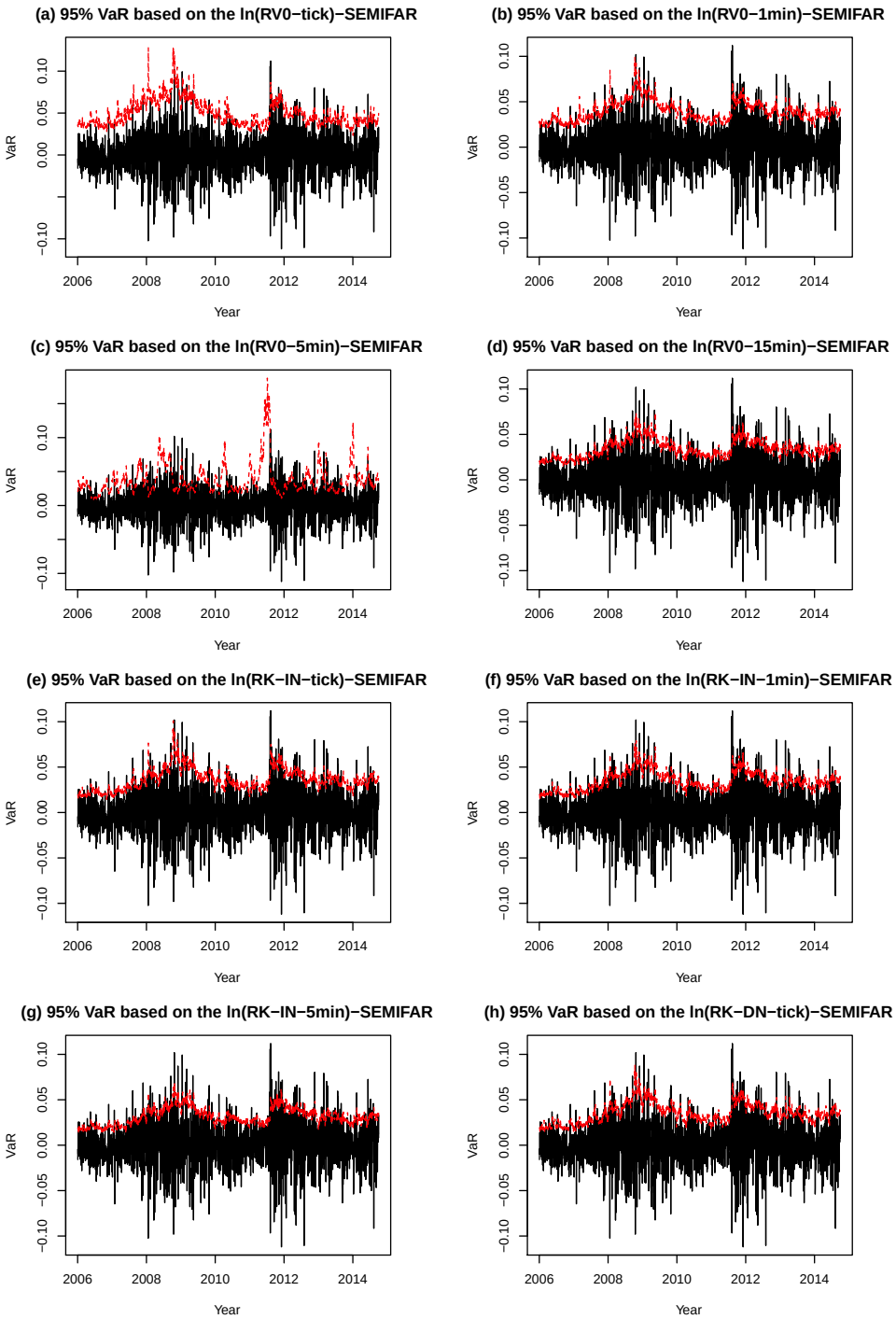

Figure 4.8: Loss and 95% VaR based on the lnRV-SEMIFAR model for Air France.

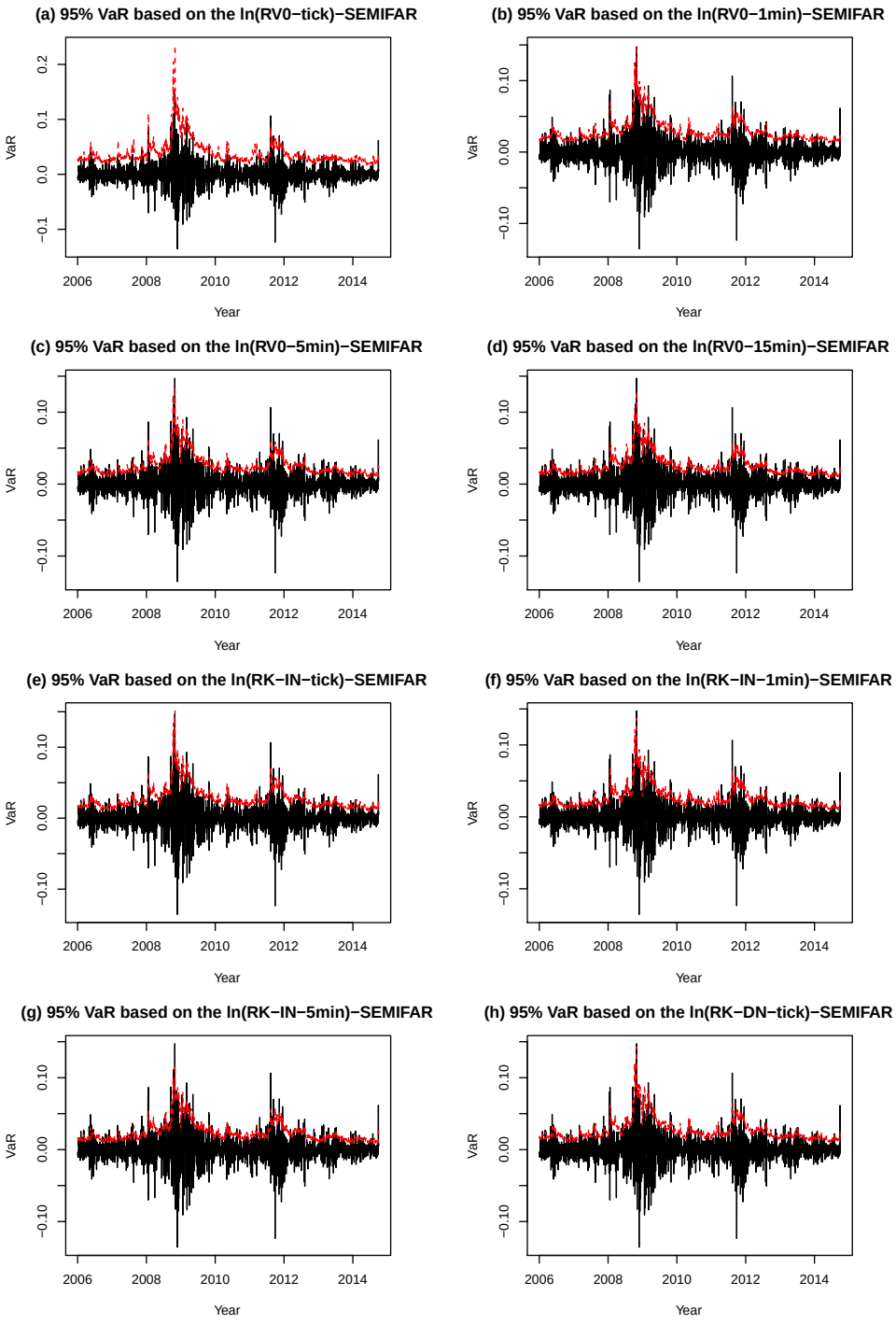

Figure 4.9: Loss and 95% VaR based on the lnRV-SEMIFAR model for Allianz.

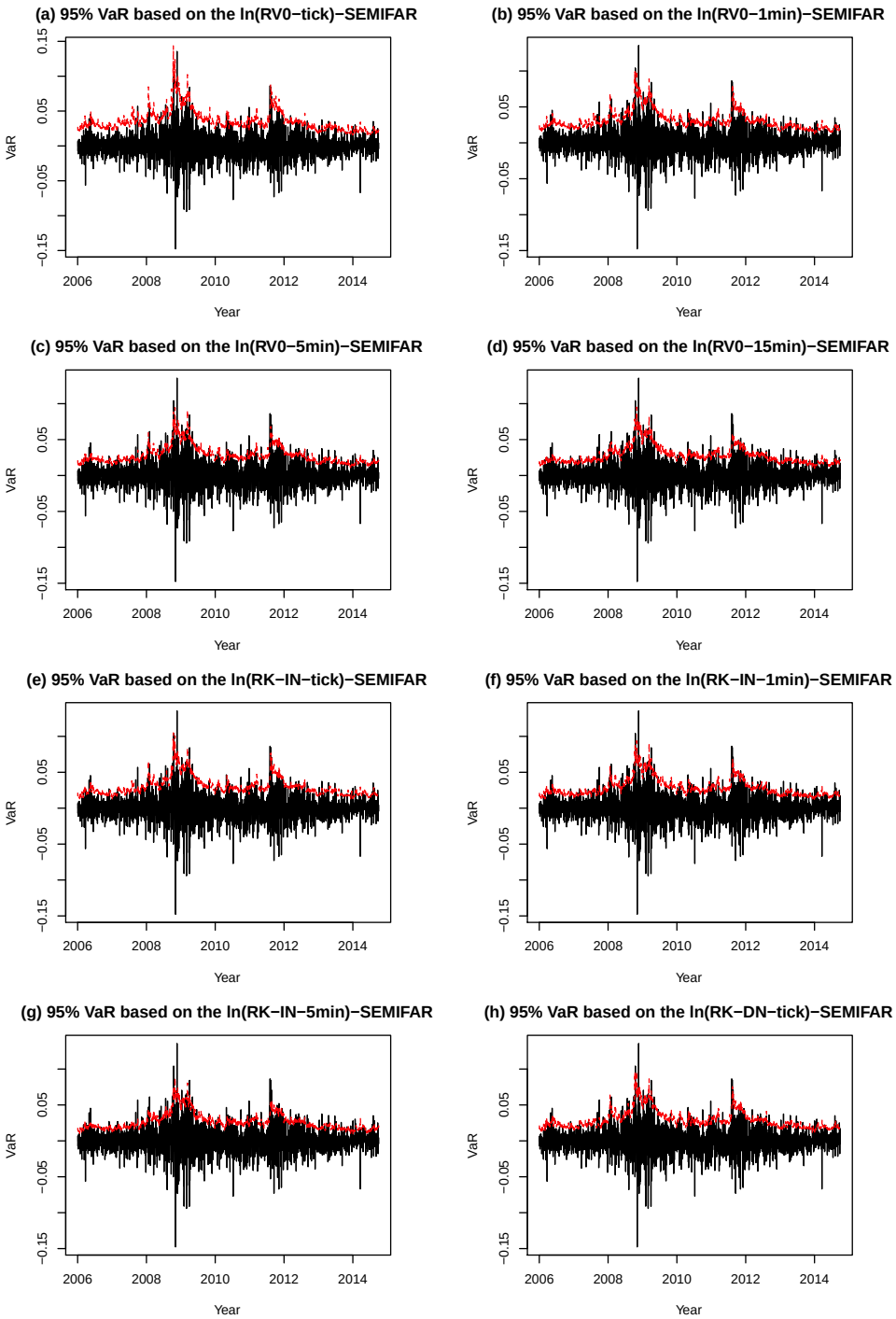

Figure 4.10: Loss and 95% VaR based on the lnRV-SEMIFAR model for BMW.

