Modelling and forecasting financial and economic time series using different semiparametric ACD models

Der Fakultät für Wirtschaftswissenschaften der
Universität Paderborn
zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
- Doctor rerum politicarum -
vorgelegte Dissertation
von

M.A. Sarah Forstinger

geboren am 13.11.1986 in Frankfurt/ Höchst

(2018)

Acknowledgements

The work presented in the following is the result of my research carried out in 5 years as research assistant at the professorship of Econometrics and Quantitative Methods of Empirical Economic Research, Paderborn University. My sincere gratitude goes to its holder and my supervisor, Professor Dr. Yuanhua Feng, who was always available for discussing my research and who provided me with helpful comments and food for thought throughout my whole dissertation project. The joint research we did was productive and also allowed me to gain deeper insight into other fields of research besides the one of analysing financial market data. A special thanks goes to the other authors I was able to work with, Professor Dr. Thomas Gries, Dr. Christian Peitz and Marlon Fritz, for their valuable input and productive collaboration. I would also like to thank Professor Dr. Yuanhua Feng, Professor Dr. Thomas Gries, Professor Dr. André Uhde and Professor Dr. Stefan Jungblut for agreeing to be the members of my doctoral committee.

Danksagung

Der größte Dank gilt meinen Eltern, Anne und Klaus, meiner Schwester Nina, meiner Oma Leni und meinem Partner Malte für ihre bedingungslose und liebevolle Unterstützung.

Abstract

The contents of this thesis are on a semiparametric extension of the ACD model of Engle and Russell (1998). The proposal of the Semi-ACD model is based on the decomposition of the data of interest into a deterministic and a stochastic part, whereby the former is assumed to be time-varying. A non-negative, time-varying, smooth scale function is included into the model to take this into account. The proposal of this thesis is its estimation with a local polynomial regression. An automatic iterative plug-in bandwidth selection algorithm is developed. No prior assumptions about a specific model are required and the scale function estimation can be used flexibly. The estimated trend is removed from the data and any parametric model can be fitted to the standardized data. A simulation study evaluates the Semi-ACD model on the basis of various criteria. In direct comparison with the cubic spline method, it is clearly superior. Non-consideration of the deterministic component is a clear misspecification. An extension of the proposal to log data shows that the estimation of the scale function is clearly simplified. In addition, decisive theoretical properties for the Semi-Log-ACD model are derived and the bandwidth selection algorithm is further automated. It is shown that this does not affect the rate of convergence of the asymptotical optimal bandwidth. To forecast non-negative financial data, the above models are combined with known and new forecasting methods. In order to not limit the flexibility of the semiparametric idea, bootstrap methods are chosen as nonparametric forecasting methods. Compared to model-based Kalman filter predictions, these give not the best forecasts, but are clearly better compared to the corresponding parametric model forecasts. The algorithm is applied to forecast the log-GDP of developing and developed countries. Random Walk models with a constant drift, a linear drift and a local linear drift are applied, as well. It is found that combining forecasting methods improves the forecasts and especially including the local linear regression method stabilizes the forecasts and enables the detection of variations in the trend process, that are typical for developing countries. Promising research questions to further improve the Semi-(Log-)ACD models are presented. In particular, first results of a simulation study show that applying the local polynomial regression IPI to log-transformed returns in a GARCH model framework works well. Pursuing this proposal further should be of great value for the research on quantitative risk management.

Zusammenfassung

Diese Arbeit stellt eine semiparametrische Erweiterung des ACD-Modells von Engle und Russell (1998) vor. Der Vorschlag des Semi-ACD-Modells basiert auf der Zerlegung der Daten in einen deterministischen und einen stochastischen Teil, wobei der Erste als zeitvariabel angenommen wird. Um dies entsprechend zu berücksichtigen wird eine nicht-negative, zeitvariable, glatte Skalenfunktion in das Modell aufgenommen. Der Vorschlag dieser Arbeit ist ihre Schätzung mit einer lokal polynomialen Regression. Ein automatischer, iterativer Plug-In Bandbreitenwahl-Algorithmus wird entwickelt. Es sind keine vorherigen Annahmen über ein bestimmtes Modell erforderlich und die Skalenfunktionsschätzung kann flexibel eingesetzt werden. Der geschätzte Trend wird aus den Daten entfernt und ein beliebiges parametrisches Modell kann an die standardisierten Daten angepasst werden. Anhand verschiedener Kriterien aus einer Simulationsstudie wird das Semi-ACD Modell bewertet. Im direkten Vergleich zur kubischen Spline-Methode ist sie deutlich überlegen. Die Nichtberücksichtigung der deterministischen Komponente ist eine eindeutige Fehlspezifikation. Eine Erweiterung des Semi-ACD Modells zur Anpassung an Log-Daten zeigt, dass die Schätzung der Skalenfunktion deutlich vereinfacht wird. Darüber hinaus werden entscheidende theoretische Eigenschaften für das Semi-Log-ACD-Modell abgeleitet und der Bandbreitenwahl-Algorithmus weiter automatisiert. Dies hat keinen Einfluss auf die Konvergenzrate der asymptotisch optimalen Bandbreite. Zur Prognose nicht-negativer Finanzdaten werden die oben genannten Modelle mit bekannten und neuen Prognosemethoden kombiniert. Bootstrap-Methoden werden als nichtparametrische Prognosemethoden gewählt. Im Vergleich zu modellbasierten Kalman-Filter-Vorhersagen liefern diese nicht die besten Vorhersagen, sind aber deutlich besser als die entsprechenden parametrischen Modellvorhersagen. Der Algorithmus wird weiter verwendet, um das Log-BIP von Entwicklungs- und Industrieländern zu prognostizieren. Auch Random Walk Modelle mit konstantem, linearem und lokal linearem Drift werden verwendet. Die Kombination von Prognosemethoden verbessert die Vorhersage und insbesondere die Einbeziehung der lokal linearen Regressionsmethode stabilisiert sie. Zudem wird die Erkennung von für Entwicklungsländer typischen Schwankungen im Trendprozess ermöglicht. Forschungsfragen zur weiteren Verbesserung der Semi-(Log-)ACD-Modelle werden vorgestellt. Insbesondere zeigen ersten Ergebnisse einer Simulationsstudie, dass die Anwendung des lokal polynomialen Regressionsalgorithmus für log-transformierte Renditen in einer GARCH-Modell Umgebung gut funktioniert. Eine tiefere Untersuchung und Fortführung dieser Idee sollten von großem Wert für die Forschung im Bereich des quantitativen Risikomanagements sein.

Contents

List of Abbreviations

ACD	Autoregressive Conditional Duration
ACDD	Autoregressive Conditional Directional Duration, Jeyasreedharan et al. (2014)
ACMD	(Autoregressive Conditional Marked Duration, Tay et al. (2004) and Kwok et al. (2009)
AIC	Akaike Information Criterion
AMISE	Asymptotic Mean Square Error
ARCH	Autoregressive Conditional Heteroskedasticity
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ASE	Average Squared Errors
BIC	Bayesian Information Criterion
BVAR	Bayesian Vector Autoregression
CS	Cubic Spline
CV	Cross-Validation
d-ACD	daily ACD, Zuccolotto (2002)
e.g.	exempli gratia
EACD	Exponential Autoregressive Conditional Duration
EF	Estimating Function
EGARCH	Exponential GARCH
EIM	Exponential Inflation Method

EoPoFi	Error of Points outside Forecasting Interval
EqCM	Equilibrium Correcting Model
FI	Forecasting Interval
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GMM	Generalized Maximum Likelihood
HF	High-frequency/ high-frequent
i.e.	id est
i.i.d.	independent identically distributed
IPI	Iterative Plug-In
JB	Jarque-Bera
LL	Local Linear
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
MD	daily average durations
MIM	Multiplicate Inflation Method
MISE	Mean Integrated Squared Error
ML	Maximum Likelihood
MSE	Mean Square Error
MSFE	Mean Squared Forecast Errors
NN	Neural Networks
NNAR	Autoregressive Neural Network
NNARLL	Autoregressive Neural Network Local Linear Regression
NW	Nadaraya-Watson
OECD	Organisation for Economic Co-operation and Development
PoT	Peaks over Threshold

QML	Quasi Maximum Likelihood
QQ	Quantile-Quantile
RASE	Root Average Squared Errors
REFF	Relative efficiencies
RMSSE	Root Mean Square Scaled Error
RSE	(Empirical) reduction of squared errors
RV	realized volatility
SACD	Semiparametric ACD model, proposed by Dungey et al. (2014)
SAS	Sinh-Arcsinh
SCoD	Semiparametric Copula Duration
SEMIFAR	Semiparametric Fractional Autoregressive
SW	Shapiro Wilk
TrNo	daily trading number
UHF	Ultra-High Frequency / ultra-high frequent
v.v.	vice versa
VAR	Vector Autoregression
VEC	Vector Error Correction
Vol	daily trading volume
YAARCH	Yet Another ARCH

List of Figures

List of Tables

1 Introduction[1]

1.1 Market microstructure theory and the ACD model

One of the many technological advancements of the 21^{st} century is the possibility to record data of interest on ultra-high frequency level (UHF), i.e. fully and usually the second it arises (see Engle, 2000). Especially for the analysis of financial market dynamics, the accessibility of data on micro level, exact to the second and containing the whole range of information corresponding to the data, gives unprecedented opportunities. It opens up new fields of research in financial economics and econometrics, such as the analysis of UHF trade durations (see Hautsch, 2004). They are defined as the time passed between two consecutive transactions. Used as a proxy for information or trading intensity, trade durations are assumed to deliver valuable information concerning the processing of information in financial markets. This is an essential idea of market microstructure theory as pointed out, for instance, by Glosten and Milgrom (1985), Admati and Pfleiderer (1988) or Easley et al. (1997). To fully utilize the opportunity of analysing UHF trade duration data, its salient features need to be addressed appropriately. The irregular time-spacing of the data is one of these features and poses a severe challenge for the theoretical development and practical application of suitable models. Time series models, such as the (Generalized) Autoregressive Conditional Heteroskedasticity ((G)ARCH) model by Engle (1982) and Bollerslev (1986), or ARMA models require the data to be equidistant and are therefore not applicable. In order to adequately address this problem without possibly alleviating the information content of UHF data through aggregation, Engle and Russell (1998) develop the Autoregressive Conditional Duration (ACD)

[1] The literature mentioned in this chapter makes no claim to completeness. Rather it is intended to put the following theoretical and applied research into context by giving selected examples of available research literature.

model. The ACD model describes trade durations as the product of their conditional expected duration and i.i.d. positive valued innovations. Treating trade durations as a sequence of different time intervals, the stochastic process that generates them is a point process. The classes of point processes are modelled with the expected duration being conditional on time as a function of past observations and its own lagged history. Besides dealing with the irregular spacing of the data, the ACD model is also able to capture the clustering of trade durations, another characteristic feature of UHF financial trade duration data. In a market microstructure theoretical frame, Easley and O'Hara (1992) explain the clusters with new information being processed at an unequal speed by the market participants. Due to asymmetric information, an increase in trading intensity is caused by informed traders wanting to take advantage of the arrival of news. Uninformed traders, however, only suspect additional information due to the increase in trading activity and follow the actions of the allegedly informed traders. Thus, trading is done over a period of time, rather than a single point in time and long durations follow long durations and short durations follow short durations. The analysis of trade durations can explain why an agent will trade by associating the durations with the intensity of liquidity demand. Also the question of when an agent will trade can be addressed by assuming that an informed trader will act on the basis of exclusive information, whereas the uninformed traders only follow. Among other research questions in the field of market microstructure theory, the analysis of trade durations to infer on the trading intensity in the financial market is a very active field of research. For example, Dufour and Engle (2000b) apply the ACD model to discuss the impact of trade durations on the price formation process. Zuccolotto (2002) introduces the daily ACD (d-ACD) model to analyse the relationship between the number of shares traded when a stock market opens and intra-daily trading intensity. Spierdijk (2004) studies the role of trade durations in the dissemination of information by applying the ACD model to specify the data generating process which underlies the trading intensity. Tay et al. (2009) aim to infer on the probability of informed trading by applying the asymmetric ACD model of Bauwens and Giot (2003). Over the course of time, with new models and methods available, results and findings are being updated. Diamond and Verrechhia (1987) state, for example, that periods of no trading activity resulting in long trade durations are due to bad news, whereas Easley and O'Hara (1992) see long trade durations as a sign of no news. Karaa et al. (2013) find that high informed trading follows low liquidity, which is in contrast to the findings of Admati and Pfleiderer (1988). Karaa et al. (2013) further find, that high trading intensity is only related to high volatility, large trading volume and narrow spreads for liquid stocks. This coincides with the results of Manganelli (2005).

For testing other market microstructure theory hypotheses, so-called marks associated with each trade are used. Marks define a subset with specific characteristics, which is considered for the analysis, so that the process is selectively thinned out (Engle and Russell, 1998). For example, only transactions are considered for the analysis for which a certain change in the price is given. Gouriéroux et al. (1999) aim to infer on liquidity by capturing the dependencies between intra-trade durations and transaction volumes or prices. They introduce duration based activity measures for the time until a fixed volume or price is traded. Giot (2000) studies intra-day volatility by applying the Log-ACD model of Bauwens and Giot (2000) and Bauwens et al. (2008) to data with a price and volume mark. Kwok et al. (2009) extend the ACMD (Autoregressive Conditional Marked Duration) model of Tay et al. (2004) by including a three-state price movement indicator into the model. They find that their results support Easley and O'Hara's (1992) result on long trade durations being due to the absence of news. However, the results of Tay et al. (2004) confirm Diamond and Verrecchia (1987) in trading intensity being low as a consequence of the arrival of bad news. Depending on the features of the data at hand, different specifications for the expected durations and/or different conditional distribution choices for the innovations make the ACD model very flexible. The contents of the following work are oriented towards the development of suitable ACD models and ACD model extensions. A market microstructure theoretical interpretation of the results is not provided, but rather an evaluation of models and methods used with regards to their practical performance. The previous brief descriptions of selected market microstructural topics and articles are used to embed the model in a financial market theory environment. It is shown in the following, though, that the area of application of the ACD model is much wider.

The ACD model as the primary focus of this work is very flexible and manifold regarding its setup and financial areas of application.[2] Saart et al. (2015) classify the development of the ACD model into three generations. The first generation extends the originally proposed standard exponential and Weibull conditional distribution assumptions. Grammig and Maurer (2000) allow for non-monotonic hazard functions with their proposal of a Burr distribution and Zhang et al. (2001) use a generalized Gamma conditional distribution for their Threshold ACD model. In order to increase the flexibility of the ACD model, mixtures of the error distributions are proposed in

[2]The main area for application of the ACD model is in finance. Due to its universal and adjustable nature it is sometimes applied in non-finance research, as well. For example, Vlahogianni et al. (2011) use it to estimate the congestion duration in urban signalized arterials in Athens. Khan and Mittnik (2018) compare the performance of different autoregressive linear and nonlinear time series models concerning their modelling and forecasting qualities in the field of earthquake seismology of the Hindu Kush region.

the literature, as well. For example, De Luca and Gallo (2004, 2008 and 2010) study the mixture of distributions with (time-varying) weights for a MEM (multiplicative error model, Engle, 2002), a generalization of the ACD model. Gómez-Déniz and Pérez-Rodríguez (2016) extend the model of De Luca and Zuccolotto (2003) from a finite and infinite mixture of exponential distributions to distribution mixtures of non-exponentials based distributions. The second generation of the ACD model development extends the trade duration specification non-linearly. The Log-ACD model, for example, contains two model specifications based on the log-transformed data. The Threshold ACD model of Zhang et al. (2001), mentioned above, is another example of the second generation ACD model, because it allows the expected conditional duration to depend nonlinearly on past information. Also including another model for different components of the data is common. Ghysels and Jasiak (1998), for instance, define the ACD-GARCH model which allows past asset return volatilities to affect trade durations and v.v.. Engle (2000) also proposes an ACD-GARCH model, where an ACD model specifies the durations and the GARCH model is applied to the volatility of returns, conditional on durations.[3] This model is taken up in either form and developed further by other authors, such as Grammig and Wellner (2002), Min et al. (2003), Racicot et al. (2008), Czado and Haug (2009, 2010) or Chung and Hwang (2016). Pacurar (2008) gives a very detailed overview over the scientific literature on the ACD model development since its first proposal. Hautsch (2011) gives some examples of further ACD models and Kaur Bhogal and Thekke Variyam (2018) give a "post-Pacurar (2008)" (Kaur Bhogal and Thekke Variyam, 2018, p.2) structured literature review on the ACD model development following the classification of Saart et al. (2015).

A final example of a recently published ACD model extension serves to summarize the core statement of this section with regard to the multiple extension possibilities. Jeyasreedharan et al. (2014) propose the autoregressive conditional directional duration (ACDD) model, which defines negative durations if the trade is bid-driven and positive durations when it is ask-driven. The mean equation of the model follows a semiparametric fractional autoregressive (SEMIFAR) formulation and the variance is modelled via GARCH. The resulting SEMIFAR-ACDD model addresses the now symmetric distribution of the directional durations, as well as persistence and long-memory. It combines several extensions and modifications and the title of the paper "Yet another ACD model: the Autoregressive Conditional Directional

[3]Meddahi et al. (2006) discuss the differences between the ACD-GARCH models of Ghysels and Jasiak (1998) and Engle (2000). Without getting into too much detail the differences are in the conditioning information, the GARCH formulation and the GARCH representation of the returns. Eventually Meddahi et al. (2006) combine the advantages of each ACD-GARCH model in another model proposal.

Duration (ACDD) model" possibly alludes to the use of the term YAARCH (Yet another ARCH; Figlewski, 1995 conference presentation UCSD), described by Engle (2002) as a "linguistic culmination" (Engle, 2002, p. 426), to express the multitude of extensions available for the class of ARCH models.

The demarcation required for the overview of the state of research presented in the following section is drawn as follows: The contents of this work are about semiparametrically extending the ACD model. This falls into the third generation specification of Saart et al. (2015) and is motivated by taking into account another specific feature of UHF data: the daily pattern for UHF and the long-term dynamics for High-Frequency (HF) data. The daily pattern of the volume of trades, spreads and their volatility was found to be U-shaped (see McInish and Wood, 1992), whereas it typically is inversely U-shaped for trade durations (Engle and Russell, 1998). In either case, the stationarity assumption of the ACD model is violated. The main idea underlying the following contents is based on decomposing the data. In a first step, the data is detrended via nonparametric methods and the resulting residuals are fitted to an ACD-type model parametrically. The next section gives an overview over the research literature available on non- and semiparametric methods in the ACD model and ACD model-related context.[4]

1.2 State of research on non- and semiparametric ACD models

Referring to the combination of parametric and nonparametric methods, there is a variety of semiparametric ACD or ACD model-related extensions, that address different aspects of the model. Solgi and Mira (2013), for instance, propose a semiparametric MEM where the nonparametric estimation applies to the innovation distribution. For a countable infinite mixture of Gamma distributions with two free parameters a Bayesian approach is described to increase the flexibility of the model. Drost and Werker (2004) propose a semiparametric duration model which contains a semiparametrically efficient score functions estimator of the unknown model parameter vector. Also, the model allows to specify the dependencies between the innovations to range from being i.i.d. to being random, as the i.i.d. assumption of the ACD model might be too restrictive in some cases. Ranasinghe and Silva-

[4]'ACD and ACD model-related' includes the ACD model as proposed by Engle and Russell (1998) and its different forms, concerning e.g. the distribution choice of the innovations, different specifications of the conditional durations or transformation of the data. Also literature on semiparametric MEM and GARCH models is considered in the following, as the ACD model is a special case of the MEM and closely related in its form and properties to the GARCH model.

pulle (2011) build on this idea of a semiparametric unknown model parameter vector estimator and develop a constrained version. The possible problem of parameter estimations lying outside the constrained parameter space and thereby violating the imposed non-negativity restrictions is addressed with this model. Zuccolotto (2003) takes another approach by claiming that traders receive more information on financial market events from the quantiles estimation than from estimating the expected trade durations. A parametric approach, assuming an exponential distribution and a semiparametric distribution-assumption free approach are proposed.

Here, the data is decomposed into a deterministic and a stochastic part to account for intra-daily seasonality or long-term dynamics nonparametrically and to model them together with the conditional dynamics of the data series. Decomposing the data is also a common approach for the analysis of financial returns with GARCH models. Engle and Lee (1999), for example, propose a volatility component model which describes the considered volatility process with an additive relationship between a stochastic long-run and a short-run component. Formally the decomposition is done by replacing the unconditional variance in the GARCH model specification of the conditional variance by the long-run component, which itself is described in an autoregressive manner. Engle and Rangel (2008) decompose the components as determined by Engle and Lee (1999) multiplicatively and use a nonparametric exponential quadratic spline approach for the estimation of the introduced trend. Similarly, Engle and Russell (1998) propose to multiplicatively decompose the trade durations for the ACD model into a stochastic and a deterministic part. For estimating the latter component cubic splines are proposed with knots chosen on every hour and additionally on the last half hour of the trading day. Cubic splines are also used by Bauwens and Giot (2000) to determine whether there is seasonality within a trading week and are still a prominent tool to eliminate intraday seasonality in financial data (see e.g. Simonsen, 2007, Czado and Haug, 2009 or Karaa et al., 2013 and 2017). The performance of the method introduced in this work is compared with the cubic spline method as a reference in chapter 2 and is shown to be better. Other suggestions for decomposition are, e.g. by Brownlees et al. (2010) which define three components to model the data: a multiplicative daily, an intra-daily periodic and an intra-daily dynamic one. The same authors propose the composite MEM in 2012, where the conditional mean is described additively by a time-varying and a zero-mean stationary component.

In general, the dynamic and stochastic components can either be estimated jointly or via a two-step procedure. Pohlmeier and Gerhard (2001) analyze intraday volatility through the estimation of transaction price changes via an ordered probit model with conditional heteroskedasticity. A two-stage estimation procedure is not applicable, so they define the discrete price jump as a count variable and extend it to the domain of negative integers. A Fourier series approximation (Andersen and Bollerslev, 1997) is used to identify the factors driving the volatility. Rodríguez-Poo et al. (2008) propose a method to jointly estimate the seasonal and dynamic components of durations by extending the generalized profile likelihood techniques by Severini and Wong (1992) to non i.i.d. observations. They compare one-step and two-step estimations with a modified Nadaraya-Watson (NW) and splines, each. They compute the bandwidth data-driven for the one-step NW method and find that this allows the method to quickly adapt to changes in the seasonal pattern. Brownlees and Gallo (2011) propose the joint estimation of the two components in their semiparametric MEM. They use a shrinkage type estimator, which shrinks the parameters of the deterministic part towards zero. A quadratic form of this component's coefficients serves as the penalty function for the penalized log-likelihood of the model. Dungey et al. (2014) also propose a two-stage semiparametric ACD (SACD) model, which aims to minimise the risk of model misspecification by subsequently correcting the fitted parametric ACD model. The methods for the two-stage semiparametric estimator of the conditional variance of Mishra et al. (2010) and Long et al. (2011) are adapted and parametric and nonparametric conditional duration estimators are combined in a multiplicative way. In their first step an ACD model is fitted parametrically. Then, nonparametric estimates of the conditional mean of the standardized parametric residuals are obtained and used as a correction factor for the parametric ACD estimator in a second step. By doing so, misspecification in the parametric model might be corrected. The results of empirically applying the model with a local exponential method and a least squares Cross-Validation (CV) approach for obtaining the optimal bandwidth show, that the duration coverage of the semiparametric model specification is superior to the parametric model. Another two-step procedure is described by Savu and Ng (2006), who propose the Semiparametric Copula Duration model (SCoD). After deseasonalising the observations with a non-linear kernel regression, the copula models are estimated in two stages. A random sample is drawn from the vector of durations and lagged durations and using the empirical distribution, the marginal distributions are estimated nonparametrically. The copula parameter vector is then estimated in a second step by maximizing a log-likelihood function.

This cross-section through parts of the literature on semiparametric ACD and ACD-related models shows, that one of the overarching objectives is an increased flexibility of the model by waiving prior assumptions on the conditional distribution or the model specifications, for example. The approach presented in this thesis fits into the current literature on the semiparametric estimation of non-negative financial data by presenting an approach, with which the data of interest can be modelled more flexibly. It is a two-stage estimation procedure in order to avoid possible numerical issues arising with a joint estimation of the deterministic and the stochastic component (Hautsch and Pohlmeier, 2001) and to allow the estimation procedure to be of a rather general nature. This proposal is derived from the close structural relationship between the GARCH and the ACD model. Feng (2004a) developed the Semi-GARCH model for simultaneously modelling the conditional heteroskedasticity and scale change in financial returns. This idea is carried over to the ACD model context by Feng (2014) and extended further throughout the course of this thesis. The principal contribution, that runs through all chapters of this work is an iterative plug-in (IPI) algorithm for estimating the scale function of the data with an automatically selected bandwidth (see Gasser et al., 1991). It is shown to work well in theory and in practice and it can also be extended and adapted concerning its setup, but also concerning its area of application. The following contents of the chapters are summarized as follows:

1.3 Summary of contents

Chapter 2 examines the quality of an IPI bandwidth selection algorithm for local linearly estimating the diurnal pattern of UHF trade duration data (see Feng, 2014). A two-step fitting procedure of the semiparametric ACD model is described, based on the decomposition of the observations into a deterministic and a stochastic part. The underlying deterministic smooth scale function is estimated nonparametrically with an automatically selected bandwidth by the data-driven IPI algorithm. The data is then standardized by the estimated scale function and a likely stationary data series is obtained to which the Exponential ACD (EACD) model is fitted. The overall aim of this chapter is to discuss the practical performance of the IPI algorithm taking into account various bandwidth influencing factors and to compare the results with the commonly used cubic spline method (Engle and Russell, 1998). For this purpose a large simulation study was carried out, with sets of different sample sizes, trends and variances. The main results of the simulation study show, that 1) the Semi-ACD model is clearly superior to the parametric ACD model, thus not taking into account the diurnal pattern before fitting the model is a severe misspecification. 2) A best combination of bandwidth influencing factors, i.e. of inflation method, inflation factor and coefficient for calculating the lag-window estimator of the sum of autocovariances is identified and 3) the proposed Semi-ACD model outperforms the cubic spline method.

Chapter 3 describes a semiparametric extension of the first-type Log-ACD model of Bauwens and Giot (2000) and Bauwens et al. (2008). Basic properties of the parametric part as a general linear process are discussed. A special case for the residuals following an ARMA model with log-normally distributed innovations are considered, as well. Conditions are derived for conditional distributions other than the log-normal one for weak stationarity of the standardized data and their ability to model heavy tails at different levels in the data. The scale function is estimated by local polynomial regression via a data-driven IPI algorithm with automatic bandwidth selection. To further automatize the IPI algorithm two parametric and one nonparametric method for estimating the variance factor are introduced into the IPI. They are practically compared and evaluated via real financial non-negative data examples. The Semi-Log-ACD and the Semi-ACD model are fitted to data examples and due to its flexibility and ease of estimating the scale function the Semi-Log-ACD is recommended over the Semi-ACD model. A possible effect of estimation errors in the semiparametric part on the parametric estimation is investigated and conditions are found under which the parametric estimation is $\sqrt{T}$-consistent.

The Semi-Log-ACD model of chapter 3 is taken up again in chapter 4. The nonparametric trend estimation methods discussed in the previous chapters are combined with bootstrap methods for forecasting different non-negative financial data. The trend estimation, as well as the actual forecasting methods do not require a prior assumption on the distribution for the innovations. A special case of the Semi-Log-ACD model under a conditional log-normal distribution is discussed, as well. The observations of this model are forecast using Kalman filters, in order to compare the results of ACD and ARMA model forecasts. The forecasts obtained with the Semi-ACD model and the Semi-Log-ACD model with and without a conditional distribution assumption are applied to different data-types of six firms. The obtained point and 90% interval forecasts are evaluated with training data sets for ten different forecast horizons with two criteria each. The results show that the semiparametric methods are on average more stable and more precise than the parametric methods. The Semi-Log-ACD model with the conditional log-normal distribution gives the best forecasts of the methods proposed.