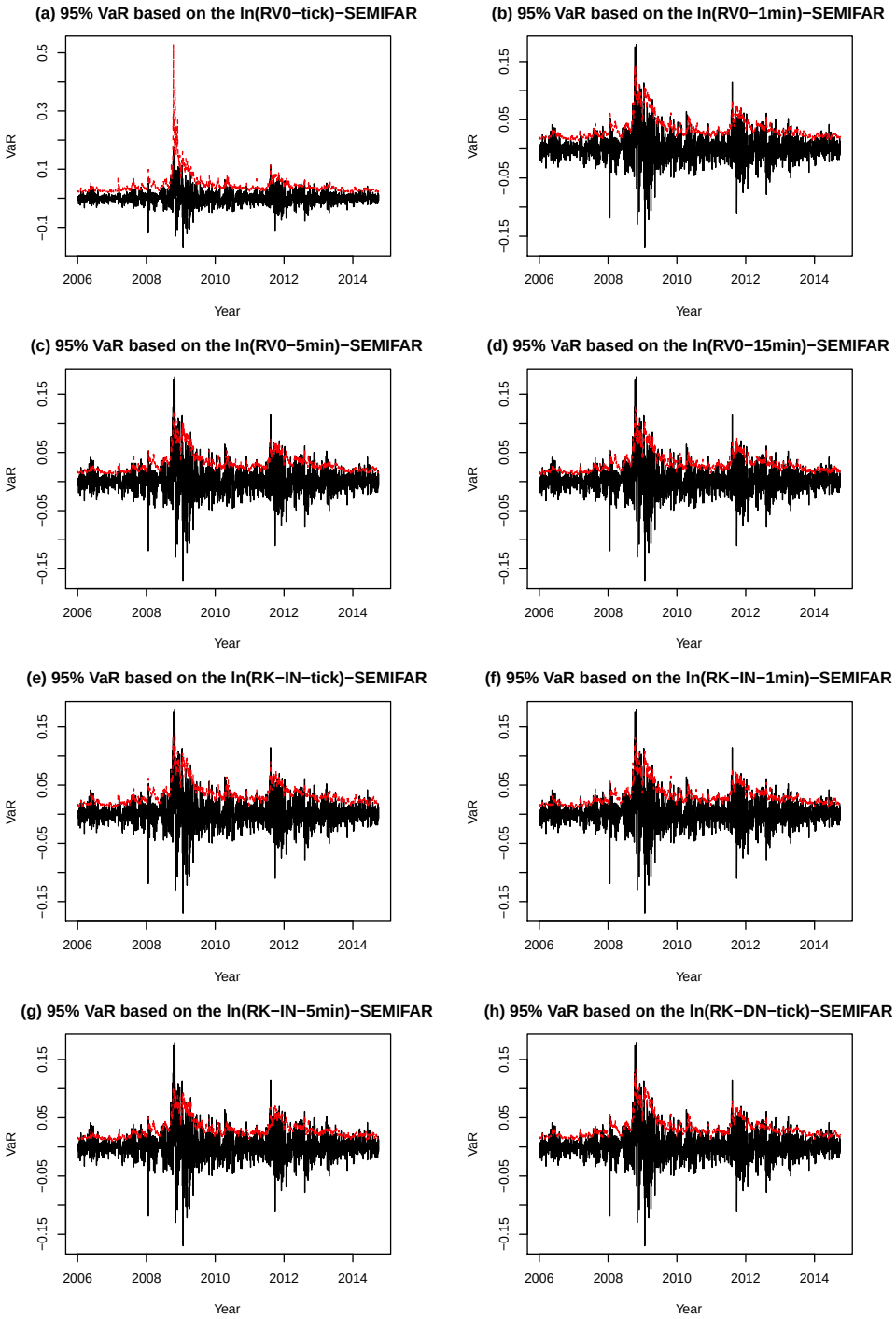

Figure 4.11: Loss and 95% VaR based on the lnRV-SEMIFAR model for Deutsche Bank.

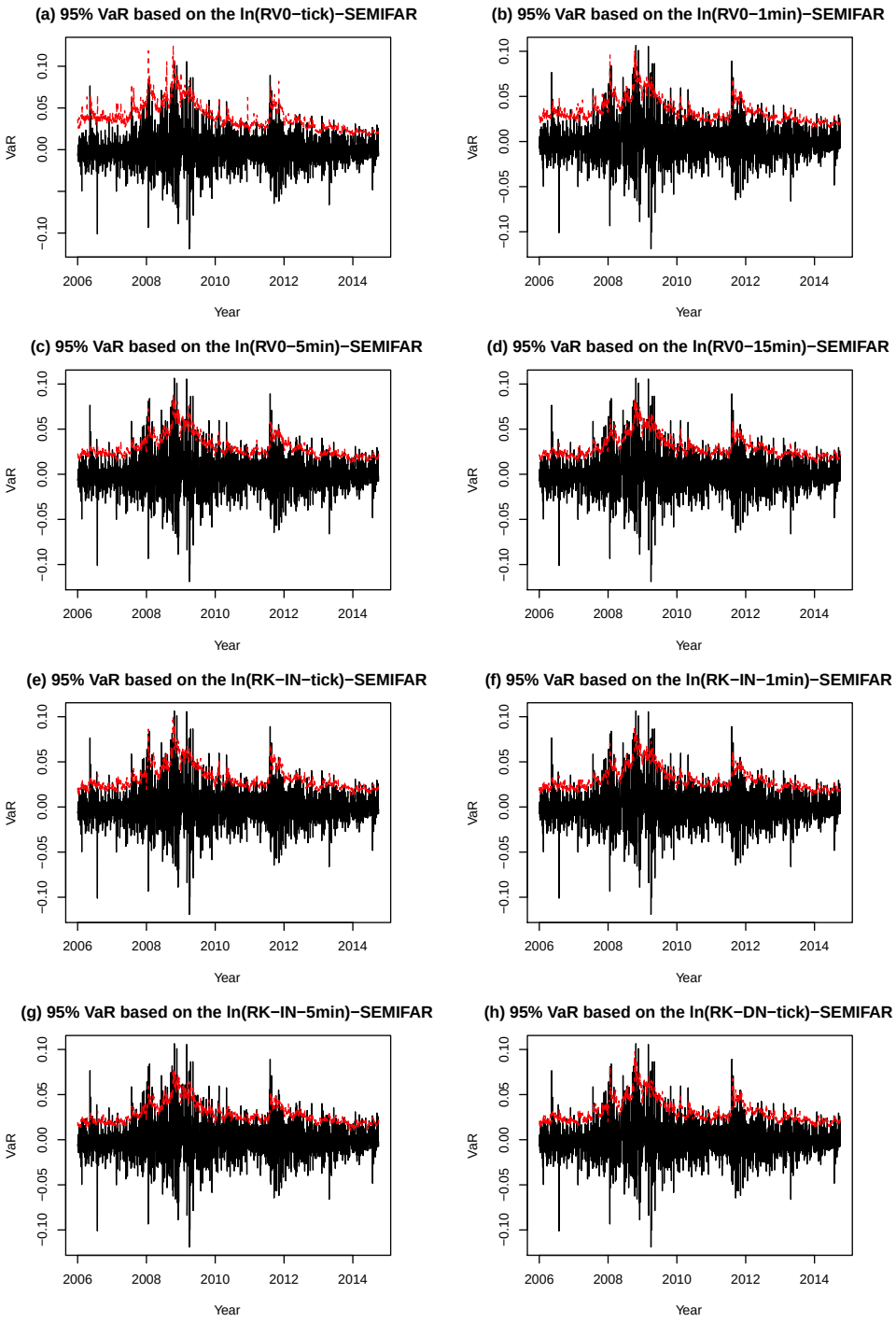

Figure 4.12: Loss and 95% VaR based on the lnRV-SEMIFAR model for Michelin.

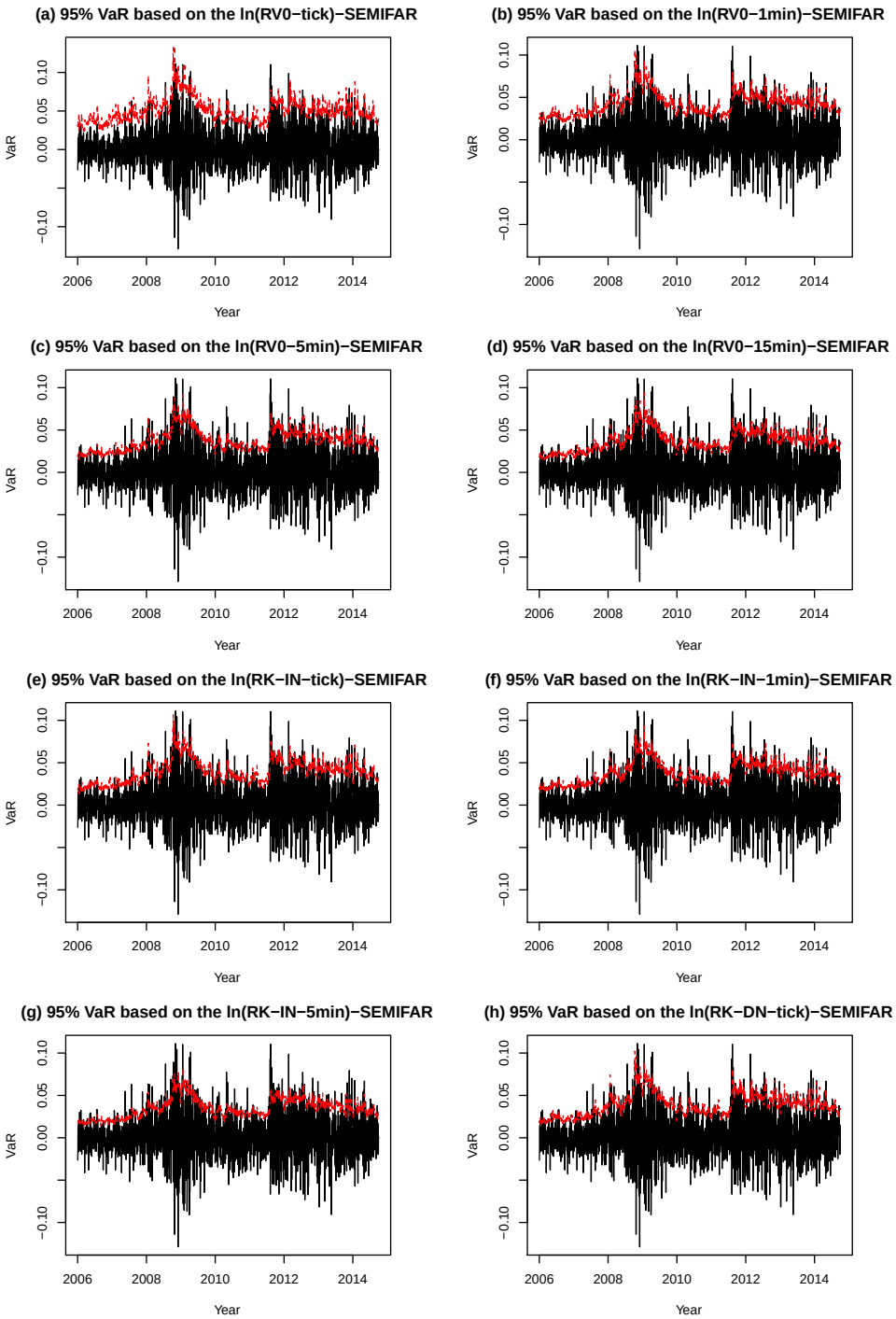

Figure 4.13: Loss and 95% VaR based on the lnRV-SEMIFAR model for Peugeot.