In chapter 5 the IPI is applied in a non-financial context. Developing economies are of major importance for global macroeconomic development. However, the empirical analysis and especially the forecasting of macroeconomic time series remain difficult due to a lack of sufficient data, data frequency, high volatility, and often highly non-linear developments. The data-driven local linear trend estimation with an extended IPI algorithm for determining the bandwidth endogenously is applied for forecasting the GDP of developing economies. This approach allows a smooth trend estimation that takes care of temporary changes in trend processes which can

often be found in macroeconomic time series. Further, for forecasting, the naïve random walk model is extended by this local linear, temporarily adjusting trend. Six selected developing economies and two advanced economies are used for showing the practical performance of the proposed methods. Different forecast combinations are evaluated and their results compared. Combining the local linear approach and the random walk with a local linear trend with other forecasting methods improves forecasting accuracy and reduces variance. The overall results of the forecasts seem rather reliable and encourage the use of this methodology.
Chapter 6 discusses further miscellaneous research questions derived from the findings of the different main chapters. The proposals include an idea to use the Semi-Log-ACD model for return series in a GARCH model framework. The algorithm developed in chapter 3 is directly applicable to return data after transformation. A simulation study is carried out to enable a first discussion on the performance. Results on the bandwidth selection, scale function estimation, model parameter estimation and the application of the Semi-Log-GARCH model to real financial return data support this idea for future research. In view of the new BASEL reform, improving the possibilities for estimating the volatility is very valuable. Another approach presented is on a new transformation, which applies to the log-data and allows to improve and refine its normality. It is applied to a data example and the results are in clear favour of this approach. To improve the quality of the growth development forecasts, this chapter also examines whether including neural networks as single methods and in combination with the methods of chapter 5 is suitable. It is found, that these additional methods do not improve the forecasting quality as compared to the results of chapter 5. Further suggestions are included without numerical results. One is on introducing a local bandwidth factor in the IPI to improve the trend estimation by allowing for different bandwidths. Another proposal is on a model parameter estimation approach using estimating functions as an alternative to the (Q)MLE. As a potential possibility for improving the quality of forecasts of non-negative financial data, block bootstrap methods are discussed. Chapter 7 summarizes the results of each chapter and embeds them in the current state of research.
The empirical analyses were carried out with *R*. The following packages were used: *ACDm* (Belfrage, 2016) and *fACD* (Perlin, 2014) for the parametric fitting of ACD models. fGARCH (Wuertz et al., 2017) for estimating GARCH model parameters. *ggpubr* (Alboukadel, 2017) for displaying figures and *forecast* (Hyndman and Khandakar, 2008) to apply NNAR and NNARLL methods for forecasting. *splines* (R-Core Team, 2017) for applying cubic splines method. *xtable* (Dahl, 2016) to import tables obtained in R into a TeX environment. The raw financial data used was retrieved from the Thomson Reuters Tick History data base and processed accordingly.

2 On the iterative plug-in algorithm for estimating diurnal patterns of financial trade durations[5]

2.1 Introduction

Since the introduction of the ACD model in the seminal work of Engle and Russell (1998), the analysis of financial market behaviour based on transaction durations became one of the most important sub-areas of financial econometrics. Numerous extensions of this model are proposed, including the Log-ACD model (Bauwens and Giot, 2000), the class of the augmented ACD models (Fernandes and Grammig, 2006) and the threshold ACD model (Zhang et al., 2001). For further information on the development in this context we refer the reader to Pacurar (2008), Russell and Engle (2010), and in particular the monograph of Hautsch (2011) and references therein.

Intraday trade durations often exhibit a nonstationary deterministic diurnal pattern (or intraday seasonality). The estimation of this component is necessary for further econometric analysis of trade durations using a stationary ACD model. Different approaches are proposed in the literature to deal with this problem. For instance, the use of a cubic spline was originally proposed by Engle and Russell (1998). A nonparametric approach is proposed by Bauwens and Giot (2000). Rodríguez-Poo et al. (2008) proposed to estimate the diurnal duration pattern and the ACD parameters jointly using generalized profile likelihood, which results in a transformed kernel estimator of the nonparametric part. Further approaches for estimating the diurnal duration pattern are e.g. the linear spline (Dufour and Engle, 2000a,b), the wavelet method (Bortoluzzo et al., 2010) and the shrinkage technique (Brownlees and Gallo, 2011). Most recently, Feng (2014) proposed a Semi-ACD (semiparametric ACD) model with a local linear estimator for the diurnal duration pattern and developed

[5]Chapter 2 is published with the same title in the *Journal of Statistical Computation and Simulation*. Apart from the adjustments of references in the text, corrections of spelling mistakes and exclusion of the abstract, nothing has been changed regarding the published paper. It is referred to as Feng et al. (2016) throughout this work.

an iterative plug-in algorithm (Gasser et al., 1991) for selecting the bandwidth. In this chapter we propose to use a design adaptive version of the asymptotically optimal bandwidth obtained by minimizing a partially weighted asymptotic MISE (mean integrated squared error), b_{A}. To reduce the computing time the two required integrals in b_{A} are calculated numerically at some equidistant evaluation points not at all observation points, without affecting the rate of convergence of $\hat{b}_{\mathrm{A}}$. Furthermore, a closed form formula of b_{A} under an EACD(1, 1) model is obtained and employed for assessing the quality of the selected bandwidth in the simulation. The proposed IPI algorithm is applicable under different ACD models, though, because the variance factor in b_{A} is obtained by a nonparametric lag-window estimator (Bühlmann, 1996). The aim of this chapter is to study the practical performance of the IPI bandwidth selector in detail and to compare the data-driven local linear approach with the commonly used cubic spline. For this purpose, a simulation study with two diurnal patterns, two EACD(1, 1) models and three sample sizes is carried out, with 400 replications in each case. For each replication the bandwidth is selected by different sub-methods. Furthermore, the effect of the window-width for calculating the lag-window estimate of the variance factor is also investigated. Cubic spline diurnal pattern estimators with two choices of knots are included for comparison purpose. In order to discuss the effect of the nonparametric estimator on further parametric estimation, EACD(1, 1) models are fitted to the original data and the standardized durations in each case, respectively. The results are then assessed according to the MSE of the selected bandwidth, the goodness-of-fit of the estimated diurnal pattern and the quality of the resulting parameter estimation. The analysis confirmed that the IPI bandwidth selector works very well. In particular, the assessment results following each of these criteria confirmed the consistency of the proposed bandwidth selector and the resulting diurnal pattern estimate. Some findings are: 1) According to the MSE of the selected bandwidth, the best bandwidth selector changes from case to case. 2) According to the goodness-of-fit of the estimated diurnal patterns, a sub-method is found which works almost overall the best. We will hence suggest the use of this sub-method in practice. 3) Concerning the estimation of the scale parameter and that of the latent variable: Empirical efficiency of the estimates using the original data decreases to zero, while that of the estimates using the standardized durations increases very quickly, as the sample size increases. This indicates that the estimates in the former case are clearly wrong but those in the latter case are consistent. Hence, the adjustment of the diurnal pattern is necessary, before a parametric ACD model is fitted. 4) The effect of the diurnal pattern on the estimated parameter of the lagged observation is not as strong as for the other two parameters, because the diurnal pattern is a long-term time-varying component. However, the improvement in this case is still very clear, as the empirical efficiency of the estimated parameters following the Semi-ACD model achieves 100% in many cases. 5) Regarding the cubic spline: As expected, the simulation results show that this approach works very well, if the cubic spline assumption is roughly fulfilled, but its behaviour can be very poor otherwise. Another disadvantage of the cubic spline is that the resulting diurnal pattern is sometimes quite unstable. Some details and the practical relevance of the proposal are illustrated by simulated and real data examples. All of the results show that our proposal usually outperforms the cubic spline.

The model and the estimator are defined in Section 2.2. The bandwidth selector is proposed in Section 2.3. Section 2.4 reports the simulation results. Practical relevance of the proposal is illustrated in Section 2.5 by examples. Concluding remarks in Section 2.6 close the chapter.

2.2 The Semi-ACD model for diurnal durations

Consider the time points $T_o = t_0 < t_1 < ... < t_N < t_{N+1} = T_c$, at which trades occur, where T_o and T_c denote the opening and closing times of a stock market, and N is the (random) number of trades on a trading day. Throughout this chapter we assume that t_i are rescaled trading time points with $T_o = 0$ and $T_c = 1$, except for Section 2.5.2, where the opening and closing times on German financial markets are employed. Let $x_i = t_i - t_{i-1}$ be the durations between two consecutive trades. A commonly used model for x_i (Engle and Russell, 1998) is

$$x_i = \phi(t_i)\psi_i\varepsilon_i, \tag{2.1}$$

where $\phi(t_i)$ is often called a (deterministic) diurnal pattern, ψ_i is the conditional expectation of the diurnally adjusted durations, which follows e.g. some stationary ACD model, and $\varepsilon_i \geq 0$ are i.i.d. random variables with $E(\varepsilon_i) = 1$. Let $y_i = \psi_i\varepsilon_i$. It is assumed that $E(y_i) = 1$ so that the model is uniquely defined, i.e. y_i follows a unit ACD with $E(\psi_i) = 1$. Engle and Russell (1998) propose to specify ψ_i following the idea of the GARCH (generalized autoregressive conditional heteroskedasticity, Engle, 1982 and Bollerslev, 1986) model:

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j y_{i-j} + \sum_{k=1}^{q} \beta_k \psi_{i-k} \tag{2.2}$$

with a standard exponential distribution of ε_i. The restriction $E(y_i) = 1$ implies that $\omega = 1 - \sum_{j=1}^{p} \alpha_j - \sum_{k=1}^{q} \beta_k$. Hence, in a Semi-ACD model the scale parameter ω is no more free.

Note that $\phi(t_i)$ is (approximately) the local mean of x_i. However, x_i and $\phi(t_i)$ depend strongly on N. Under regularity conditions we have indeed $x_i = O_p(N^{-1})$. Hence, it is more convenient to study the deterministic pattern in the rescaled durations $z_i = Nx_i$, because the local mean of z_i is (approximately) a fixed deterministic function. For given N, the estimation of the local mean of z_i is equivalent to that of $\phi(t_i)$. Furthermore, we assume that trades on a day occur according to some design density $0 < f(t) < \infty$ on $t \in [0, 1]$ and define $m(t) = 1/f(t)$ and $\phi_N(t) = m(t)/N$. According to Feng (2014), it holds $E[z_i|N] = m(t_{i-1})[1 + O(N^{-1/2})]$. This ensures that $\phi(t)$ can be equivalently estimated through $\hat{\phi}(t) = \hat{m}(t)/N$.

2.2.1 Local linear estimation of the scale function

Note that x_i and z_i can be rewritten as special nonparametric regression models as follows:

$$x_i = \phi(t_i) + \phi(t_i)(y_i - 1) \tag{2.3}$$

and

$$z_i \approx m(t_i) + m(t_i)(y_i - 1). \tag{2.4}$$

Now, the derivatives $m^{(\nu)}(t)$ can be estimated by minimizing the weighted least squares

$$Q = \sum_{i=1}^{N} \{z_i - a_0(t) - a_1(t)(t_i - t) - ... - a_d(t)(t_i - t)^d\}^2 K\left(\frac{t_i - t}{b}\right), \tag{2.5}$$

where $K(u)$ is a kernel function and b is the bandwidth. We obtain the estimates $\hat{m}^{(\nu)}(t) = \nu!\hat{a}_\nu$, for $\nu \leq d$, and accordingly $\hat{\phi}^{(\nu)}(t) = \frac{\nu!\hat{a}_\nu}{N}$. If we put $d = 1$ and $\nu = 0$, this leads to the local linear estimates $\hat{m}(t) = \hat{a}_0$ and $\hat{\phi}(t) = \hat{a}_0/N$, which will be used in this chapter. The asymptotic properties of $\hat{m}(t)$ and $\hat{\phi}(t)$ are obtained by Feng (2014). Let $\gamma(k)$ denote the autocovariances of y_i and $S = \sum \gamma(k)$ be their sum. Furthermore, let $R(K) = \int K^2(u)du$ and $I(K) = \int u^2 K(u)du$ for a kernel function K. At an interior point $0 < t < T$ the asymptotic variance and asymptotic bias of $\hat{m}(t)$ are given by

$$\text{var}\left[\hat{m}(t)\right] \approx \frac{R(K)S}{Nbf(t)} m^2(t) = \frac{R(K)S}{Nb} m^3(t) \tag{2.6}$$

and

$$B[\hat{m}(t)] = b^2 \frac{m''(t)I(K)}{2}. \tag{2.7}$$

Accordingly, we have $\text{var}\,[\hat{\phi}(t)] \approx \text{var}\,[\hat{m}(t)]/N^2$ and $B[\hat{\phi}(t)] \approx B[\hat{m}(t)]/N$. Based on equations (2.6) and (2.7) the asymptotic mean integrated squared error (AMISE), an approximation of $\text{MISE}(\hat{m}) = \int E\{[\hat{m}(t) - m(t)]^2\}dt$, is given by

$$\text{AMISE}(\hat{m}) = b^4 \frac{\int [m''(t)]^2 dt I(K)}{4} + \frac{R(K)S \int m^3(t)dt}{Nb}. \tag{2.8}$$

By minimizing the AMISE we obtain the asymptotically optimal bandwidth

$$\tilde{b}_{\text{A}} = \left(\frac{R(K)S}{I^2(K)} \frac{I(m^3)}{I([m'']^2)} \right)^{1/5} N^{-1/5}, \tag{2.9}$$

where $I(m^3) = \int m^3(t)dt$ and $I([m'']^2) = \int [m''(t)]^2 dt$. One problem with the above formula is that the $I(m^3)$ term may cause unnecessary instability of the selected bandwidth. To solve this problem we propose to use the following formula of the optimal bandwidth

$$b_{\text{A}} = \left(\frac{R(K)S}{I^2(K)} \frac{I(m^2)}{I([m'']^2)} \right)^{1/5} N^{-1/5}, \tag{2.10}$$

which minimizes the dominating part of the partially weighted MISE $\int \{B[\hat{m}(t)]^2 + f(t)V[\hat{m}(t)]\}dt$. Note that a Semi-ACD model can also be applied to model other financial variables such as daily average durations and daily trade volumes. The formula of b_{A} in (2.10) is design adaptive, i.e. it is the same for equidistant, non-equidistant fixed design as well as for random design. Hence, an algorithm developed based on this formula works for Semi-ACD models in all of these cases. This fact also ensures that many known results on the IPI bandwidth selector with dependent errors can be easily adapted to the one developed in the next section. Furthermore, we will see that by means of this idea the computing time can be reduced clearly without affecting the rate of convergence of the proposed bandwidth selector.

To assess the simulation results, we need to calculate $\tilde{b}_{\text{A}}$ or b_{A} under given design. Note that $R(K)$ and $I(K)$ are two known constants. The terms $I([m'']^2)$, $I(m^2)$ or $I(m^3)$ can also be calculated easily. However, the formula of the sum of $\gamma(k)$ for a given ACD model is still unknown in the literature. In the simulation in Section 2.4, EACD(1, 1) models will be used. In this case, S can be calculated according to the following lemma.

Lemma 2.1 *If y_i follow an EACD(1, 1) with $\epsilon_i \sim \exp(1)$ and $\psi_i = (1 - \alpha - \beta) + \alpha y_{i-1} + \beta \psi_{i-1}$, then the sum of all $\gamma(k)$ of y_i is given by*

$$S = \left(\frac{1 - \beta}{1 - (\alpha + \beta)} \right)^2 \frac{1 - (\alpha + \beta)^2}{1 - (\alpha + \beta)^2 - \alpha^2}. \tag{2.11}$$

The proof of Lemma 2.1 is given in the appendix. Note that the above formula only holds for an EACD(1, 1). More general results will not be discussed here. For a given diurnal pattern and given EACD(1, 1), it can be shown that the difference between $\tilde{b}_\mathrm{A}$ and b_A is quite small. This confirms that the use of b_A is theoretically and practically reasonable.

2.2.2 Estimation of the ACD parameters

Having estimated the scale function and diurnally adjusted the original duration series, an ACD model can be fitted to the diurnally adjusted durations. Let θ denote the vector of the unknown ACD(p, q) parameters, $\theta = (\omega, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)'$. Assume that $\hat{m}(t)$ and $\hat{\phi}(t)$ are consistent estimates of $m(t)$ and $\phi(t)$, then θ can be estimated from $\hat{y}_i = x_i/\hat{\phi}(t_i)$ using the QML (quasi maximum likelihood) method under an EACD(p, q) assumption as proposed by Engle and Russell (1998) and Engle (2000). If the type of the distribution of ϵ_i is assumed, fully efficient ML estimates of θ can also be employed. For a detailed description on these topics we refer the reader to Chapter 5.3 of Hautsch (2011) and references therein. The resulting parameter estimate will be denoted by $\hat{\theta}$. Now, assume that $y_i = \psi_i \epsilon_i$ were observable. The parameter vector θ could also be estimated from y_i using the same method. Denote this (practically unavailable) estimate by $\tilde{\theta}$. It is well known that $\tilde{\theta}$ is $\sqrt{N}$-consistent and asymptotically normal. According to the similarity between the GARCH and the ACD models, consistency and asymptotic properties of $\hat{\theta}$ can be obtained following the ideas of Lemma 1 and Theorem 3 in Feng (2004a). These results indicate that $\hat{\theta}$ is also $\sqrt{N}$-consistent and asymptotically normal up to a bias term. Following the proof of Theorem 3 in Feng (2004a), we can see that this bias term is the same for kernel and local linear estimates of $m(t)$. Moreover, it is easy to see that this conclusion does not depend on N. Hence we have $B(\hat{\theta}) = E[\hat{\theta} - \tilde{\theta}] = O[b^2 + (Nb)^{-1}]$, where the $O(b^2)$ term is due to the integral of the bias $E[\hat{m}(t) - m(t)]$ and the $O[(Nb)^{-1}]$ term is caused by the variance of $\hat{m}(t)$. If a bandwidth $b = O(n^{-a})$ with $1/4 < a < 1/2$ is used, this bias term is asymptotically negligible. If a bandwidth of the optimal order $O(b_\mathrm{A})$ is used, we have $B(\hat{\theta}) = O(N^{-2/5})$. Furthermore, if x_i follow a parametric ACD model with $\phi(t)$ to be a constant, then $\hat{\theta}$ is $\sqrt{N}$-consistent and asymptotically normal, if b is of a larger order than $O(N^{-1/2})$. This is particularly true, when b is selected by the proposed data-driven algorithm in the next section. This means that the Semi-ACD model also works well in the case when the data do not have a diurnal pattern, but with some loss in the efficiency. Proofs of those results are omitted to save space.

For the practical implementation we propose to fit an EACD(1, 1) or another suitable ACD model to $\hat{y}_i$ using the fACD package in R. Other available ACD packages in the literature can also be employed for this purpose. As in the parametric case, model selection using the AIC or BIC can also be applied to $\hat{y}_i$.

2.3 The bandwidth selection procedure

The IPI bandwidth selector to be proposed extends the original idea of Gasser et al. (1991) in different ways. Let b_0 denote the starting bandwidth. In the j-th iteration, $m''(t)$ will be estimated using the bandwidth b_{2j} calculated from b_{j-1}, the bandwidth selected in the $(j-1)$-th iteration. The formula for calculating b_{2j} from b_{j-1} is called the inflation method. Gasser et al. (1991) used the following MIM (multiplicative inflation method) inflation form

$$b_{2j} = b_{j-1} N^{1/10}. \tag{2.12}$$

On the other hand, Beran and Feng (2002a) proposed to use a faster EIM (exponential inflation method) inflation form

$$b_{2j} = b_{j-1}^{\lambda}, \tag{2.13}$$

where $0 < \lambda < 1$ denotes the inflation factor, which determines the rate of convergence of $\hat{b}_A$.

Assume that the MIM or the EIM with a suitable value of λ is used and that $\hat{S}_j$ is calculated from $\hat{\gamma}(k)$ using the Bartlett window $w_k = 1 - k/(L+1)$ with $L = c_f N^{1/3}$. Let $\sqrt{N} < M < N$ be an odd integer. Define $t_r^* = (r-1)/(M-1)$ to be M equidistant evaluation points, and $m_1 = [0.05 * (M-1)]$ and $m_2 = [0.95 * M]$, where $[\cdot]$ denotes the integer part. The proposed IPI algorithm processes as follows:

Step 1a. In the j-th iteration estimate $\hat{m}_j(t_r^*)$, $r = 1, ..., M$, by b_{j-1}. Calculate $\hat{I}_j(m^2) = \{\sum_{r=m_1}^{m_2} [\hat{m}_j(t_r^*)]^2\}/(m_2 - m_1 + 1)$ and $\hat{y}_{ji} = x_i/\hat{\phi}_j(t_i)$, $i = 1, ..., N$. Then calculate $\hat{\gamma}_j(k)$ from $\hat{y}_{ji}$ and obtain $\hat{S}_j = \sum_{|k|<L} w_k \hat{\gamma}_j(k)$.

Step 1b. Calculate b_{2j} using the chosen method, estimate $\hat{m}_j''(t_r^*)$ by local cubic regression and calculate $\hat{I}_j([m'']^2) = \{\sum_{r=m_1}^{m_2} [\hat{m}_j''(t_r^*)]^2\}/(m_2 - m_1 + 1)$.

Step 2. Insert the values of $\hat{I}_j(m^2)$, $\hat{S}_j$ and $\hat{I}_j([m'']^2)$ into (2.10) to obtain b_j.

Step 3. Increase j by one and repeatedly carry out Steps 1 and 2 until convergence or until a given number of iterations is reached. Put $\hat{b} = b_j$.

The idea to estimate $\hat{m}_j(t)$ and $\hat{m}_j''(t)$ only at M evaluation points will reduce the computing time very clearly. In particular note that $\hat{m}_j''(t)$ is a third order local polynomial estimator, which has to be carried out in each iteration. This simplification will not affect the rate of convergence of $\hat{b}$, if $M > \sqrt{N}$, because the highest rate of convergence of an IPI bandwidth selector in the current context is of the order

$O(N^{-2/7})$. Our empirical experience shows that bandwidths selected by different M values are almost the same. In the simulation in the next section, $M = 201$ is fixed to ensure that the large simulation can be finished in an adequate time. Note that even for the smallest sample size there, i.e. $N = 8000$, M is just about 2.5% of the whole observation time points. Our simulation results show that this simplification works very well in practice. Although it is well known that local polynomial regression has automatic boundary correction, the curve estimation quality at a boundary point is still worse than that at an interior point. This problem was dealt with in two ways. Firstly, at any boundary point, the total bandwidth used is kept to be the same as at an interior point. For instance, for a given bandwidth b, the estimation at a point $t < b$ is carried out with all observations within the interval $t_i \in [0, 2b]$. Secondly, the integrals $\hat{I}_j(m^2)$ and $\hat{I}_j([m'']^2)$ are calculated without the 5% estimates at each boundary to avoid their effect on the bandwidth selection.
For calculating the standardized durations in the j-th iteration, $\hat{\phi}_j(t_i)$ are obtained from $\hat{m}_j(t_r^*)$ by means of linear interpolation. At the beginning, we propose to fix $b_0 = 1/10$, so that a rather large number of observations, i.e. 20% of the observations, is used for estimating the scale function in the first iteration. In general, the finally selected bandwidth does not depend on b_0, if set to any reasonable value, because the IPI algorithm is a fix-point search procedure. It can be shown that with the above starting bandwidth b_j will become a consistent estimator of b_{A} in a few iterations. The choice of the inflation method is more important. In this chapter we will mainly consider the use of the EIM. Although, as shown by Beran and Feng (2002a), an inflation factor of $\lambda = 5/7$ will lead to the highest $O(N^{-2/7})$ convergence rate, Feng (2014) proposed the use of $\lambda = 1/2$ so that $\hat{b}$ is most stable but with a lower rate of convergence of the order $O(N^{-1/5})$, because the variation of the intraday durations is very large. This idea was confirmed by the simulation results in the next section. A further choice of λ is $\lambda = 5/9$ to minimize the MSE of $\hat{m}''$ with a rate of convergence of the order $O(N^{-2/9})$. A bandwidth selected by the MIM of Gasser et al. (1991) is also most stable with the rate of convergence of the order $O(N^{-1/5})$. The error in $\hat{S}$ will cause an additional error term in $\hat{b}/b_{\mathrm{A}}$ of the order $O_p(N^{-1/3})$, which is asymptotically negligible. This fact is not affected by the choice of c_f. For practical application, we propose to use $\hat{m}(t_i)$ and $\phi(t_i)$ obtained by using the selected bandwidth $\hat{b}$ at all observation points t_i as the final estimates.

2.4 The simulation study

In the simulation study different cases were constructed to examine the practical performance of the bandwidth selector in detail and to see, whether a relatively better combination of the control parameters exists and how the algorithm can be further improved.

2.4.1 Description of the simulation study

Firstly, two diurnal patterns, $m_1(t)$ and $m_2(t)$, were chosen, where $m_1(t)$ exhibits a typical inverse U-shape and $m_2(t)$ shows an atypical duration pattern with long durations in the morning and afternoon and comparatively short durations around noon. These two patterns are displayed in Figure 2.1, which were indeed designed based on the estimated diurnal duration patterns of the BMW stocks on two trading days in August 2011 (see Feng, 2014). The closed function forms are very complex and are hence omitted. For each diurnal pattern, data were generated using two EACD(1, 1) models with:

$$\text{ACD}_1 : \psi_{1i} = 0.04 + 0.09x_{i-1} + 0.87\psi_{1i-1} \qquad (2.14)$$

and

$$\text{ACD}_2 : \psi_{2i} = 0.04 + 0.14x_{i-1} + 0.82\psi_{2i-1} \qquad (2.15)$$

with $\omega = 1 - \alpha - \beta$. The simulation was carried out with three different sample sizes $N_1 = 8000$, $N_2 = 16000$ and $N_3 = 32000$. The combinations of pattern, model and sample size define 12 main cases of the simulation in total. For each main case 400 replications were generated.[6] Local linear estimators with bandwidth selection using four inflation methods, i.e. the MIM and the three EIM with $\lambda = 5/7, 5/9$ and $1/2$, are considered. For each method, $\hat{S}$ was then calculated with five values of c_f, namely $c_f = 2$, 4, 6, 8 and 10, respectively. Hence, for each replication the bandwidth was selected by 20 sub-methods separately. The weighting function used in (2.5) is the bi-square kernel. Furthermore, two cubic splines (called CSA and CSB) with $\text{Knots}_\text{A} = \{0, 2/17, 4/17, ..., 16/17, 1\}$ and $\text{Knots}_\text{B} = \{0, 1/17, 2/17, ..., 15/17, 16/17, 1\}$ are included as comparisons. Knots_A is chosen following Engle and Russell (1998), which corresponds to those set on each hour on German financial markets, except for the last one. Knots_B, corresponding to those set on each half an hour, is chosen to discuss the effect of the choice of knots.

[6]For a few replications in some main cases the estimated diurnal pattern at some boundary points or some parameter estimates in the second stage were negative. Replications with this problem were replaced by the next simulated data set until 400 replications were carried out successfully.

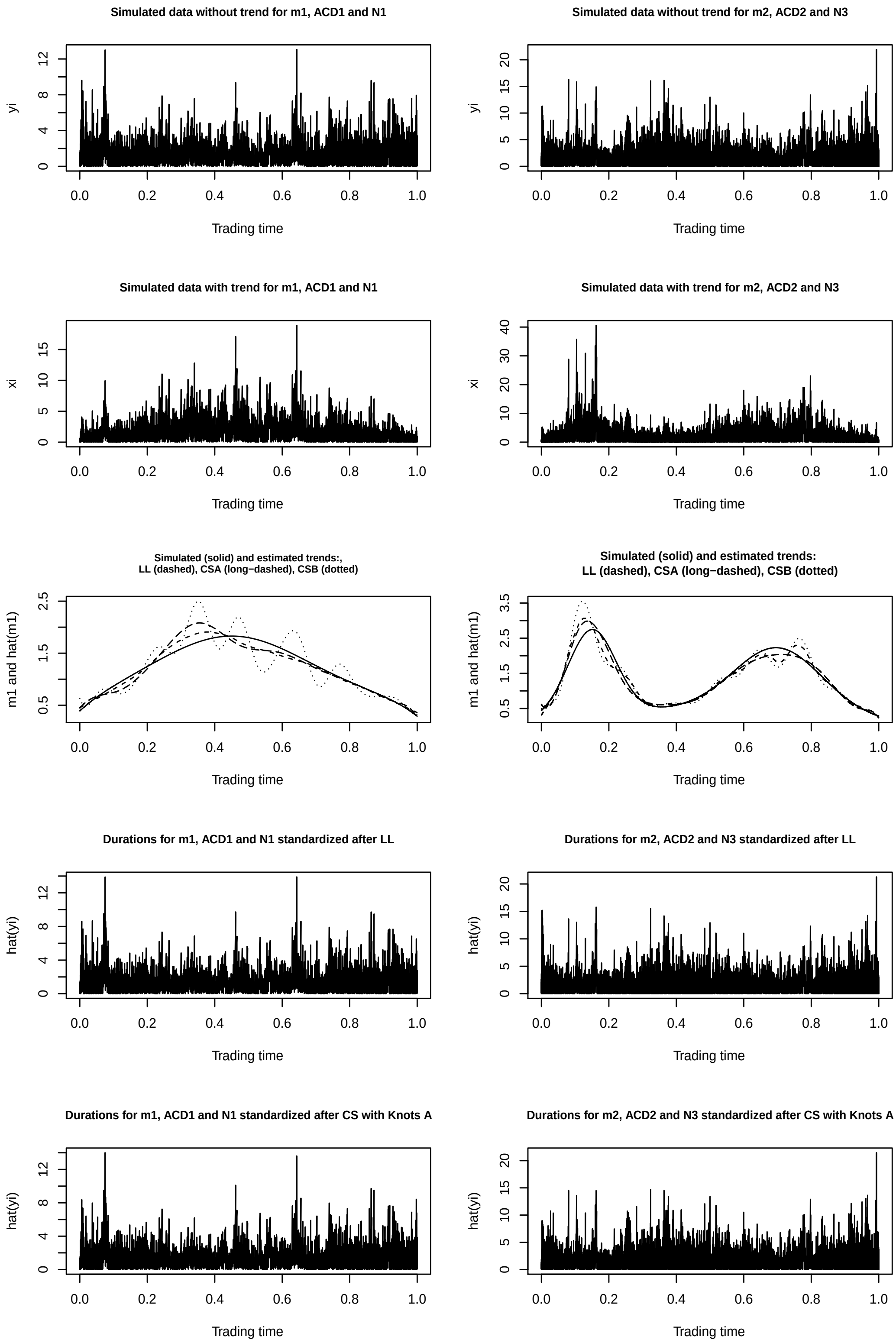

Figure 2.1 – Estimation results for the selected examples Case 111 (left) and Case 223 (right).