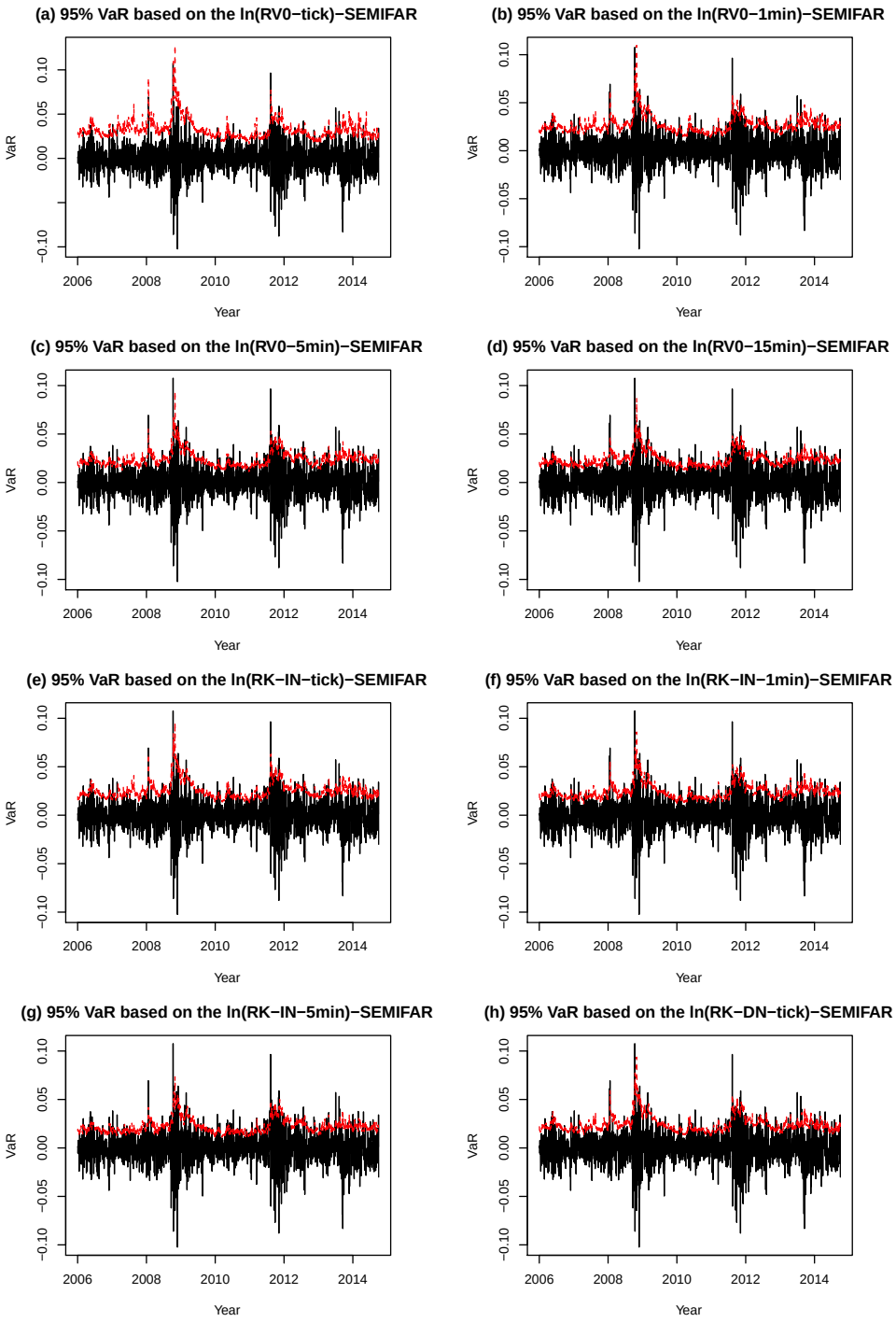

Figure 4.14: Loss and 95% VaR based on the lnRV-SEMIFAR model for RWE.

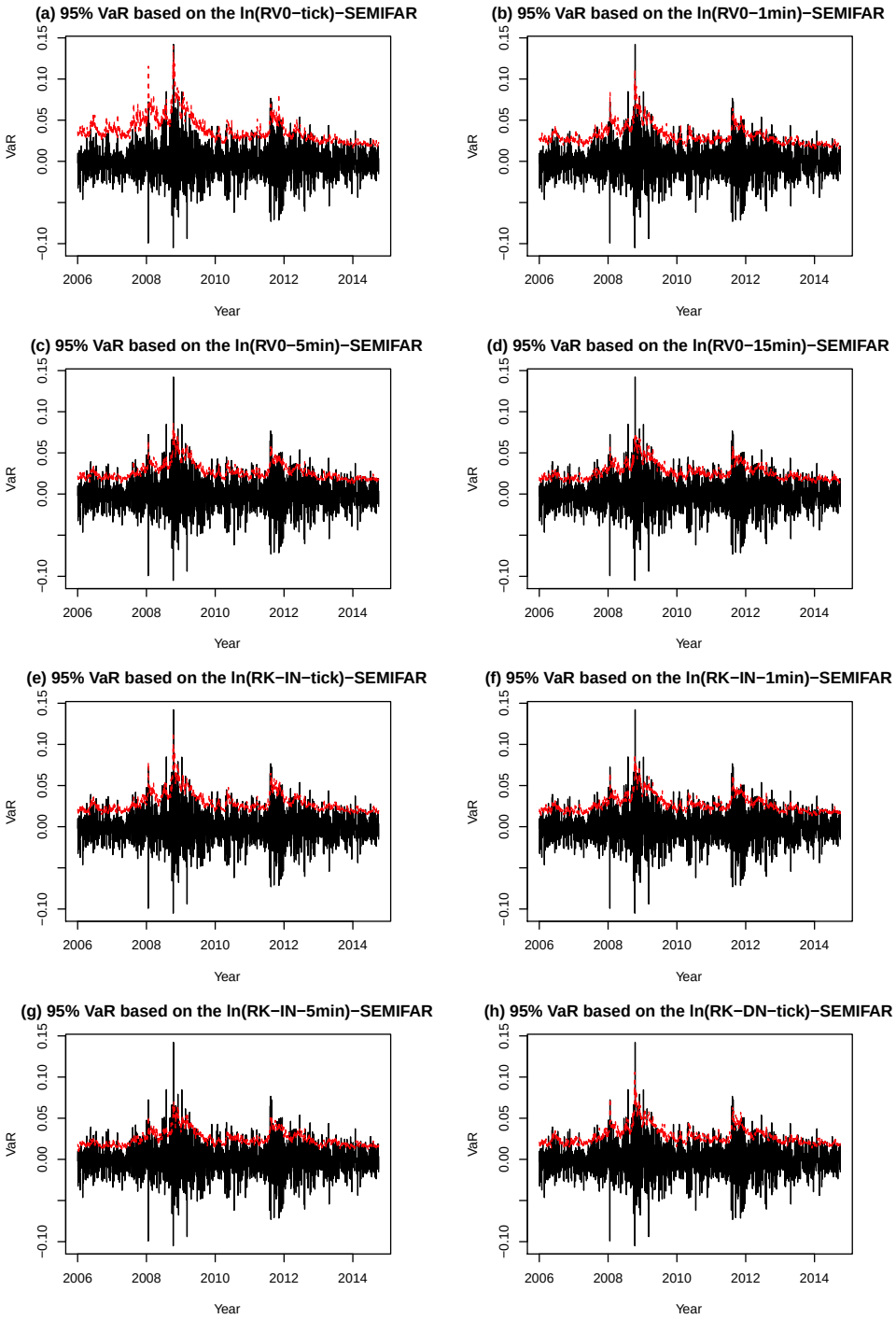

Figure 4.15: Loss and 95% VaR based on the lnRV-SEMIFAR model for Schneider Electric.

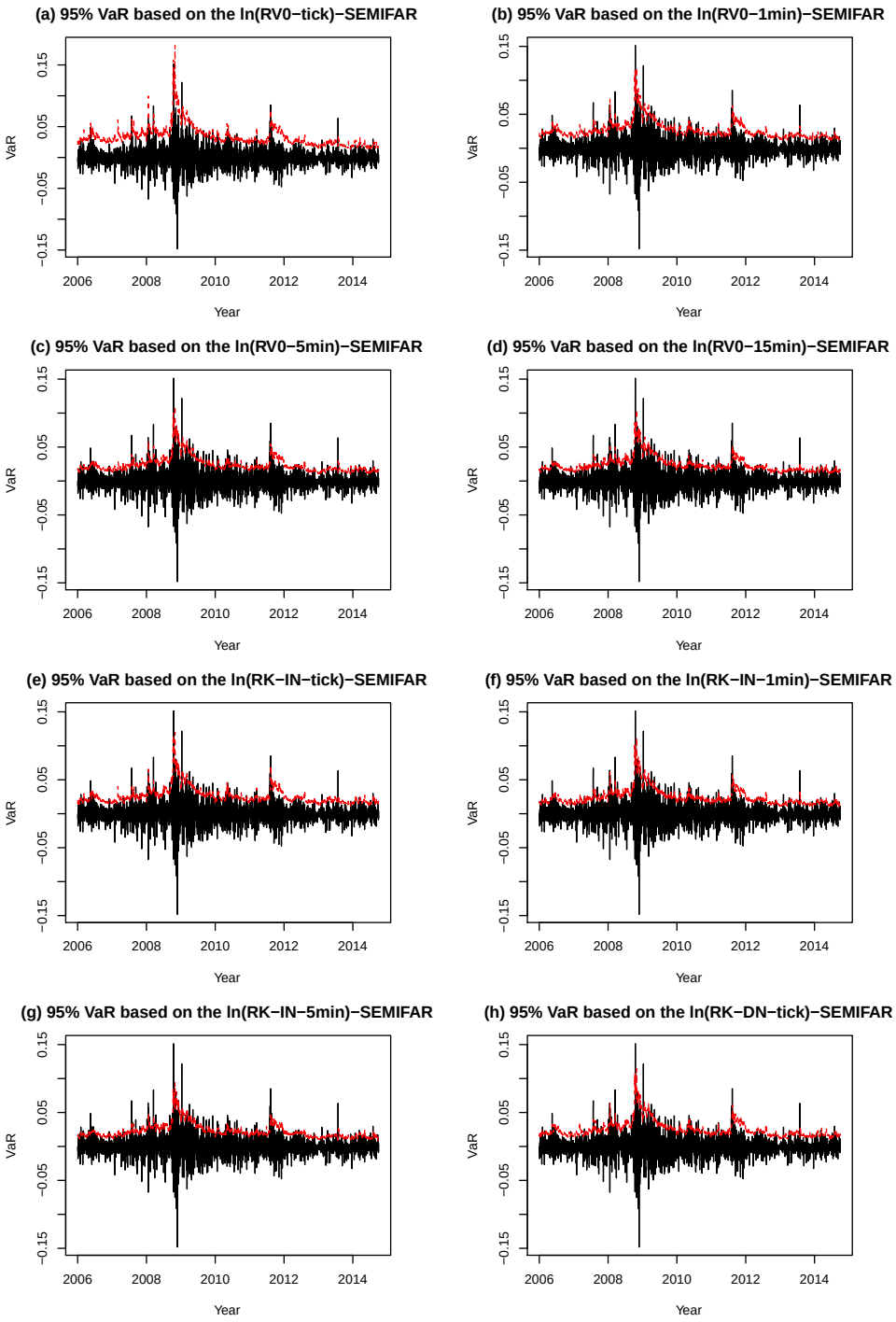

Figure 4.16: Loss and 95% VaR based on the lnRV-SEMIFAR model for Siemens.