2.4.2 Performance of the selected bandwidth

The quality of the bandwidth selection is first discussed according to its bias, variance and MSE, and then assessed by the goodness-of-fit, i.e. the corresponding MSEs of the estimated diurnal patterns using the selected bandwidths. Finally, the simulation results are evaluated by the quality of the estimated ACD parameters in each case. Tables 2.1 to 2.3 show the means, standard deviations as well as the mean squared errors of the bandwidths (multiplied by 100) selected in the 400 replications in the corresponding sub-cases for $N_1 = 8000$, $N_2 = 16000$ and $N_3 = 32000$, respectively, together with the true values of b_{A} calculated following Lemma 2.1 (also multiplied by 100). Firstly, we can see that the MSE decreases strongly, when N increases, which indicates that the proposed bandwidth selector is consistent. It is clear that the performance of $\hat{b}$ depends on the form of the diurnal pattern and the properties of the ACD model very strongly. It is the easiest to select the bandwidth for the second diurnal pattern with the first ACD model, while the bandwidth is very difficult to select in the combination of the first diurnal pattern with the second ACD model. In the former case, the bandwidth can already be selected very well with $N_1 = 8000$. In the latter case the quality of the selected bandwidth with $N_2 = 16000$ is still not good enough. Furthermore, we can see that, if the bandwidth is difficult to select, the effect caused by increasing the sample size is usually more clear. Another more important question, that was to be addressed is whether an overall superior inflation method for selecting a bandwidth can be identified. If the trend is simple, the results suggest to apply the EIM with $\lambda = 1/2$ but for large sample sizes the MIM also works well. If the trend is more complicated no clear statement can be made on which inflation method is generally superior to the others, as it seems to depend on the features of the ACD model as well as the number of observations.

Concerning the choice of c_f, the results are ambiguous, as well. For all cases with first trend and sample size N_1 the optimal c_f is 6. For N_2 one optimal c_f cannot be clearly identified, however $c_f = 8$ and $c_f = 10$ can be ruled out to be optimal. For N_3 as well as almost of the cases simulated with the second trend the majority of smallest MSE values are achieved by $c_f = 2$. Thus, it is not possible to find an overall superior choice of c_f. But it seems that the performance of a moderate c_f is more stable. Hence we will propose to use $c_f = 4$, 6 or 8. If N is large enough, $c_f = 2$ can also be chosen.

Table 2.1 – Statistics from the 400 replications for all cases with $N_1 = 8000$

				EIM, $\lambda = 5/7$			EIM, $\lambda = 5/9$			EIM, $\lambda = 1/2$			MIM		
$m(t)$	ACD	c_f	b_A	mean	SD	MSE	mean	SD	MSE	mean	SD	MSE	mean	SD	MSE
		2		10.51	2.24	34.85	13.95	1.69	6.94	14.40	1.50	4.70	13.58	2.15	10.36
		4		11.33	2.37	27.21	14.71	1.79	4.82	15.17	1.58	3.14	14.54	2.17	6.79
1	1	6	15.98	11.49	2.45	26.12	14.91	1.84	4.54	15.40	1.62	2.96	14.75	2.20	6.33
		8		11.43	2.51	26.95	14.94	1.89	4.62	15.44	1.67	3.07	14.77	2.25	6.52
		10		11.24	2.58	29.07	14.88	1.95	5.00	15.37	1.72	3.34	14.69	2.32	7.03
		2		10.46	2.87	78.77	15.07	2.69	21.55	15.84	2.45	15.14	14.49	3.20	29.38
		4		11.39	3.10	65.40	16.01	2.76	15.73	16.69	2.58	11.36	15.71	3.24	20.44
1	2	6	18.86	11.52	3.19	64.04	16.26	2.83	14.74	16.96	2.74	11.11	15.96	3.29	19.24
		8		11.49	3.28	65.06	16.32	2.90	14.84	17.03	2.81	11.23	16.00	3.36	19.47
		10		11.36	3.36	67.62	16.29	2.94	15.25	17.01	2.87	11.64	15.95	3.43	20.19
		2		6.22	0.63	0.76	6.60	0.76	0.63	7.11	0.91	0.90	6.52	0.72	0.61
		4		6.56	0.70	0.57	6.93	0.94	0.89	7.56	1.19	1.93	6.83	0.87	0.76
2	1	6	6.83	6.60	0.83	0.75	6.96	0.95	0.92	7.63	1.21	2.11	6.87	0.92	0.84
		8		6.53	0.89	0.88	6.91	1.01	1.02	7.60	1.29	2.25	6.80	0.94	0.88
		10		6.42	0.97	1.10	6.81	1.00	1.00	7.52	1.41	2.47	6.72	1.04	1.10
		2		7.08	1.27	2.59	8.34	2.58	6.72	9.16	2.75	8.79	7.85	2.35	5.55
		4		7.60	1.72	3.17	8.84	2.86	8.81	9.73	2.94	11.41	8.34	2.81	7.99
2	2	6	8.06	7.73	2.00	4.10	9.03	3.22	11.33	9.93	3.24	13.99	8.49	3.08	9.67
		8		7.71	2.13	4.66	9.06	3.42	12.65	9.97	3.48	15.75	8.46	3.20	10.39
		10		7.64	2.32	5.55	9.05	3.64	14.18	9.93	3.61	16.48	8.41	3.28	10.90

Table 2.2 – Statistics from the 400 replications for all cases with $N_2 = 16000$

				EIM, $\lambda = 5/7$			EIM, $\lambda = 5/9$			EIM, $\lambda = 1/2$			MIM		
$m(t)$	ACD	c_f	b_A	mean	SD	MSE	mean	SD	MSE	mean	SD	MSE	mean	SD	MSE
		2		10.73	1.65	12.85	13.16	1.13	1.83	13.52	0.96	1.08	13.28	1.28	2.03
		4		11.46	1.74	9.02	13.81	1.20	1.45	14.12	1.03	1.10	13.97	1.29	1.66
1	1	6	13.91	11.61	1.79	8.47	13.99	1.25	1.56	14.29	1.07	1.29	14.13	1.34	1.85
		8		11.62	1.82	8.52	14.02	1.29	1.68	14.33	1.11	1.40	14.16	1.38	1.97
		10		11.58	1.83	8.77	14.00	1.34	1.80	14.32	1.14	1.46	14.15	1.42	2.08
		2		10.97	2.35	35.26	14.28	1.79	7.76	14.77	1.71	5.61	14.57	2.04	7.55
		4		11.74	2.48	28.00	15.08	1.86	5.24	15.50	1.79	4.06	15.38	2.02	5.17
1	2	6	16.42	11.94	2.56	26.55	15.29	1.91	4.94	15.71	1.84	3.89	15.56	2.05	4.91
		8		11.95	2.59	26.68	15.33	1.95	4.98	15.76	1.87	3.95	15.59	2.09	5.03
		2		5.59	0.35	0.26	5.82	0.20	0.05	6.07	0.24	0.07	5.93	0.23	0.05
		4		5.85	0.38	0.15	6.04	0.25	0.07	6.39	0.57	0.51	6.17	0.44	0.25
2	1	6	5.95	5.88	0.40	0.17	6.09	0.53	0.30	6.44	0.60	0.61	6.20	0.52	0.33
		8		5.84	0.43	0.19	6.05	0.46	0.22	6.41	0.59	0.57	6.16	0.51	0.31
		10		5.77	0.46	0.24	6.00	0.56	0.31	6.36	0.73	0.70	6.09	0.39	0.17
		2		6.51	1.04	1.34	7.19	1.61	2.62	7.79	1.77	3.73	7.08	1.60	2.56
		4		6.93	1.32	1.74	7.51	1.66	3.00	8.19	1.70	4.27	7.35	1.63	2.75
2	2	6	7.02	7.04	1.47	2.17	7.66	1.95	4.22	8.39	2.28	7.07	7.44	1.87	3.67
		8		7.02	1.49	2.22	7.70	2.26	5.56	8.43	2.40	7.74	7.51	2.21	5.11
		10		7.00	1.68	2.80	7.66	2.30	5.70	8.38	2.40	7.60	7.49	2.28	5.42

Table 2.3 – Statistics from the 400 replications for all cases with $N_3 = 32000$

				EIM, $\lambda = 5/7$			EIM, $\lambda = 5/9$			EIM, $\lambda = 1/2$			MIM		
$m(t)$	ACD	c_f	b_A	mean	SD	MSE	mean	SD	MSE	mean	SD	MSE	mean	SD	MSE
		2		10.11	1.20	5.42	12.02	0.75	0.57	12.34	0.61	0.42	12.26	0.72	0.54
		4		10.62	1.24	3.75	12.53	0.78	0.78	12.83	0.64	0.92	12.78	0.69	0.93
1	1	6	12.11	10.76	1.26	3.42	12.67	0.80	0.96	12.96	0.66	1.16	12.93	0.69	1.15
		8		10.79	1.28	3.38	12.72	0.82	1.04	13.01	0.67	1.26	12.97	0.71	1.25
		10		10.77	1.29	3.44	12.72	0.83	1.07	13.01	0.69	1.29	12.98	0.72	1.28
		2		10.85	2.00	15.84	13.58	1.35	2.32	13.97	1.23	1.63	13.92	1.33	1.90
		4		11.49	2.03	11.99	14.18	1.43	2.05	14.55	1.30	1.75	14.49	1.31	1.75
1	2	6	14.29	11.68	2.08	11.16	14.34	1.47	2.15	14.71	1.33	1.93	14.67	1.30	1.83
		8		11.72	2.12	11.10	14.40	1.49	2.24	14.77	1.34	2.03	14.71	1.33	1.94
		10		11.70	2.15	11.32	14.41	1.52	2.33	14.78	1.37	2.11	14.70	1.36	2.02
		2		5.03	0.24	0.08	5.24	0.15	0.03	5.32	0.17	0.05	5.37	0.16	0.07
		4		5.23	0.26	0.07	5.39	0.17	0.07	5.51	0.21	0.16	5.55	0.18	0.17
2	1	6	5.18	5.26	0.28	0.08	5.41	0.19	0.09	5.55	0.23	0.19	5.57	0.19	0.19
		8		5.24	0.29	0.09	5.40	0.20	0.09	5.54	0.25	0.20	5.56	0.21	0.19
		10		5.21	0.31	0.09	5.37	0.21	0.08	5.52	0.26	0.18	5.53	0.20	0.17
		2		5.83	0.51	0.34	6.16	0.70	0.49	6.47	0.68	0.60	6.23	0.73	0.55
		4		6.14	0.75	0.56	6.38	0.73	0.60	6.80	0.87	1.22	6.43	0.72	0.62
2	2	6	6.11	6.22	0.89	0.81	6.44	0.88	0.88	6.89	1.11	1.84	6.49	0.98	1.10
		8		6.21	0.94	0.89	6.47	1.02	1.16	6.93	1.27	2.28	6.48	0.96	1.06
		10		6.18	0.94	0.89	6.44	1.03	1.17	6.89	1.21	2.06	6.46	1.01	1.14

2.4.3 Goodness of fit of $\hat{m}(t)$

To assess the goodness-of-fit of the data-driven estimate of the diurnal pattern directly, we will define the RASE (the root of the average of the averaged squared errors) as follows. For a given diurnal pattern and sample size, the ASE for the j-th replication is defined by

$$\text{ASE}_j = \frac{1}{0.9N} \sum_{k=0.05N+1}^{0.95N} (\hat{m}(t_k) - m(t_k))^2, \tag{2.16}$$

where again 5% estimates at each boundary are not used for calculating this criterion. The RASE is then defined as the root of the average of ASE_j over all 400 replications:

$$\text{RASE} = \sqrt{\frac{1}{400} \sum_{j=1}^{400} \text{ASE}_j}. \tag{2.17}$$

Results of RASE (multiplied by 100) for all sub-methods mentioned above, as well as CSA and CSB are displayed in Table 2.4. An important empirical finding is that these results indicate a clear order of the goodness-of-fit of the four methods for calculating b_{2j}. Now the sub-method EIM with $\lambda = 1/2$ performs the best overall. The MIM sub-method is the second best one and the EIM with $\lambda = 5/7$ is the worst. These results also suggest that $c_f = 2$ should not be used. For the best sub-method, the difference between the results with the other values of c_f is unclear, although $c_f = 6$ or $c_f = 8$, or sometimes $c_f = 10$, is usually the best. Note that the main purpose of nonparametric estimation of the diurnal pattern is to fit $m(t)$ as well as possible. Also note that the MIM was proposed to achieve a most stable bandwidth. Our simulation results seem to indicate that the stability of the bandwidth selection is more important than the rate of convergence of the bandwidth itself. The simulation results indicate further that, following the RASE criterion, a relatively larger value of c_f is more preferable. This shows again that the stability of the selected bandwidth plays a very important role for the goodness-of-fit of the resulting curve estimation. Hence, we will suggest the use of the EIM with $\lambda = 1/2$ and $c_f = 6$. In the following, this special sub-method will be simply called the LL estimator. Furthermore, these results indicate that the estimation of $m_1(t)$ under ACD_1 is the easiest, while the estimation of $m_2(t)$ under ACD_2 is most difficult. Finally, conclusions obtained following the RASE are quite different to those drawn from the MSE of $\hat{b}$.

Hereafter, the behaviour of the cubic spline will be compared with that of the LL estimator. Note that a local cubic function with locations defined by the knots is used in the former and a local linear with a total bandwidth of $2^*\hat{b}$ in the latter. Hence, the distance between two sequential knots plays a role similar to the one of the total bandwidth, up to possible effects caused by different polynomial orders and further different features between both approaches. The key difference is that the knots in CSA and CSB are fixed beforehand, but the bandwidth of the LL estimator is selected by a data-driven algorithm. It is clear that the cubic spline works very well, if the specification is roughly fulfilled by the underlying trend. Otherwise, its performance is usually not good and can even be very poor, if the distance between the knots is too far from the required bandwidth. These facts are confirmed by the simulation results: We see, in the six atypical (main) cases (with $m_2(t)$) that the performance of CSA is clearly better than that of the LL estimator, because now the distance between each two sequential knots is very close to 2^*b_{A}. Hence, CSA is really a reasonable choice sometimes. But in the six typical cases (with $m_1(t)$), the LL estimator performs clearly better than CSA, because the estimation of $m_1(t)$ requires a much larger bandwidth. In all of the 12 main cases, the performance of CSB is much poorer than that of the other two approaches. Note in particular that CSB performs in all six cases with $m_1(t)$ very poorly, because the distance between the knots is now too far from the total optimal bandwidth. Finally, note that an atypical diurnal duration pattern like $m_2(t)$ occurs in practice rarely and $\hat{b}$ is often close to b_{A} for $m_1(t)$. See the data examples in Section 2.5.2 or in Feng (2014). This indicates that in practice the LL estimator will usually outperform CSA.