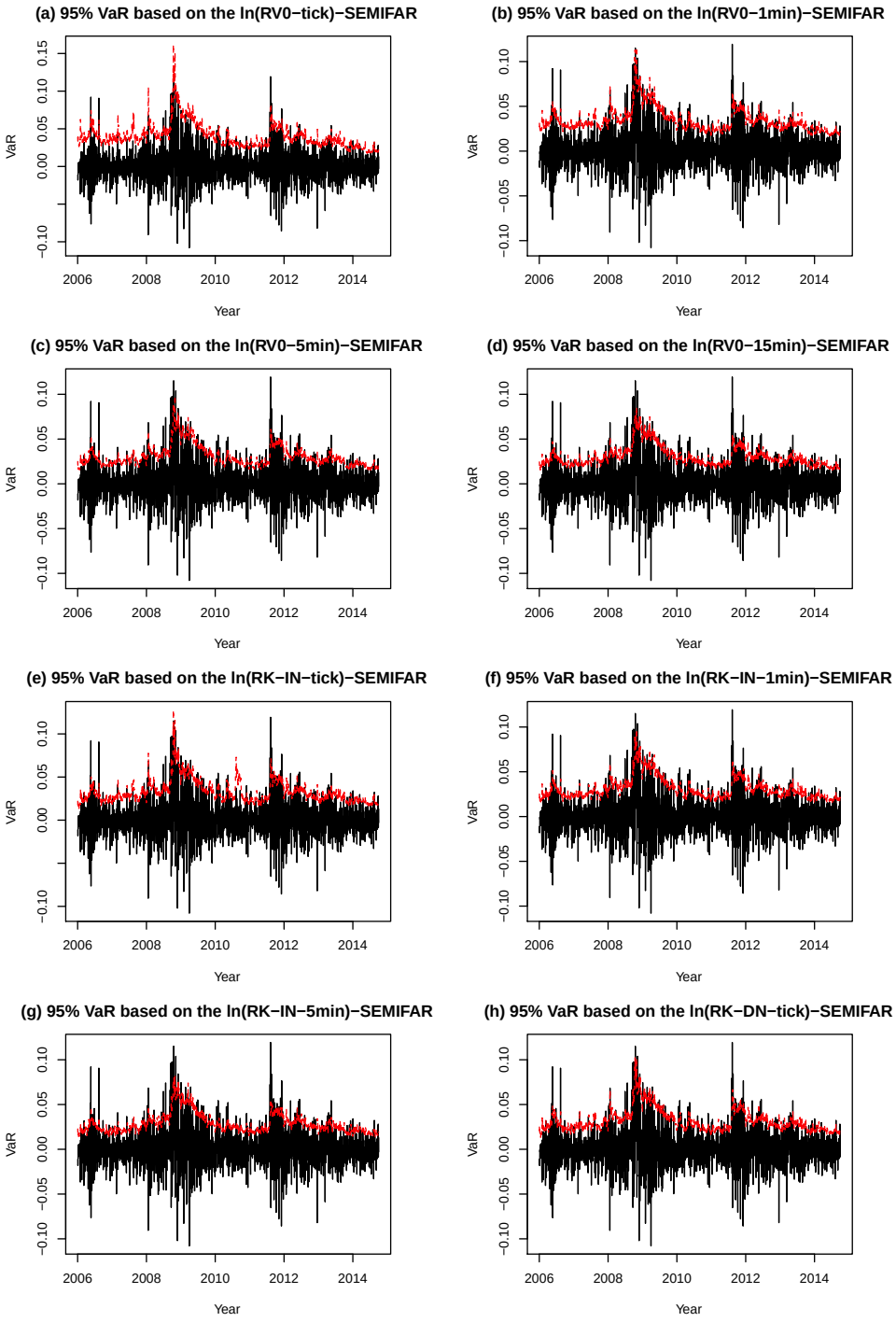

Figure 4.17: Loss and 95% VaR based on the lnRV-SEMIFAR model for Thyssenkrupp.

Appendix to Chapter 4

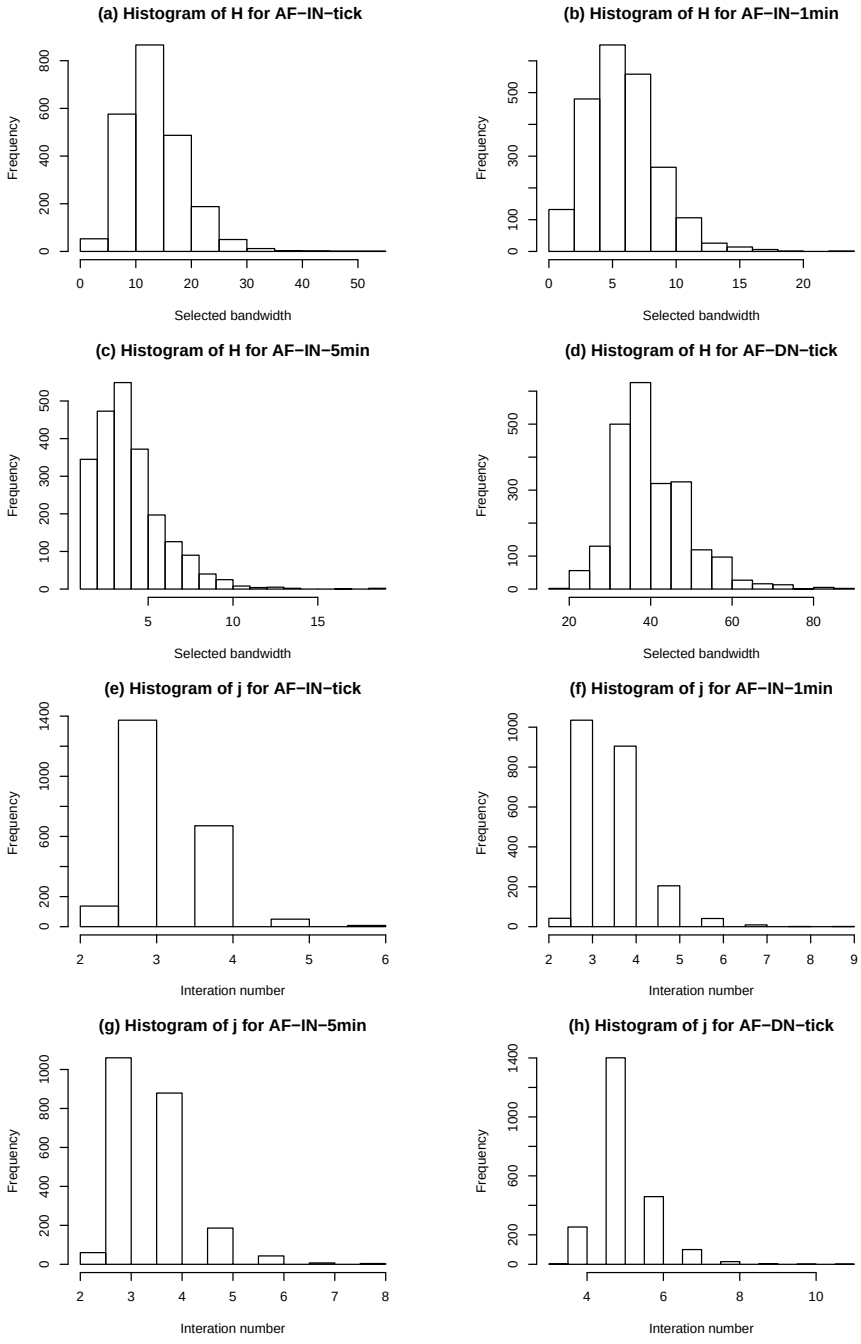

Figure A.4.1: Histograms of selected bandwidth and interation number for Air France.

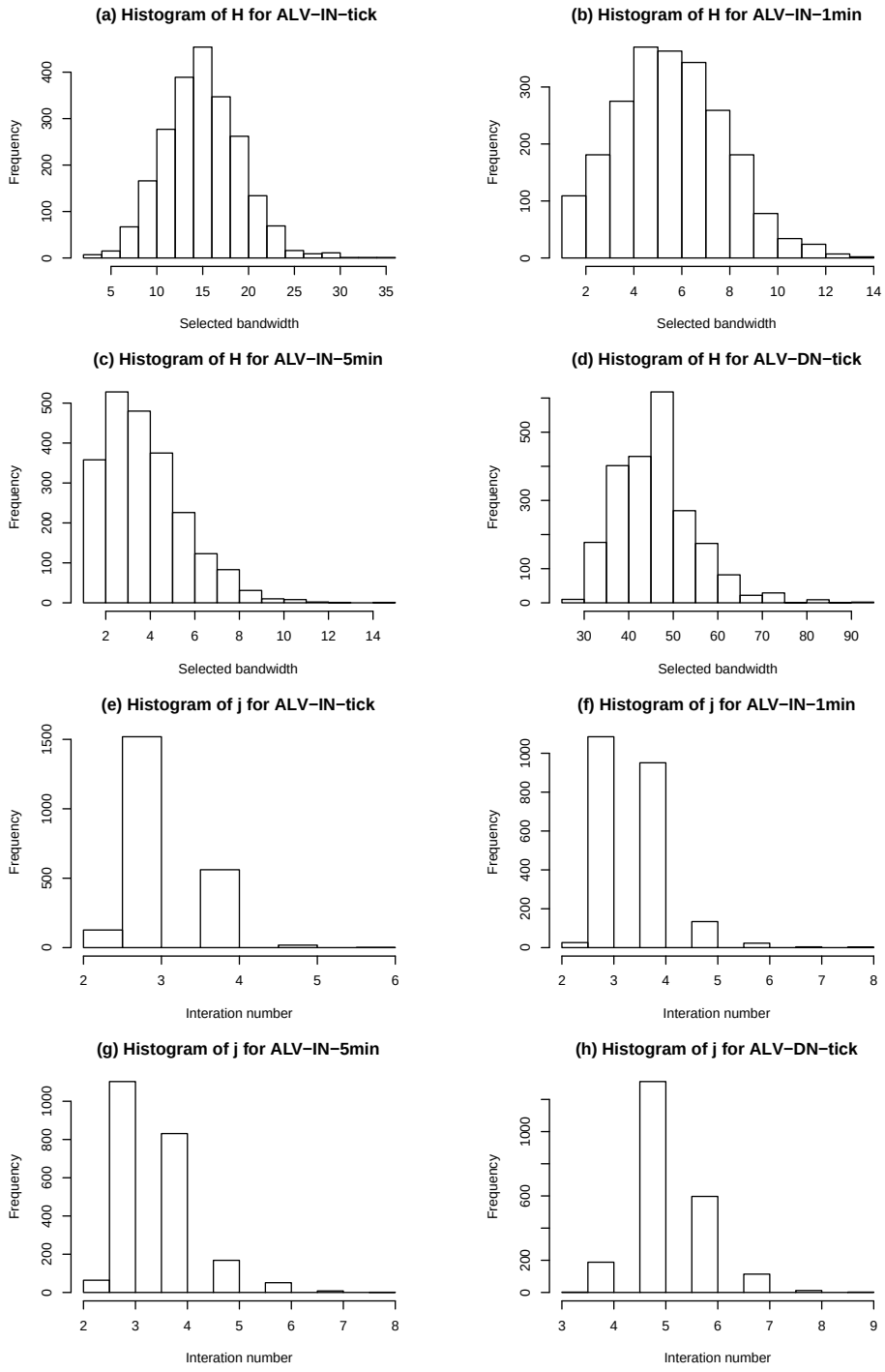

Figure A.4.2: Histograms of selected bandwidth and interation number for Allianz.

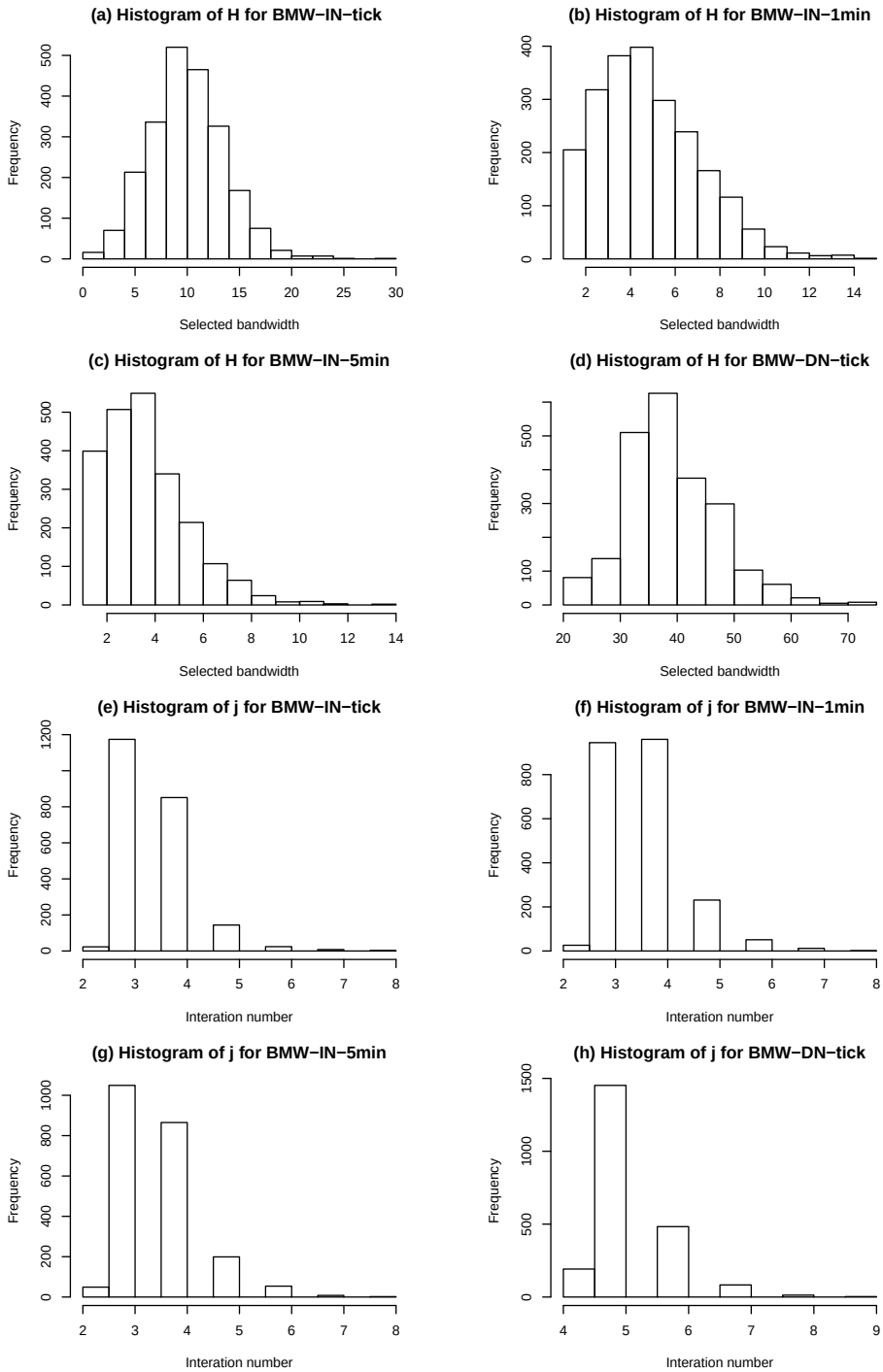

Figure A.4.3: Histograms of selected bandwidth and interation number for BMW.

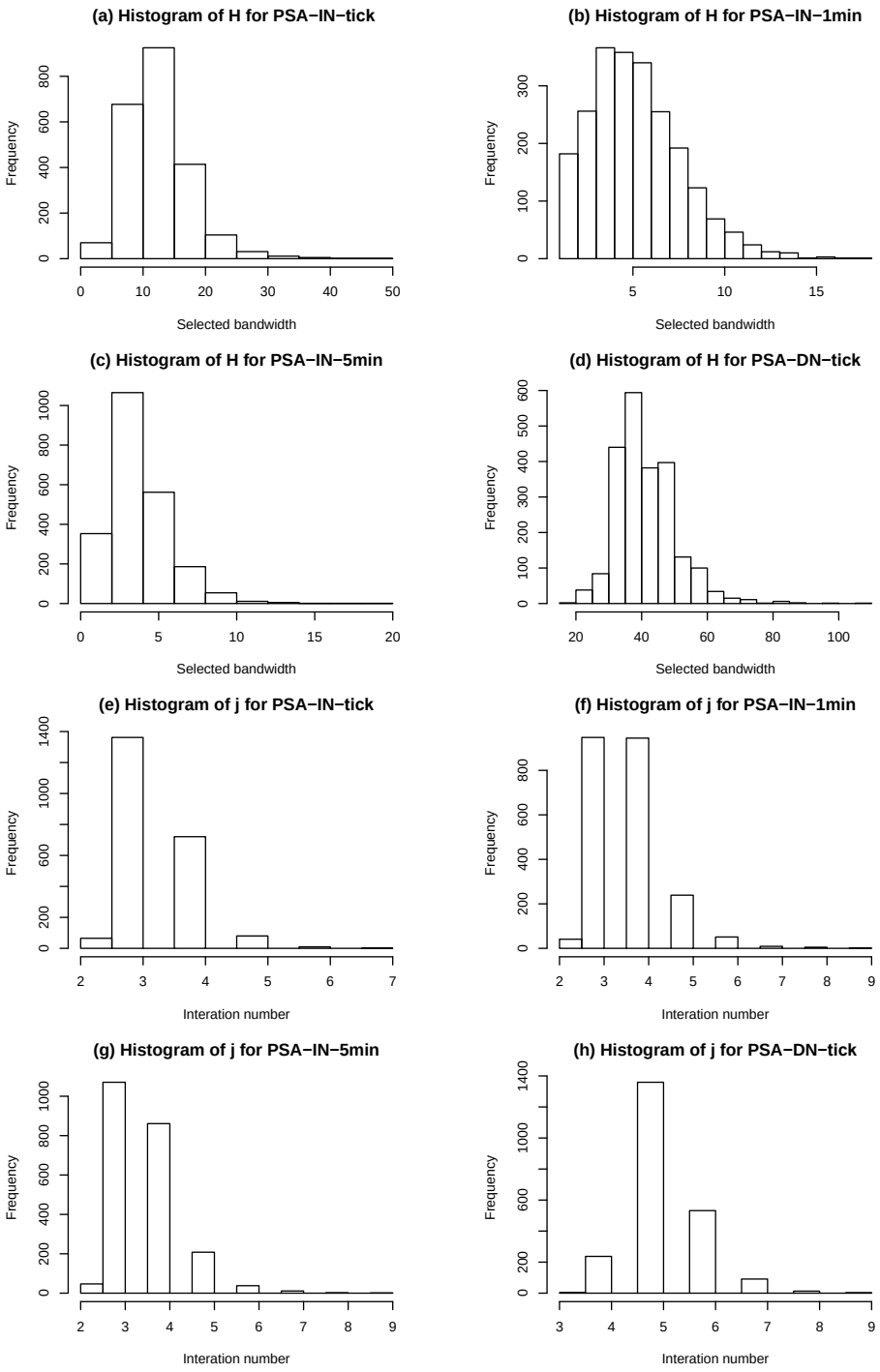

Figure A.4.4: Histograms of selected bandwidth and interation number for Peugeot.

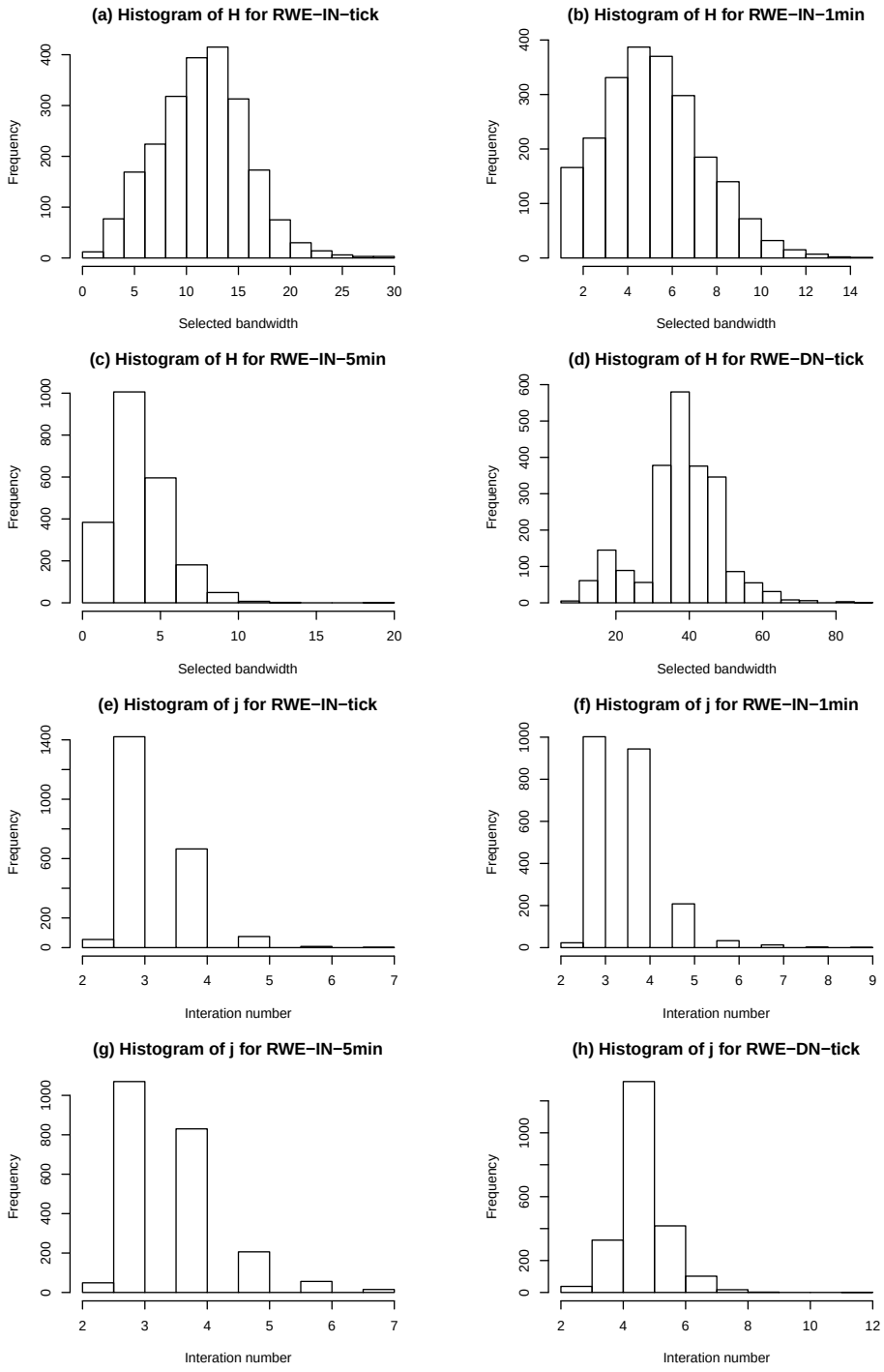

Figure A.4.5: Histograms of selected bandwidth and interation number for RWE.