Table 2.4 – RASE from the 400 replications for all simulated cases ($m_k(t)$, ACD_s, N_i)

λ	InfM	c_f	(1,1,1)	(1,1,2)	(1,1,3)	(1,2,1)	(1,2,2)	(1,2,3)	(2,1,1)	(2,1,2)	(2,1,3)	(2,2,1)	(2,2,2)	(2,2,3)
		2	16.52	9.82	6.99	35.39	19.63	14.81	23.05	17.15	12.02	39.33	28.83	22.81
		4	15.79	9.42	6.79	33.95	18.65	14.01	22.39	16.71	11.77	36.86	27.26	21.56
5/7	EIM	6	15.73	9.35	6.75	33.87	18.50	13.85	22.28	16.67	11.73	35.46	26.87	21.42
		8	15.86	9.37	6.74	34.10	18.57	13.95	22.43	16.75	11.75	35.54	26.94	21.47
		10	16.19	9.37	6.74	34.60	18.75	14.05	22.61	16.85	11.80	36.01	27.11	21.57
		2	12.44	8.31	6.23	22.48	13.73	9.67	21.79	16.42	11.69	30.06	24.59	19.42
		4	12.05	8.08	6.10	21.54	13.26	9.42	21.18	16.09	11.53	28.75	23.97	18.96
5/9	EIM	6	11.93	8.03	6.07	21.33	13.20	9.37	21.15	15.97	11.50	28.69	23.72	18.89
		8	11.88	8.02	6.06	21.20	13.14	9.35	21.21	16.01	11.51	28.44	23.75	18.78
		10	11.92	8.03	6.06	21.21	13.16	9.36	21.38	16.09	11.54	28.62	23.89	18.83
		2	11.87	8.13	6.13	20.99	13.37	9.36	20.76	16.03	11.60	28.05	23.31	18.64
		4	11.50	7.93	6.02	20.54	12.95	9.11	20.02	15.55	11.39	27.25	22.57	18.02
1/2	EIM	6	11.35	7.88	5.99	20.03	12.84	9.04	19.91	15.48	11.35	26.84	22.47	17.93
		8	11.33	7.87	5.98	19.66	12.82	9.02	19.94	15.52	11.35	26.86	22.46	17.75
		10	11.41	7.87	5.98	19.64	12.83	9.02	20.3	15.60	11.38	27.03	22.49	17.94
		2	13.17	8.30	6.17	26.58	13.65	9.54	22.20	16.38	11.55	32.09	28.31	19.65
		4	12.63	8.05	6.3	24.17	13.13	9.27	21.54	15.93	11.37	30.62	24.58	19.14
—	MIM	6	12.46	8.00	5.80	23.89	13.03	9.19	21.43	15.89	11.35	30.24	24.66	19.06
		8	12.48	8.00	5.99	23.90	13.03	9.19	21.67	15.94	11.36	30.25	24.42	19.09
		10	12.60	8.00	5.99	24.04	13.03	9.20	21.72	16.07	11.38	30.49	24.48	19.12
CS (Knots_A)			15.02	10.03	7.22	23.59	16.52	11.49	18.13	13.02	9.39	24.59	19.78	14.53
CS (Knots_B)			22.29	14.66	10.32	36.77	23.43	16.63	24.97	17.86	12.08	40.36	29.25	19.99

Note: $k = 1, 2$ for $m_1(t)$ and $m_2(t)$, $s = 1, 2$ for ACD_1 and ACD_2 and $i = 1, 2, 3$, for $N_1 = 8000$, $N_2 = 16000$ and $N_3 = 32000$, respectively.

2.4.4 Performance of the ACD parameter estimation

For each of the 400 replications of a main case three EACD(1, 1) models were fitted to the duration data simulated without a trend, y_i, the duration data simulated with a trend, x_i, and the diurnally adjusted durations, $\hat{y}_i = x_i/\hat{\phi}(t_i)$. Let θ denote the true parameter vector $\theta = (\omega, \alpha, \beta)'$. Denote by $\tilde{\theta}$, $\hat{\theta}^x$ and $\hat{\theta}^{\hat{y}}$ the estimated parameter vector based on y_i, x_i and $\hat{y}_i$, respectively. For assessing the quality of the parameter estimation, the relative efficiencies (REFF) of $\hat{\theta}^x$ and $\hat{\theta}^{\hat{y}}$ with respect to $\tilde{\theta}$ are defined as follows:

$$\text{REFF}(\hat{\theta}^{\hat{y}}) = \frac{\text{MSE}(\tilde{\theta})}{\text{MSE}(\hat{\theta}^{\hat{y}})} * 100(\%) \text{ and } \text{REFF}(\hat{\theta}^{x}) = \frac{\text{MSE}(\tilde{\theta})}{\text{MSE}(\hat{\theta}^{x})} * 100(\%). \tag{2.18}$$

These results are listed in Tables 2.5 to 2.7 for the three parameters, respectively. Theoretically, if there is a deterministic trend in the data x_i, $\hat{\theta}^x$ is obtained under misspecification and is hence inconsistent. As indicated before, $\hat{\theta}^{\hat{y}}$ is however consistent. These facts can be seen clearly from REFF($\hat{\theta}^{\hat{y}}$) and REFF($\hat{\theta}^x$) and the comparison between them will indicate the gain in parameter estimation by means of the Semi-ACD model.

Some general findings which we can draw from these results are as follows: 1) The larger N, the higher the REFF of the estimated parameters from $\hat{y}_i$ but the lower the REFF of those obtained from x_i. As $N \to \infty$, $\hat{\theta}^{\hat{y}}$ will tend to 100% but REFF($\hat{\theta}^x$)

will however tend to zero. This fact can be seen more clearly, if the estimation of β is considered. See Table 2.7. 2) The quality of $\hat{\omega}^{\hat{y}}$ is the poorest, because ω is the scale parameter and $\phi(t)$ is the scale function. Indeed, the proposed Semi-ACD model can be asymptotically rewritten as an ACD model with only one time varying scale parameter, while its α and β are constant, as in a parametric ACD model. 3) The highest REFFs are achieved by $\hat{\alpha}^{\hat{y}}$, where these efficiencies are about 100% in most cases. Now, the REFFs of $\hat{\alpha}^{x}$ are also high, because α reflects the short term dependence and is not affected by the diurnal pattern so much.

Furthermore, the quality of the parameter estimation based on $\hat{y}_i$ depends on the combination of the diurnal pattern and the ACD model very strongly. The case, where the estimation of ω is the easiest seems to be the combination of $m_1(t)$ with ACD_2. By the combination of $m_2(t)$ and ACD_1, ω is very difficult to estimate. Now the REFF of $\hat{\omega}^{\hat{y}}$ for $N_1 = 8000$ using any sub-method is clearly smaller than 50%. Similar conclusions can be drawn for $\hat{\beta}^{\hat{y}}$. The difference is only that the REFFs of $\hat{\beta}^{\hat{y}}$ are usually clearly higher than those of $\hat{\omega}^{\hat{y}}$ in corresponding cases.
Concerning the difference caused by the sub-methods for the bandwidth selection we can find that the EIM with $\lambda = 1/2$ performs usually the best, except for the combination of $m_2(t)$ and ACD_2. In this case the EIM with $\lambda = 5/7$ performs slightly better than the other methods. However, we will suggest the use of the EIM with $\lambda = 1/2$ again, because it seems to be more stable. Note in particular that by the combination of $m_2(t)$ and ACD_1, the EIM with $\lambda = 5/7$ performs clearly poorer than all of the other methods. This sub-method is hence not a suitable choice. When the sub-method EIM with $\lambda = 1/2$ is chosen, the difference caused by the choice of c_f is usually unclear. In general, all of the c_f values perform well. However, we will still suggest the use of $c_f = 6$ or $c_f = 8$, because now the proposed algorithm performs more stable than with the other c_f values.
In order to compare the performances of the LL and CS estimators based on the REFF, the following simplification is applied. Assuming that the REFF of Method A is greater than that of Method B, we will say that there is an unclear, clear or very clear gain of Method A compared to Method B, if the difference between both REFFs is less than 5, between 5 and 10 or larger than 10 percentage points, respectively. Analogous simplification applies to the loss. Consider the estimation of ω first, where in two of the six typical cases there is a very clear gain of the LL estimator compared to CSA, in two cases the gain is clear and in the other two cases the gain is unclear. In four of the six atypical cases, there is a very clear loss of the LL estimator compared to the CSA and in one of those cases the gain (or loss) is unclear. However, in the last case the LL estimator performs clearly better than CSA. In all of the first eleven cases the gain of the LL estimators compared to CSB

Table 2.5 – Empirical efficiencies (%) of $\hat{\omega}$ from the 400 replications for all cases ($m_k(t)$, ACD_s, N_i)

λ	InfM	c_f	(1,1,1)	(1,1,2)	(1,1,3)	(1,2,1)	(1,2,2)	(1,2,3)	(2,1,1)	(2,1,2)	(2,1,3)	(2,2,1)	(2,2,2)	(2,2,3)
		2	40.1	67.6	78.7	51.2	74.4	87.7	33.4	51.1	60.4	52.5	67.6	82.6
		4	42.9	69.5	79.8	55.4	72.3	88.2	36.1	54.5	62.5	55.1	64.5	81.4
5/7	EIM	6	42.8	69.7	79.0	55.4	72.3	88.6	35.8	54.8	63.2	54.1	62.8	79.9
		8	41.9	69.6	80.2	55.0	72.3	88.8	34.7	54.7	63.6	52.8	52.3	79.0
		10	40.5	69.2	80.0	53.4	72.2	88.7	33.1	52.8	62.5	50.6	60.4	79.1
		2	56.9	74.6	84.3	66.7	75.0	89.6	37.6	55.7	64.1	52.6	59.5	81.5
		4	58.6	75.9	85.4	67.6	75.1	89.8	40.0	58.2	66.1	52.1	58.6	80.7
5/9	EIM	6	58.9	76.2	85.6	67.3	75.2	89.9	39.6	57.6	66.4	51.4	55.8	78.1
		8	59.0	76.3	84.9	67.1	75.2	89.9	38.8	57.6	65.8	50.5	53.6	77.5
		10	58.7	75.9	85.6	67.0	75.1	89.8	37.6	56.4	65.5	49.2	53.7	77.0
		2	58.5	76.3	84.9	67.3	75.3	89.8	42.5	59.0	65.0	54.9	58.6	80.3
		4	60.3	77.2	86.4	67.8	76.0	90.1	45.1	61.6	67.3	55.0	58.2	78.9
1/2	EIM	6	60.8	77.5	86.2	68.8	76.0	90.0	45.0	61.4	67.5	53.9	53.7	74.0
		8	61.0	77.6	86.5	68.9	76.1	90.2	44.1	61.3	67.7	53.4	52.7	71.9
		10	60.4	77.4	86.7	68.8	76.1	90.2	42.4	59.4	67.2	52.5	55.2	72.6
		2	53.3	74.7	85.1	66.2	75.6	89.9	36.8	57.2	65.9	52.1	64.3	81.6
		4	56.2	76.3	85.6	67.3	76.0	89.7	39.7	59.9	68.8	52.6	60.1	83.6
—	MIM	6	56.9	76.4	86.4	67.4	76.1	90.1	39.2	59.8	68.8	51.5	56.4	79.0
		8	56.4	76.1	86.1	67.5	76.0	90.1	38.1	59.3	68.7	50.4	54.5	77.7
		10	54.3	76.1	86.3	67.5	76.2	90.3	36.7	58.2	67.9	49.7	54.5	77.2
CS (Knots_{A})			51.0	72.6	82.7	56.4	75.0	84.2	71.3	77.4	89.9	71.0	55.3	62.4
CS (Knots_{B})			25.2	49.2	67	37.4	65	75.8	36.6	55.9	64.7	50.5	55.7	73.3
x			3.9	2.0	0.85	4.5	2.3	1.0	3.7	1.7	0.87	4.9	1.7	0.80

Table 2.6 – Empirical efficiencies (%) of $\hat{\alpha}$ from the 400 replications for all cases ($m_k(t)$, ACD_s, N_i)

λ	InfM	c_f	(1,1,1)	(1,1,2)	(1,1,3)	(1,2,1)	(1,2,2)	(1,2,3)	(2,1,1)	(2,1,2)	(2,1,3)	(2,2,1)	(2,2,2)	(2,2,3)
		2	100.9	100.4	100.3	99.2	100.0	99.5	95.5	98.7	100.4	96.8	99.6	100.3
		4	101.4	100.5	100.0	99.4	100.0	99.7	95.6	99.1	100.0	95.4	99.0	100.3
5/7	EIM	6	101.4	100.3	99.2	99.4	100.1	99.7	95.3	89.9	99.8	96.7	10.2	100.5
		8	101.4	100.2	100.6	99.4	100.0	99.7	95.5	98.6	100.5	96.9	99.5	100.4
		10	101.1	100.3	100.0	99.4	99.9	99.6	94.8	98.8	100.3	96.6	99.0	100.3
		2	101.0	100.5	99.3	99.3	100.3	100.0	95.6	99.1	100.2	97.5	100.2	100.1
		4	101.0	100.4	99.7	99.9	100.3	100.0	95.3	98.9	99.9	98.1	100.1	100.2
5/9	EIM	6	101.3	100.2	99.8	99.3	100.3	100.1	92.9	98.9	99.7	98.3	100.0	100.1
		8	101.2	100.1	99.4	99.2	100.2	100.0	94.4	98.6	100.6	98.2	100.2	99.9
		10	101.3	100.2	99.5	99.1	100.3	100.1	94.8	98.7	100.1	98.3	100.4	100.0
		2	101.5	100.4	98.7	98.9	99.9	100.1	95.4	98.5	100.3	99.2	99.1	100.1
		4	101.5	100.4	99.2	99.9	100.3	99.9	96.1	98.7	100.2	98.8	100.1	100.2
1/2	EIM	6	101.0	100.3	99.5	99.5	100.2	99.8	95.6	98.6	100.6	99.0	100.1	100.3
		8	101.3	100.2	99.3	99.7	100.2	99.9	96.2	98.8	100.2	99.1	99.8	100.2
		10	101.3	100.2	99.8	99.7	100.2	99.9	97.0	98.6	99.8	98.7	99.7	100.4
		2	101.1	100.1	99.6	99.2	100.1	99.6	96.3	99.0	99.8	95.7	99.7	95.9
		4	101.0	100.1	99.9	99.8	100.4	99.2	96.3	98.8	100.6	98.1	98.9	100.1
—	MIM	6	101.1	100.1	99.5	99.7	100.4	99.8	95.0	98.9	100.3	98.3	100.0	100.4
		8	100.9	100.3	99.5	99.7	100.1	99.9	92.1	98.8	100.3	98.4	100.5	100.2
		10	100.9	100.2	99.9	99.9	100.4	99.9	94.7	98.2	99.9	96.8	99.7	100.1
CS (Knots_{A})			102.3	100.3	99.8	101.0	98.9	99.5	99.5	98.7	99.9	100.2	99.5	98.9
CS (Knots_{B})			99.9	99.8	100.3	97.8	99.1	99.6	80.9	99	100.3	97.6	100.3	99.7
x			88.7	79.9	67.4	54.2	90.5	91.1	90.6	74.5	63.5	91.7	89.8	90.2

Table 2.7 – Empirical efficiencies (%) of $\hat{\beta}$ from the 400 replications for all cases ($m_k(t)$, ACD_s, N_i)

λ	InfM	c_f	(1,1,1)	(1,1,2)	(1,1,3)	(1,2,1)	(1,2,2)	(1,2,3)	(2,1,1)	(2,1,2)	(2,1,3)	(2,2,1)	(2,2,2)	(2,2,3)
		2	68.9	87.7	93.2	82.4	93.4	99.3	60.5	75.6	76.8	84.6	93.6	94.8
		4	70.8	89.0	93.3	84.5	91.6	99.6	62.4	78.0	77.6	85.9	92.6	95.1
5/7	EIM	6	70.7	88.9	91.9	84.6	91.7	99.7	62.1	78.1	78.1	85.2	92.4	94.9
		8	69.9	88.8	94.1	84.5	91.7	99.9	61.2	77.7	79.1	84.7	91.1	94.3
		10	68.9	88.5	93.4	83.7	91.6	99.8	59.4	77.2	78.0	83.2	90.5	94.4
		2	82.0	91.3	95.0	91.8	92.5	100.8	63.8	78.8	78.9	83.3	91.3	96.2
		4	83.0	91.9	95.7	91.9	92.3	100.9	65.5	80.3	80.0	80.9	90.3	96.0
5/9	EIM	6	83.4	91.9	96.1	91.8	92.3	100.8	64.3	80.1	80.0	80.8	89.0	95.0
		8	83.3	91.8	95.1	91.9	92.3	100.8	64.4	80.6	79.9	80.3	88.0	94.7
		10	83.2	91.7	95.7	91.8	92.3	100.8	63.5	79.9	79.5	79.7	88.4	94.4
		2	83.1	92.2	94.5	91.9	92.0	101.0	67.9	80.9	79.5	82.9	89.9	96.0
		4	84.4	92.4	95.6	92.1	92.7	100.7	70.1	83.3	80.8	82.2	90.9	95.9
1/2	EIM	6	84.6	92.5	95.6	92.3	92.6	100.6	69.8	83.0	81.0	81.2	89.0	94.0
		8	84.9	92.5	95.8	92.4	92.7	100.7	69.1	83.1	81.1	81.4	88.5	93.2
		10	84.3	92.4	96.2	92.4	92.7	100.7	67.8	82.1	80.5	80.8	91.0	93.6
		2	79.7	91.3	95.4	90.7	93.1	100.5	63.1	79.5	79.6	83.9	94.9	95.9
		4	81.8	92.0	95.7	91.3	92.9	99.8	66.2	81.2	81.8	84.0	90.2	96.9
—	MIM	6	82.4	92.0	96.0	91.6	93.0	100.5	64.9	81.5	81.6	83.4	88.6	95.6
		8	82.1	91.9	95.6	91.5	92.6	100.7	63.3	81.2	81.7	83.2	87.7	94.7
		10	80.9	91.8	96.1	91.4	93.0	100.7	62.9	80.5	80.8	81.9	87.6	94.5
CS (Knots$_A$)			78.7	90.9	94.5	87.9	92.7	99.6	91.6	91.1	96.9	93.7	81.6	84.0
CS (Knots$_B$)			55.3	79	88.7	80.6	90.9	94.5	64.2	82.6	84	90.0	86.7	90.2
x			10.0	4.7	2.1	4.7	5.9	2.9	9.7	4.3	2.0	11.3	5.3	2.5

is very clear. In the last case the gain of the LL estimator is unclear. As expected, there are no clear differences between the performances of $\hat{\alpha}$ obtained by different approaches. Concerning the quality of $\hat{\beta}$, the gain achieved by the LL estimator compared to CSA is unclear in four of the six typical cases, but in the other two cases there is a clear gain. Now, in three of the six atypical cases there is a very clear loss again, and in one of those cases there is a clear loss of the LL estimator compared to the CSA. In the last two cases, though, there is a clear or even very clear gain of the LL estimator compared to the CSA, again. The performance of $\hat{\beta}$ using CSB compared to that of $\hat{\omega}$ is clearly improved. Now, $\hat{\beta}$ obtained by CSB even outperforms that obtained by CSA in one case, and outperforms that obtained by the LL estimator in another case. The assessment according to the REFF of the parameter estimation is more in favour of the use of the LL estimator, because now it performs in a few atypical cases even better than CSA.

2.5 Application to simulated and real data examples

The practical relevance of the proposed LL estimator with the IPI algorithm will be illustrated and compared with that of the cubic spline by simulated and real data examples.

2.5.1 Estimation results for two simulated data examples

At first, two simulated data sets were chosen, for which the fitted results using the proposed LL estimator, CSA and CSB are shown in more detail. The first example is the first simulated data set in the case with $m_1(t)$, ACD$_1$ and N_1, called Case 111. The second example is the last simulated data set in the case with $m_2(t)$, ACD$_2$ and N_3, called Case 223. The left panels of Figure 2.1 display the simulated data without trend, y_i, the simulated data with trend, x_i, the true trend $m_1(t_i)$ together with the estimates of $m_1(t)$ obtained by the LL estimator, CSA and CSB respectively, and the standardized duration series according to the first two trend estimates for Case 111. Those for Case 223 are shown in the right panels of Figure 2.1.

We see, for Case 111 the LL estimator works a little bit better than CSA, whereas CSA works better than the LL estimator for Case 223. In both examples the trend estimated by each of the two methods fits the true trend well. The standardized durations obtained by both approaches look very close to the simulated data. That is the nonstationarity caused by the trend is well removed. Finally, the fitted ACD models in the two cases are with $\psi_{1i}^{\mathrm{LL}} = 0.038 + 0.082x_{i-1} + 0.879\psi_{1i-1}^{\mathrm{LL}}$ and $\psi_{2i}^{\mathrm{LL}} = 0.040 + 0.139x_{i-1}^{\mathrm{LL}} + 0.820\psi_{2i-1}$, for the LL estimator, and $\psi_{1i}^{\mathrm{CSA}} = 0.039 + 0.082x_{i-1} + 0.878\psi_{1i-1}^{\mathrm{CSA}}$ and $\psi_{2i}^{\mathrm{CSA}} = 0.040 + 0.140x_{i-1} + 0.821\psi_{2i-1}^{\mathrm{CSA}}$ for CSA, respectively. We see, in both examples the estimates of corresponding parameter are always almost the same.

In both cases CSB performs the worst. From Case 111 we see in particular that if the distance between each two sequential knots is (relatively) too small, some random conditional fluctuation of the data may be wrongly estimated as a part of the deterministic trend. This will in turn affect the further parameter estimation strongly. These examples confirm again that the cubic spline only works well, if this specification with given knots is roughly fulfilled. Otherwise its performance can be very poor. Hence, Engle and Russell (1998) proposed the use of CSA assuming that $m(t)$ can be approximated by this cubic spline specification.

2.5.2 Application to some real data examples

In the following, data sets of two German firms BMW and Deutsche Telekom (DTE) on five days from 22. to 26. Sep. 2014 are chosen to compare the practical performance of our proposal to that of the cubic spline. In Germany financial markets open from 9:00 to 17:30. The two choices of knots correspond to Knots$_{\mathrm{A}}$ and Knots$_{\mathrm{B}}$ mentioned in Section 2.4 are now Knots$_{\mathrm{A}}^*$={9:00, 10:00, ..., 16:00, 17:00, 17:30} and Knots$_{\mathrm{B}}$ is Knots$_{\mathrm{B}}^*$={9:00, 9:30, ..., 17:00, 17:30}, respectively. The selected bandwidths for those examples are $\hat{b}$=1.46, 1.62, 1.44, 1.01 and 1.23 for BMW, and $\hat{b}$=1.38, 1.32, 1.31, 1.00 and 1.58 for DTE, respectively.

We see that the selected bandwidths vary strongly from one example to another. The biggest is over 1.5 (hour) and the smallest about one hour. The selected bandwidths correspond to total windows with a length of about 2 to 3.2 hours. These results indicate that CSB does not work at all now. The performance of the LL estimator will hence only be compared to that of CSA. Observations on two selected days from each firm are displayed in the left panels of Figure 2.2, respectively. The fitted trends by the LL and CSA estimators are shown in the corresponding right panels. Those obtained by CSB are clearly unreasonable and are hence not given here. Examples of DTE on the 24. Sep. 2014 and that of BMW on 25. Sep. 2014 are chosen to show that CSA does work well sometimes. We see that in the former example results of both estimators are quite similar. This is also about true for the latter. The example of BMW on the 23. Sep. 2014 is chosen to show the typical difference between diurnal patterns estimated by the LL estimator and CSA. In this case, the two estimated trends are similar to each other during some sub-periods. Otherwise, the trend estimated by CSA varies strongly around the trend obtained by the LL estimator in a random periodic way, which seems to depend on the chosen knots. This feature is shown more clearly by the example of DTE on the 23. Sep. 2014, where the estimated trend by CSA runs in a random periodic way around that of the LL estimator through the whole support, similar to the behaviour of CSB in the simulated example Case 111 discussed in Section 2.5.1. Discussion on the standardized durations and the parameter estimation in the second stage is omitted.
In summary, the simulation results and real data examples show that our proposal outperforms CSA in a general case, although CSA works sometimes quite well. The key difference is that the LL estimator adapts to the features of each data set, whereas CSA with fixed knots is not data-adaptive. The distance between two sequential knots by CSA is sometimes not big enough. Now, some random fluctuations can be wrongly estimated as a part of the diurnal pattern. It can be shown that the use of too few knots, e.g. those set on every two hours, may cause strong bias and a new problem, namely that the estimated trend at the boundary is often negative. Detailed discussion on this topic is beyond the aim of this chapter.

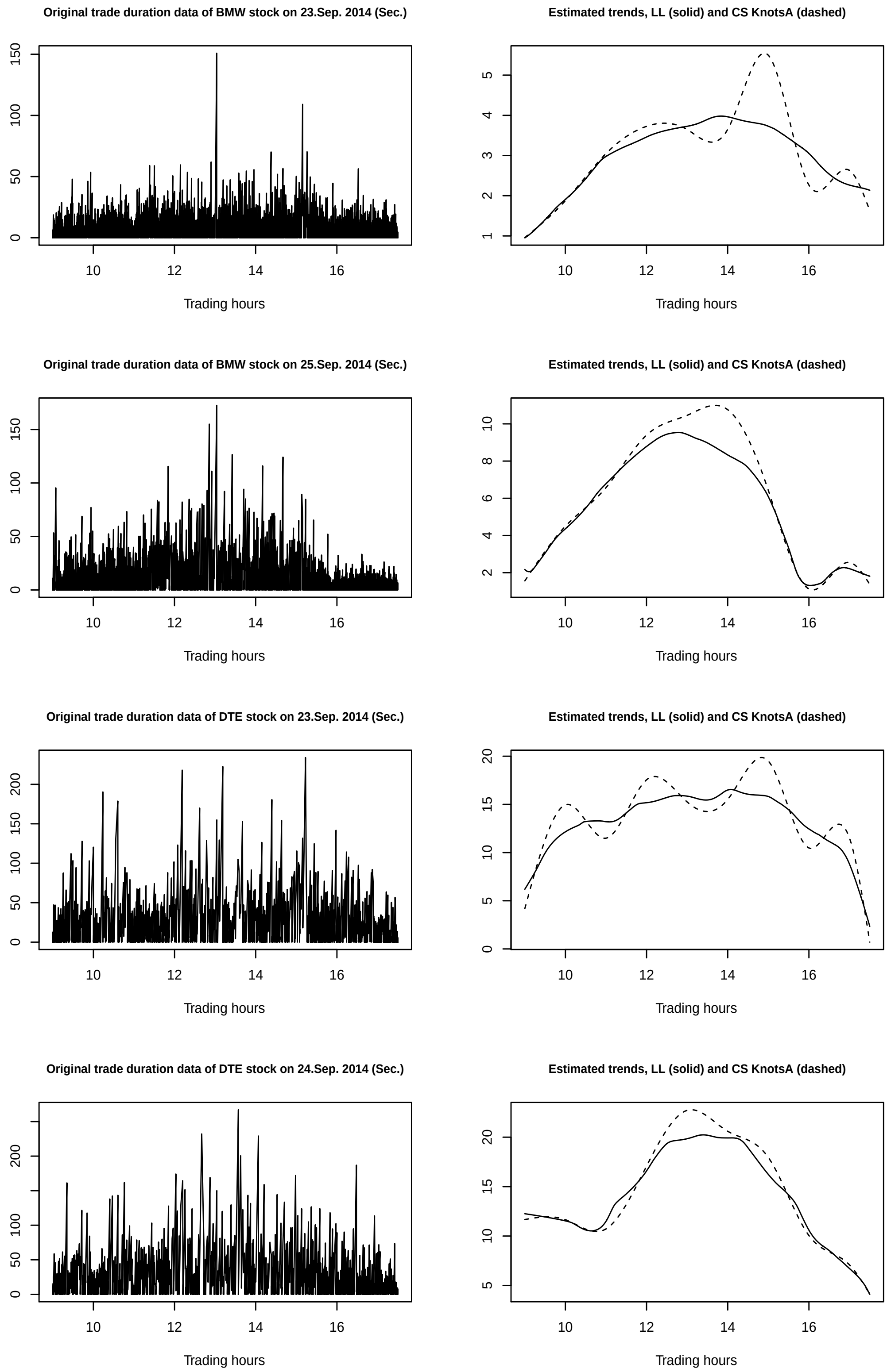

Figure 2.2 – Trends estimated by local linear and cubic spline approaches for four real data sets.