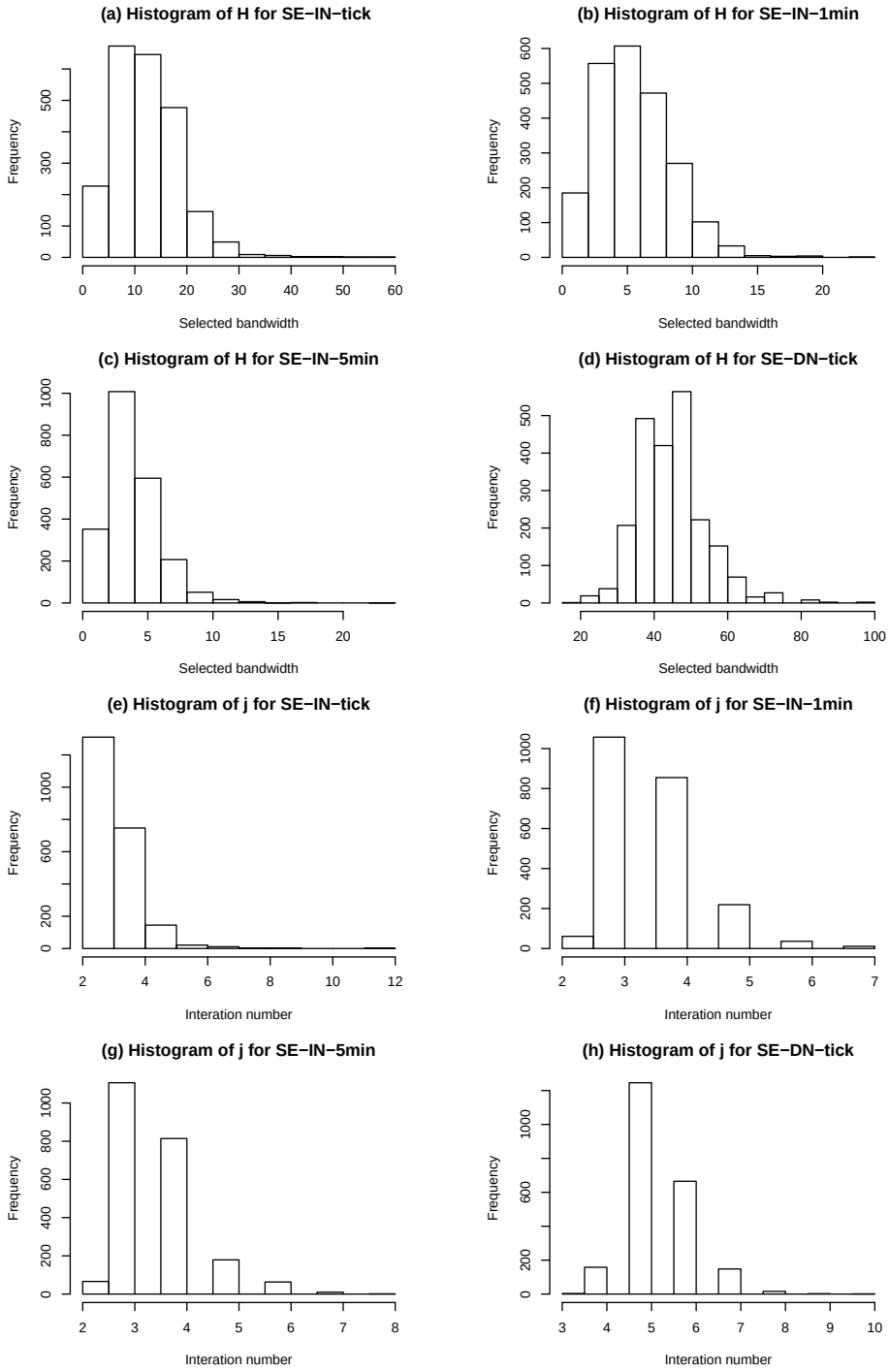

Figure A.4.6: Histograms of selected bandwidth and interation number for Schneider Electric.

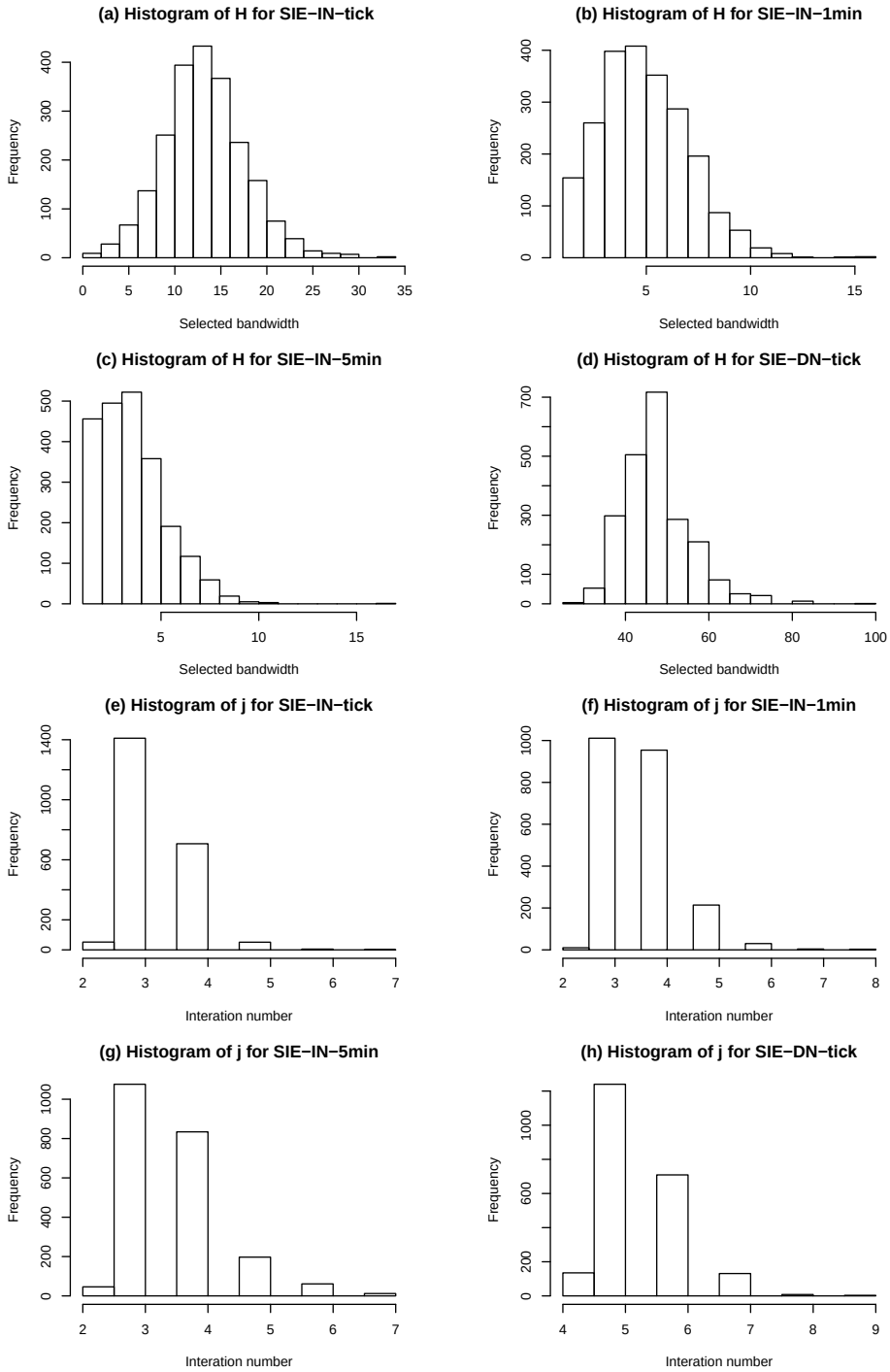

Figure A.4.7: Histograms of selected bandwidth and interation number for Siemens.

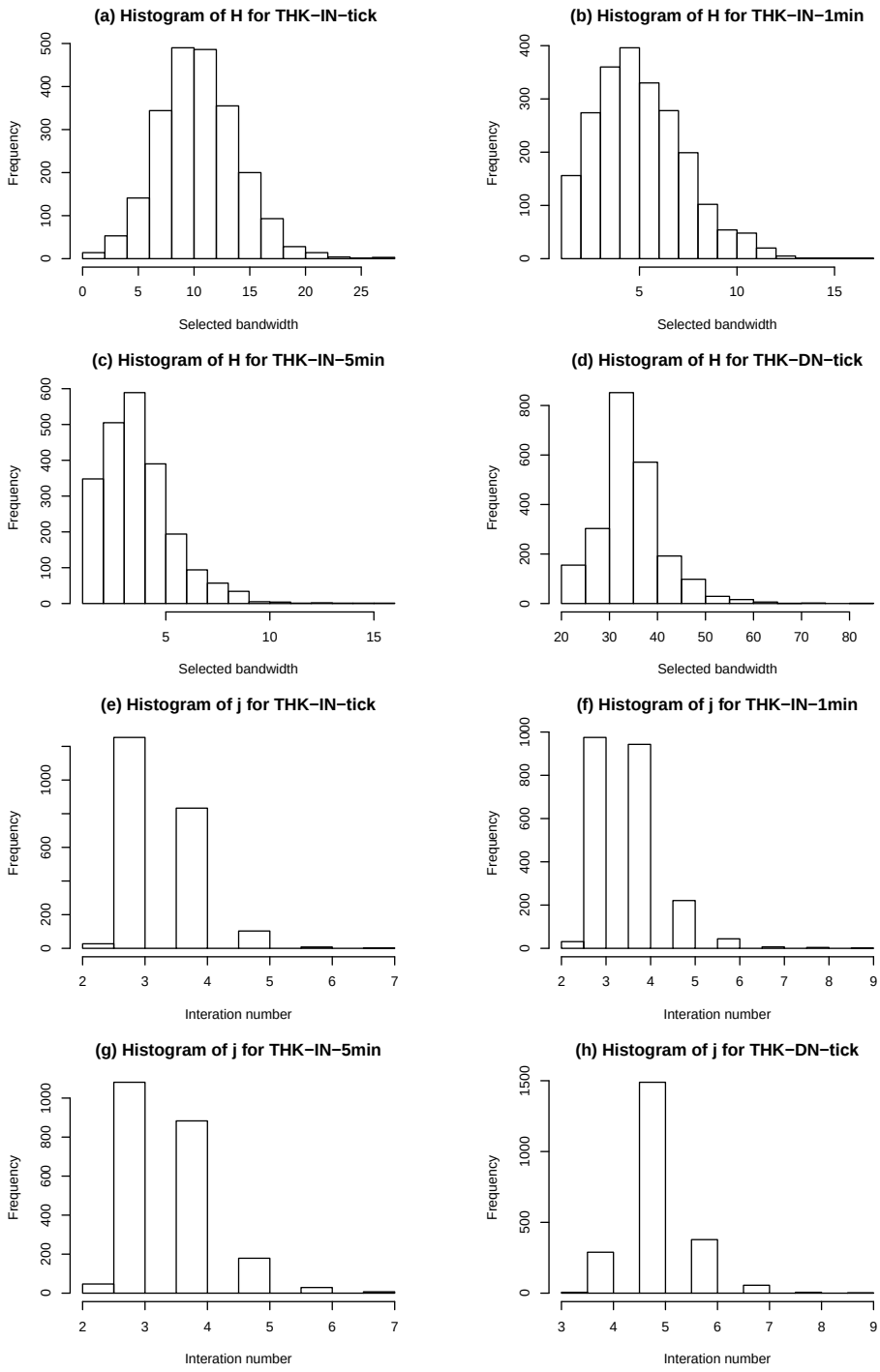

Figure A.4.8: Histograms of selected bandwidth and interation number for Thyssenkrupp.

Chapter 5

Conclusion

This thesis focuses on a data-driven realized kernels and its further analysis using the Semi-FI-Log-ACD model. It can be primarily divided into three parts.

In chapter 2 modeling and forecasting of long memory and a smooth scale function in different nonnegative financial time series aggregated from high-frequency data based on a fractionally integrated Log-ACD (FI-Log-ACD) and its semi-parametric extension (Semi-FI-Log-ACD) are discussed. It is shown that the EFARIMA model is equivalent to the FI-Log-ACD model, which means that the proposed FI-Log-ACD is the application of the well known FARIMA model to the log-process. Furthermore, necessary and sufficient conditions for the existence of a stationary solution of the FI-Log-ACD are obtained. These conditions are fulfilled, if η_t are log-normal innovations with $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Further examples which fulfill those conditions are the log-logistic and log-Laplace innovations with suitable restriction on the parameters. Detailed properties of the FI-Log-ACD under the log-normal assumption as obtained in Beran et al. (2012) are summarized. All of the processes X_t^*, Z_t, λ_t as well as $\zeta_t = \ln(\lambda_t)$ exhibit long memory with the same memory parameter, if $d > 0$. The long memory parameter in Z_t and in X_t^* under the log-normal assumption is the same. However, the constant in the asymptotic formula of $\rho_{X^*}(k)$ is smaller than that in $\rho_Z(k)$. If $d \leq 0$, X_t^* does not have antipersistent, even if Z_t is antipersistent. The study on the long-memory property of the conditional means ζ_t and λ_t in Z_t and X_t^* is also provided. Fore-

casting based on the Semi-FI-Log-ACD is then discussed in detail. The local linear regression is used to extrapolate the trend function. Because the sample size is very large, an approximately best linear predictor $\hat{Z}^*_{n+k}$ based on $Z_1, \ldots Z_n$ is defined. Its truncated version $\hat{Z}_{n+k}$ is proposed for the practical implementation. It is shown, that under given assumptions the two linear predictors $\hat{Z}_{n+k}$ and $\hat{Z}^*_{n+k}$ are asymptotically equivalent to each other and that the linear predictor $\hat{Z}_t$ is asymptotically equivalent to the exactly best linear predictor $\breve{Z}_t$ proposed by Beran and Ocker (1999). The proposed predictor is still an approximately best linear predictor in the presence of long memory. Asymptotic variances of the prediction errors for an individual observation and for the conditional mean are obtained. Calculation of approximate forecasting intervals under log-normal assumption is discussed. Effect of the errors in the estimated trend on the asymptotic properties of the proposed predictor is also investigated. The Semi-FI-Log-ACD is then applied for modeling and forecasting daily trading volumes, daily trading numbers, average durations and realized volatility of four European Stocks from Jan. 2, 2006 to Jun. 30, 2012. The results indicate that this model is widely applicable and the proposed linear predictor works very well in practice. It is also shown that the log-normal distribution is a suitable choice for different kinds of aggregated financial data.

The selection of the bandwidth is a crucial problem for applying the RK. RK work well only if the bandwidth is chosen properly. The main purpose of chapter 3 is to propose a simple, fast and fully data-driven consistent bandwidth selector for RK based on the iterative plug-in idee (Gasser et al., 1991). To simplify the estimation procedure we use a biased version of the asymptotically optimal bandwidth of the RK, called H_{B}, which is of the same order as H_{A} but with a biased factor in the constant. The difference between H_{A} and H_{B} is an constant factor $H_{\mathrm{B}}/H_{\mathrm{A}} = (T\int_0^T \sigma_u^4 du/\mathrm{IV}^2)^{1/5}$, which is usually slightly bigger than one. The selected bandwidth $\hat{H}_{\mathrm{B}}$ is obtained by means of an iterative procedure. An R code is programmed for practical implementation of the proposed bandwidth selector. In each iteration, the resulting RK is used as an estimate of the IV, and the variance of the microstructure noise ω^2 is estimated based on the difference

between RV_0 and RK. In the first iteration RV_Z is used as the initial value of RK. It is shown that $\hat{\omega}^2$ defined in this way is $\sqrt{n}$-consistent in each iteration. Both of RK and $\hat{H}_B$ become consistent form in the third iteration, while their rate of convergence can still be improved in the fourth iteration. Thereafter, RK achieves its optimal rate of convergence of the order $O(n^{-1/5})$ and this rate of convergence is also shared by $(\hat{H}_B - H_B)/H_B$. The robust practical performance of the proposal is illustrated by application to data of two German and two French firms within a period of several years. The proposed IPI algorithm converges very quickly. In most of the cases the proposed algorithm for the four companies converges within four iterations, which confirmed the results of Theorem 3.3.1. The numerical comparison of RV_0, RV_Z and RK shows that the use of the proposed RK will lead to a clear reduction of the bias and the variation. In addition, three special cases are considered as challenging days, where the proposed data-driven algorithm does not work well. The histograms of the selected optimal bandwidths for the four companies is nearly asymptotically normally distributed. In view of this, we use the ESEMIFAR model to further analyze the obtained RK. Possible structure breaks cased by the financial crisis in 2008 may have a clear effect on the estimation results. Using piecewise ESEMIFAR model can improve the quality of estimation results.

In chapter 4 we improve the IPI algorithm mentioned in chapter 3. Now, this IN algorithm works well for all cases. Furthermore, based on the conclusions of Ikeda (2015) and Wang (2014) a fully automatic data-driven algorithm to the dependent MS noise assumption, namely the DN algorithm is given. For both algorithms the end effect in the computation of the RK are considered. The non-flat-top Parzen kernel is used to guarantee the nonnegativity of RK. Theoretically, in the fourth iteration G and H have achieved to the optimal rate of convergence, respectively. After several iterations the resulted RK achieves its optimal rate of convergence of the order $O(n^{-1/5})$. It is also shown, that no matter what the starting bandwidths are, the selected bandwidths as of the second iteration are very close and the final selected bandwidths are indeed the same. The robust practical performance of both algorithms are illustrated by application to the data from 10 European

firms within 9 years. In addition, we study a comparison of RK using different sampling frequencies (tick-by-tick, 1-minute, 5-minute and 15-minute) calculated by the IN algorithm as well as RK using tick-by-tick returns calculated by the DN algorithm. Meanwhile, RV_0 computed for different frequency returns are also compared with these RK estimators mentioned above. It is ascertained, that the IN algorithm does not work well for the data/frequencies over 5 minute, because on some days the number of observations is smaller than the selected bandwidth. The mean values and standard deviations for tick-by-tick RV_0 are larger than any other volatility estimators and that diminish with the decline of sampling frequency. Further comparison is carried out by assessing their performances in the computation of Value-at-Risk (VaR) based on the Semi-FI-Log-ACD model. A Backtesting to examine the observed amount of exceptions (points over VaR) in excess of the benchmark is utilized. It is found that RK-IN-tick and RK-DN-tick estimators have good performances and are hence recommended using as the estimators of IV in practice.

There are still some open questions in this thesis. For instance, the idea of the piecewise ESEMIFAR model in chapter 3 can be applied properly. A suitable approach for detecting structural breaks under the SEMIFAR model can be developed. The proposed bandwidth selector in chapter 4 can be improved, if an unbiased estimator of the integrated quarticity can be developed. A possible modification of rating different realized measures is by means of comparing their MSE values.

Bibliography

[1] Allen, D.; Chan, F.; McAleer, M. and Peiris, S. (2008). Finite sample properties of the QMLE for the Log-ACD model: Application to Australian stocks. *Journal of Econometrics*, 147, 163-185.

[2] Aït-Sahalia, Y. and Mancini, L. (2008). Out of sample forecasts of quadratic variation. *Journal of Econometrics*, 147(1), 17-33.

[3] Aït-Sahalia, Y.; Mykland, P. A. and Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160, 160-175.

[4] Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4, 115-158.

[5] Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39 (4), 885 -905.

[6] Andersen, T. G.; Bollerslev, T.; Diebold, F. X. and Ebens, H. (2001a). The distribution of stock return volatility. *Journal of Financial Economics*, 61, 43-76.

[7] Andersen, T. G.; Bollerslev, T.; Diebold, F. X. and Labys, P. (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96, 42-55.

[8] Andersen, T. G.; Bollerslev, T.; Diebold, F. X. and Labys, P. (2003). Modelling and forecasting realized volatility. *Econometrica*, 71, 579-625.

[9] Andersen, T. G.; Bollerslev, T. and Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, 160, 220-234.

[10] Andersen, T. G.; Dobrev, D. and Schaumburg, E. (2009). Duration-based volatility estimation. Working Paper.

[11] Andersen, T. G.; Bollerslev, T. and Dobrev, D. (2007). No-arbitrage semimartingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications. *Journal of Econometrics*, 138(1), 125-180.

[12] Andersen, T. G.; Bollerslev, T.; Frederiksen, P. and ørregaard Nielsen, M. (2010). Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics*, 25(2), 233-261.

[13] Asai, M.; McAleer, M. and Yu, J. (2006). Multivariate stochastic volatiltiy: A review. *Econometric Reviews*, 25(2-3), 145-175.

[14] Baillie, R. T.; Bollerslev, T. and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 74, 3-30.

[15] Baillie, R. T and Morana, C. (2009). Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach. *Journal of Economic Dynamics & Control*, 33, 1577-1592.

[16] Bandi, F. M. and Russell, J. R. (2006). Separating market microstructure noise from volatility. *Journal of Financial Economics*, 79, 655-692.

[17] Bandi, F.M. and Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies*, 75, 339-369.

[18] Bandi, F. M. and Russell, J. R. (2011). Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *Journal of Econometrics*, 160, 145-159.

[19] Barndorff-Nielsen, O. E.; Hansen, P.R.; Lunde, A. and Shephard, N. (2008). Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. *Econometrica*, 76, 1481-1536.

[20] Barndorff-Nielsen, O. E.; Hansen, P.R.; Lunde, A. and Shephard, N. (2009). Realised kernels in practice: trades and quotes. *Econometrics Journal*, 12, C1-C32.

[21] Barndorff-Nielsen, O. E.; Hansen, P.R.; Lunde, A. and Shephard, N. (2011a). Subsampling realised kernels. *Journal of Econometrics*, 160 (1), 204-219.

[22] Barndorff-Nielsen, O. E.; Hansen, P.R.; Lunde, A. and Shephard, N. (2011b). Multivariate realized kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and nonsynchronous trading. *Journal of Econometrics*, 162, 149-169.

[23] Barndorff-Nielsen, O. E. and N. Shephard (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17 (5), 457-477.

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

[25] Bauwens, L. and Giot, P. (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.

[26] Beltratti, A. and Morana, C. (2005). Statistical Benefits of Value-at-risk with long memory. *Journal of Risk*, 7, 21-45.

[27] Beran, J.; Feng, Y. and Ghosh, S. (2015). Modelling long-range dependence and trends in duration series: an approach based on EFARIMA and ES-EMIFAR models. *Statistical Papers*, 56(2), 431.

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

[29] Beran, J. and Feng, Y. (2002b). Local polynomial fitting with long-memory short-memory and antipersistent errors. *Annals of the Institute of Statistical Mathematics*, 54, 291-311.

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

[31] Beran, J.; Feng, Y.; Ghosh, S. and Kulik, R. (2013). Long memory processes - probabilistic properties and statistical methods, Berlin: Springer.

[32] Beran, J. and Ocker, D. (1999). SEMIFAR forecasts, with applications to foreign exchange rates. *Journal of Statistical Planning and Inference*, 80, 137-153.

[33] Beran, J. and Ocker, D. (2001): Volatility of stock-market indexes - An analysis based on SEMIFAR models. *Journal of Business & Economics Statistics*, 19(1), 103-116.

[34] Bollen, B. and Inder, B. (2002). Estimating daily volatility in financial markets utilizing intraday data. *Journal of Empirical Finance*, 9, 551-562.

[35] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.

[36] Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices rates of return. *Review of Economics and Statistics*, 69, 542-547.

[37] Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, 72, 498-505.

[38] Bollerslev, T.; R. Engle and J. Wooldridge (1988). A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96, 116-131.

[39] Brockwell, P. J. and Davis, R. A. (2006). Time series: theory and methods (2nd edition), New York : Springer

[40] Broto, C. and Ruiz, E. (2004). Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys*, 18, 613-649.

[41] Carnero, M. A.; Peña, D. and Ruiz, E. (2004). Persistence and kurtosis in GARCH and stochastic volatiltiy models. *Journal of Financial Econometrics*, 2(2), 319-342.

[42] Corsi, F.; Zumbach, G.; Müler, U. A.; Dacorogna, M. M. (2001). Consistent high-precision volatility from high-frequency data. *Economic Notes*, 30(2), 183-204.

[43] Choi, K.; Yu, W. and Zivot, E. (2010). Long memory versus structural breaks in modeling and forecasting realized volatility. *Journal of International Money and Finance*, 29, 857-875.

[44] Corsi, F. (2009). A simple approximate long memory model of realized volatility. *Journal of Financial Econometrics*, 7, 174-196.

[45] Christensen, K.; Kinnebrock, S. and Podolskij, M. (2010a). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159, 116-133.

[46] Christensen, K. and Podolskij, M. (2007). Realized range-based estimation of integrated variance. *Journal of Econometrics*, 141, 323-349.

[47] Christensen, K.; Oomen, R. and Podolskij, M. (2010b). Realised quantile-based estimation of the integrated variance. *Journal of Econometrics*, 159(1), 74-98.

[48] Dacorogna, M.; Gencay, R.; Müller, U.; Olsen, R. B. and Pictet, O. V. (2001). An introduction to high-frequency finance. *Academic Press, London.*

[49] Deo, R.; Hsieh, M. and Hurvich, C. M. (2010). Long memory in intertrade durations, counts and realized volatility of NYSE stocks. *Journal of statistical planning and inference*, 140, 3715-3733.

[50] Deo, R.; Hurvich, C. and Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model. *Journal of Econometrics*, 131, 29-58.

[51] Ding, Z.; Granger, C. W. J.; and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1 (1993) 83-106, North-Holland.

[52] Dittman, I. and Granger, C. (2002). Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics*, 110, 113-133.

[53] Ebens, H. (1999). Realized stock volatility. *Working Paper* No. 420, Department of Economics, Johns Hopkins University.

[54] Engle, R. F. and Lee, G. G. J. (1999). A permanent and transitory component model ofstock return volatility. In R. Engle, & H. White (Eds.), Cointegration, causality, and forecasting: A festschrift in honor of Clive W. J. Granger(pp. 475-497). Oxford: Oxford University Press.

[55] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1007.

[56] Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, 17, 425-446.

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

[58] Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new approach for irregularly spaced transaction data. *Econometrics*, 66, 1127-1162.

[59] Feng, Y. (2004). Simultaneously modelling coditional heteroskedasticity and scale change. *Econometric Theory*, 20, 563-596.

[60] Feng, Y. and Beran J. (2013). Optimal convergence rates in non-parametric regression with fractional time series errors. *Journal of Time Series Analysis*, 34, 30-39.

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

[62] Feng, Y. and Zhou, C. (2015b). An iterative plug-in algorithm for realized kernels. *Preprint*, University of Paderborn.

[63] Fleminga, J. and Kirbyb, C. (2011). Long memory in volatility and trading volume. *Journal of Banking & Finance*, 35, 1714-1726.

[64] French, K. R.; Schwert, G. W. and Stambaugh, R. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19, 3-29.

[65] Gasser, T.; Kneip, A. and Köhler, W. (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association*, 86(415), 643-652.

[66] Geweke, J. (1986). Modeling the Persistence of Conditional Variances: A Comment. *Econometric Review*, 5, 57-61.

[67] Ghysels, E.; Harvey, A. C. and Renault, E. (1996). Stochastic volatility. *Handbook of Statistics*, 14, 119-191.

[68] Giot, P. and Laurent, S.(2004). Modeling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, 11(3), 379-398.

[69] Glosten, L.; Jagannathan, R. and Runkle, D. (1993). On the relation between expected value and the volatilty of the nominal excess return on stocks. *Journal of Finance*, 48, 1779-1801.

[70] Gramming, J. and Maurer, K. O. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *The Econometrics Journal*, 3, 16-38.

[71] Granger, C. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227-238.

[72] Griffin, J. E. and Oomen, R. C. A.(2008). Sampling returns for realized variance calculations: tick time or transaction time? *Econometric Reviews*, 27(3), 230-253.

[73] Hansen, P. R.; Large, J. and Lunde, A. (2008). Moving average-based estimators of integrated variance. *Econometric Reviews*, 27, 79-111.

[74] Hansen, P. R. and Lunde, A. (2004). An unbiased measure of realized variance. Working Paper, Department of Economics, Stanford University.

[75] Hansen, P. R. and Lunde, A. (2006a). Realized variance and market microstructure noise (with discussion). *Journal of Business and Economic Statistics*, 24, 127-218.

[76] Hansen, P. R. and Lunde, A. (2006b). Consistent ranking of volatility models. *Journal of Econometrics*, 131, 97-121.

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

[78] Higgins, M. and Bera, A. (1992). A class of nonlinear ARCH models. *International Economic Review*, 33, 137-158.

[79] Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68, 165-176.

[80] Hsieh , D. A. (1991). Chaos and nonlinear dynamics: Application to financial markets. *Journal of Finance*, 46(5), 1839-1877.

[81] Ikeda, S. (2015). Two scale realized kernels: a univariate case. Boston University. *Journal of Financial Econometrics*, 13(1), 126-165.

[82] Jasiak, J. (1998). Persistence in intertrade durations. *Finance*, 19, 166-195.

[83] Karanasos, M. (2004). The statistical properties of long-memory ACD models. *WSEAS Transactions on Business and Economics*, 2, 169-175.

[84] Karanasos, M. (2008). The statistical properties of exponential ACD models. *Quantitative and Qualitative Analysis in Social Sciences*, 2, 29-49.

[85] Karanasos, M.; Psaradakis, Z. and Sola, M. (2004). On the autocorrelation properties of long-memory GARCH processes. *Journal of Time Series Analysis*, 25, 265-281.

[86] Koopman, S. J.; Jungbacker, B. and Hol, E. (2005). Forecasting daily variability of the S&P100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, 12, 445-475.

[87] Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3, 73-84.

[88] Large, J. (2007). Estimating quadratic variation when quoted prices jump by a constant increment. *SSRN Electronic Journal*, 160(1), 2-11.

[89] Leipus, R.; Philippe, A.; Puplinskaite, D. and Surgailis, D. (2014). Aggregation and long memory: recent developments. *Journal of the indian statistical association*, 52, 71-101.

[90] Liu, L.; Patton, A. and Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187, 293-311.

[91] Lunde, A.; Shephard, N. and Sheppard, K. (2016). Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business & Economic Statistics*, 34(4), 504-518.

[92] Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets*, 8, 377-399.

[93] Martens, M. (2002). Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. *Journal of Futures Markets*, 22, 497-518.

[94] Martens, M. and Dijk, D. (2007). Measuring volatility with realized range. *Journal of Econometrics*, 138, 187-207.

[95] Mikosch, T. and Stăarică, C. (2004). Change of structure in financial time series, long range dependence and the GARCH model. *REVSTAT-Statistical Journal*, 2(1), 41-73.

[96] Nelson, D. B. (1991), Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347-370.

[97] Oomen, R. C. A. (2005). Properties of bias-corrected realized variance under alternative sampling schemes. *Journal of Financial Econometrics*, 3, 555-577.

[98] Oomen, R. C. A. (2006). Properties of realized variance under alternative sampling schemes. *Journal of Business and Economic Statistics*, 24, 219-237.

[99] Pantula, S. (1986). Modeling the persistence of conditional variances: A comment. *Econometric Review*, 5, 71-74.

[100] Peitz, C. (2016). Die parametrische und semiparametrische Analyse von Finanzzeitreihen- Neue Methoden, Modelle und Anwendungsmöglichkeiten. Wiesbaden: Springer.

[101] Peters, R. T. and Vilder, R. G. (2006). Testing the continuous semimartingale hypothesis for the S&P 500. *Journal of Business & Economic Statistics*, 24, 444-454.

[102] Podolskij, M. and Vetter, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3), 634-658.

[103] Schwert, W. (1990). Stock volatility and the crash of 87. *Review of Financial Studies*, 3, 77-102.

[104] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V. and Barndorff-Nielsen, O.E., Eds., Time Series Models in Econometrics, Finance and Other Fields, Monographs on Statistics and Applied Probability, Vol. 65, Chapman and Hall, 1-67.

[105] Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices 1961-79. In: Anderson, O.D., Ed., Time Series Analysis: Theory and Practice 1, North-Holland, Amsterdam, 203-226.

[106] Taylor, S. J. (1986). Modelling financial time series. Wiley, New York.

[107] Wang, J. (2014). Research on ACD model and its application. *Doctorial Dissertation*. Zhongnan University of Economics and Law.

[108] Wasserfallen, W. and Zimmermann, H. (1985). The behavior of intraday exchange rates. *Journal of Banking and Finance*, 9, 55-72.

[109] Zaffarroni, P. (2007). Contemporaneous aggregation of GARCH processes. *Journal of Time Series Analysis*, 28, 521-544.

[110] Zakolan, J. M. (1994). Threshold heteroskedasticity models. *Journal of Economic Dynamics and Control*, 15, 931-955.

[111] Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12, 1019-1043.

[112] Zhang, L.; Mykland, P. A. and AÏt-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association*, 100, 1394-1411.

[113] Zhou, B. (1996). High-frequency data and volatility in foreign exchange rates. *Journal of Business and Economic Statistics*, 14, 45-52.