2.6 Conclusion

In this chapter a data-driven estimation of the diurnal pattern in a recently proposed Semi-ACD model is discussed. Detailed results on the bandwidth selection are obtained. A large simulation was carried out to discuss the practical performance of the proposed bandwidth selector in different cases. The results are then assessed in different ways. It is shown that the IPI bandwidth selector works well in general. One of the sub-methods using the EIM inflation form, an inflation factor $\lambda = 1/2$ and a coefficient $c_f = 6$ for calculating the lag-window estimator of the sum of all autocovariances seems to outperform the others in most of the cases, in particular if the performance is assessed using the goodness-of-fit of the estimated diurnal pattern. The results of the parameter estimation further showed that if a significant daily pattern is not removed from the data, the fitted ACD model is inconsistent. Hence, in practice the Semi-ACD model instead of the stationary parametric ACD model should be used. Furthermore, it is also shown that our proposal usually outperforms the commonly used cubic spline, because the latter is indeed still a parametric method and is not data-adaptive. Further experiments show that the smoothing spline, the nonparametric counterpart of the cubic spline, with a suitable smoothing parameter might be another attractive approach. Hence, it is worthy to develop a data-driven smoothing spline in the current context e.g. by adapting the proposal of Krivobokova and Kauermann (2007).

A.2 Appendix of Chapter 2

Proof of Lemma 2.1. Assume that the true scale functions and ACD model parameters ω, α and β are known. Let $\eta_i = y_i - \psi_i$ be a martingale difference sequence by construction. Following Engle and Russell (1998), the EACD(1, 1) model can be represented as an ARMA(1, 1) model: $y_i = \omega + (\alpha+\beta)y_{i-1} - \beta\eta_{i-1} + \eta_i$. Based on well known results on the sum of all autocovariances of an ARMA(1, 1) model we have

$$S = \left(\frac{1-\beta}{1-(\alpha+\beta)}\right)^2 \sigma^2_{\eta_i}. \tag{A2.1}$$

Straightforward calculation leads to

$$\sigma^2_{\eta_i} = \mathrm{var}\,(y_i) + \mathrm{var}\,(\psi_i) - 2\mathrm{cov}\,(y_i, \psi_i). \tag{A2.2}$$

Following Engle and Russell (1998) we have

$$\mathrm{var}\,(y_i) = \frac{1-\beta^2-2\alpha\beta}{1-\beta^2-2\alpha\beta-\alpha^2}.$$

Bauwens and Giot (2000) showed that $\mathrm{var}\,(\psi_i) = \frac{\alpha^2}{1-\beta^2-2\alpha\beta}$ and

$$\begin{aligned}\mathrm{cov}\,(y_i, \psi_i) &= E[(y_i - E(y_i))(\psi_i - E(\psi_i))] \\ &= E[y_i\psi_i] - E[y_i]E[\psi_i]. \end{aligned} \tag{A2.3}$$

Under the weakly stationarity assumption we have $E(y_i) = E(\psi_i)$. Furthermore, following Bauwens and Giot (2000), we have $E[y_i\psi_i] = E[\psi_i^2]$. This leads to $\mathrm{cov}\,(y_i, \psi_i) = E[\psi_i^2] - \mu^2 = \mathrm{var}\,(\psi_i)$ and

$$\sigma^2_{\eta_i} = \frac{1-(\alpha+\beta)^2}{1-(\alpha+\beta)^2-\alpha^2} \tag{A2.4}$$

Inserting (A2.4) into equation (A2.1) gives

$$S = \left(\frac{1-\beta}{1-(\alpha+\beta)}\right)^2 \frac{1-(\alpha+\beta)^2}{1-(\alpha+\beta)^2-\alpha^2}. \tag{A2.5}$$

Lemma 2.1 is proved. ◇

3 A semiparametric multiplicative error model for modelling long-term and conditional dynamics in non-negative financial data[7]

3.1 Introduction

The Autoregressive Conditional Duration (ACD) model of Engle and Russell (1998) for analysing irregularly spaced data underwent an extensive development since its initial proposal. Saart et al. (2015) refer to the different stages of development as generations. Extending the first generation baseline model of Engle and Russell (1998) parametrically constitutes the second generation and contains, for example, the Threshold ACD model (Zhang et al., 2001) or the augmented ACD model (Fernandes and Grammig, 2006) with the Box-Cox ACD model (Dufour and Engle, 2000a) and the Log-ACD model (Bauwens and Giot, 2000 and Bauwens et al., 2008) as special cases. The application of non- and semiparametric methods makes up the third generation. Wongsaart et al. (2010), for example, propose a semiparametric regression approach to nonlinear duration modelling. Brownlees and Gallo (2011) develop a shrinkage type estimator used for a semiparametric MEM (Engle, 2002) or Cosma and Galli (2006) introduce a nonparametric ACD model with an iterative algorithm proposed by Bühlmann and McNeil (1999) for a nonparametric GARCH model. Feng (2014) and Feng et al. (2016) propose a semiparametric ACD (Semi-ACD) model, based on data decomposition. Furthermore, alternatives to the baseline ACD model are proposed in the literature, such as the Birnbaum-Saunders ACD model, which uses the conditional median instead of the conditional mean to specify the time-varying model dynamics (Bhatti, 2010). Other extensions, such as the fractionally integrated ACD model of Jasiak (1999) or the long-memory stochastic duration model of Deo et al. (2010) are developed to take into account long-memory of the data.

[7]This chapter is available as a pre-print and was written in joint work with Professor Dr. Yuanhua Feng (Paderborn University). It is referred to as Forstinger and Feng (2018) throughout this work.

In this chapter we propose a semiparametric extension of the (first-type) Log-ACD model by Bauwens and Giot (2000) and Bauwens et al. (2008). Because the non-negativity constraints on the parameters can be dropped for the Log-ACD model, its semiparametric extension serves the aim of providing a more flexible alternative to the Semi-ACD model. Furthermore, the methods of the Semi-ACD model can be generalized to local polynomial regression for log-data and are not restricted to local linear regression anymore. The log-transformation also solves the problem of observations being close to or possibly touching the zero bound (see Engle, 2002) and it simplifies estimation, because the model becomes an additive nonparametric regression with stationary linear time series errors (see Shumway and Stoffer, 2011). For "the volatility counterpart of the Log-ACD model" (Allen et al., 2008, p.163), a semiparametric extension of the EGARCH model of Nelson (1991) is studied with a local polynomial link function estimation by Yang and Wu (2011). It is found to perform better than the EGARCH(1, 1) model, which supports our intention of proposing the Semi-Log-ACD model.
Properties of the proposal are discussed briefly. Firstly, conditions for the existence of a strictly and weakly stationary solution of the proposal as a general linear process are found. Some suitable examples fulfilling those conditions with possible heavy-tails at different levels are provided. Further detailed results are given for the special case, when the innovations are log-normally distributed. It is shown that now any power of the process is still log-normal, which has a similar dependence structure than that of the generating Gaussian process. Closed form formulas for their auto-correlations are obtained, which extend the results in Beran et al. (2015) and hold for long-memory processes. In particular, it is indicated that the autocorrelations of any power of such a process decay exponentially, if the log-data follows a normal stationary and invertible ARMA model.
An iterative plug-in (IPI) algorithm (Gasser et al., 1991) for selecting the bandwidth in a local polynomial regression is further proposed here. The aim is to develop a fully nonparametric and automatic IPI algorithm independent of any parametric assumption. Generalizing the IPI to local polynomial regression increases the range of applicability of the IPI, as well as the semiparametric model fitting to log-data, in general. The estimation of the variance factor in the asymptotically optimal bandwidth is the key point of the IPI. Here, we propose to use a lag-window estimator for the variance factor and adapt another iterative plug-in procedure for choosing the window-width in this context (see Bühlmann, 1996). It is evaluated through application to different data sets. Compared to the estimation of the variance factor by means of an ARMA model it is shown that the current proposal is more stable. It can also be extended to the case with long-memory, but this is not straightfor-

ward and beyond the aim of this chapter. The effect of estimation errors in the nonparametric part on the further parametric estimation is investigated. Conditions are found under which the parametric estimation is still $\sqrt{T}$-consistent. They are fulfilled by an automatically selected bandwidth, as proposed here.
The remaining part of this chapter is organized as follows: Section 3.2 defines the model and discusses the existence conditions. Further properties of a general linear process and the specified Semi-Log-ACD model are provided in section 3.3. The estimation procedure is described in section 3.4. Section 3.5 discusses the practical implementation of the algorithm including the methods for estimating the variance factor. The proposals are applied to different real financial data examples and evaluated in section 3.6 and section 3.7 concludes.

3.2 The proposed models

The models are described for non-negative daily average financial data to simultaneously model its long-term and conditional dynamics. Extending the proposals and results to irregularly spaced data is straightforward, though. Compared to the ACD model, the Semi-ACD model by Feng (2014) and Feng et al. (2016) and the Log-ACD model of Bauwens and Giot (2000) and Bauwens et al. (2008) are each more flexible. The Semi-ACD model estimates the deterministic and the stochastic data parts separately and can thereby account for long-term dynamics or daily patterns in the data. The Log-ACD model does not require the non-negativity constraints on the model parameters. Combining these models to the Semi-Log-ACD model aims to further increase the flexibility. The scale function estimation IPI developed for the Semi-ACD model is general and can be adapted to the Log-ACD model. It is not limited to local linear regression, though, so the IPI is described for local polynomial regression later. The additive structure of the log-data in this model, in fact, even simplifies the scale function estimation. In the following, the general framework is set for the ACD model and its semiparametric extension. Their log-transformation gives the Log-ACD and Semi-Log-ACD models, which can be defined as linear processes. Latter enable a general discussion on the asymptotic properties for processes which allow long-memory, short-memory or antipersistence.

3.2.1 The Semi-ACD model

Let $X_t \geq 0$ denote the observations for $t = 1, ..., T$. The baseline ACD model is extended semiparametrically by including a time-varying smooth scale function,

$\nu(\tau_t) > 0$, to model the local, deterministic changes of the mean level of X_t:

$$X_t = \nu(\tau_t)\psi_t\varepsilon_t. \tag{3.1}$$

The conditional mean, $\psi_t = \omega + \sum_{j=1}^{p} \alpha_j X^*_{t-j} + \sum_{r=1}^{q} \beta_r \psi_{t-r}$, follows the linear parametrisation as suggested by Engle and Russell (1998) with $X^*_t = \frac{X_t}{\nu(\tau_t)} = \psi_t\varepsilon_t$. The innovations, ε_t, are i.i.d. non-negative random variables and $\tau_t = t/T$ denotes the rescaled time. For a general sequence of regularly spaced non-negative random variables (3.1) is a semiparametric extension of the MEM (Engle, 2002), so X_t is not limited to being transaction durations. It will still be denoted by the ACD model terminology here. The non-negativity constraints to ensure that $\psi_t > 0$ are $\omega > 0$, $\alpha_1, ..., \alpha_p \geq 0$ for $j = 1, ..., p$ and $\beta_1, ..., \beta_q \geq 0$ for $r = 1, ..., q$. The Semi-ACD model as described by (3.1) and ψ_t is non-stationary, but is assumed to be a locally stationary process following Dahlhaus (1997). This allows to model the long-term dynamics of the data together with its conditional dynamics, expressed by the scale function and the model parameters, respectively.

3.2.2 Linear processes and the Semi-Log-ACD model

Within the framework of (3.1), define $Y_t = \log(X_t)$, $\mu(\tau_t) = \log(\nu(\tau_t))$, $\zeta_t = \log(\psi_t)$, $Y^*_t = \log(X^*_t) = \zeta_t + \eta_t = Y_t - \mu(\tau_t)$ and $\eta_t = \log(\varepsilon_t)$ with $E(\eta_t) = 0$ and $var(\eta_t) = \sigma^2_\eta$. The multiplicative model in (3.1) can be written as an additive nonparametric regression with linear stationary errors of the following form:

$$Y_t = \mu(\tau_t) + Y^*_t. \tag{3.2}$$

The description of Y_t given in (3.2) is the first part of the Semi-Log-ACD model. In addition, the log-conditional mean is linearly parametrised as $\zeta_t = \omega + \sum_{j=1}^{p} \alpha_j Y^*_{t-j} + \sum_{r=1}^{q} \beta_r \zeta_{t-r}$ (first-type Log-ACD model; Bauwens and Giot, 2000 and Bauwens et al., 2008). No non-negativity constraints on the parameters are required. In the following, the asymptotic properties are discussed generally. For this purpose let Y^*_t be a linear process and X^*_t a log-linear process accordingly. Denote the MA(∞) representation of Y^*_t by:

$$Y^*_t = \alpha(\mathrm{B})\eta_t = \sum_{i=0}^{\infty} \alpha_i \eta_{t-i}, \tag{3.3}$$

where $\alpha(\mathrm{B}) = \sum_{i=0}^{\infty} \alpha_i \mathrm{B}^j$, $\alpha_0 = 1$ and $\sum_{i=0}^{\infty} |\alpha_i| < \infty$. Following e.g. Prado and West (2010), (3.3) is the stationary solution of Y^*_t. Assuming invertibility of Y^*_t, it can also be represented as an AR(∞) model with $\sum_{j=0}^{\infty} |b_j| < \infty$, implying that

$\sum_{i=0}^{\infty} \alpha_i \neq 0$:

$$Y_t^* = \sum_{j=1}^{\infty} b_j Y_{t-j}^* + \eta_t. \tag{3.4}$$

Here, α_i and β_j can be obtained by matching the powers of B in $\phi(\mathrm{B})\psi(\mathrm{B})^{-1}$ and $\psi(\mathrm{B})\phi(\mathrm{B})^{-1}$, respectively (see e.g. Property 3.1 and 3.2 of Shumway and Stoffer, 2011). From (3.3) a closed form expression of the residual autocovariances of Y_t^* and their sum can be derived, i.e. $\gamma(k) = \sigma^2 \sum_{i=0}^{\infty} \alpha_i \alpha_{i+|k|}$ and $c_f = \frac{1}{2\pi} \sum_{-\infty}^{\infty} \gamma(k)$. The above assumptions ensure that $0 < \sum_{-\infty}^{\infty} \gamma(k) < \infty$. The autocorrelations of Y_t^* are independent of the conditional distribution of X_t^*, which are given by $\rho(k) = \sum_{i=0}^{\infty} \alpha_i \alpha_{i+k} / \sum_{i=0}^{\infty} \alpha_i^2$ for $k = 0, 1, \ldots$. The existence of moments and results on the autocorrelations for X_t^* are studied for Log-ACD models by Bauwens et al. (2008) and Karanasos (2008). By comparing $\rho(k)$ with corresponding results in aforementioned works the tail behaviour of X_t^* and Y_t^* can be discussed. A thorough discussion on this topic is beyond the aim of this chapter, though, but the results on the existence of the moments will be extended to the here discussed generalized linear process definition. For this purpose, let $\alpha_{\sup} = \lim_{k\to\infty} \max(\alpha_0, \alpha_1, \ldots, \alpha_k) \geq \alpha_0 = 1$ and $\alpha_{\inf} = \lim_{k\to\infty} \min(\alpha_0, \alpha_1, \ldots, \alpha_k)$. We introduce the following conditions:

A1. Y_t^* in (3.3) is stationary and invertible with $\alpha_0 = 1$ and $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$.

A2. Both, $E(\varepsilon_t^{m\alpha_{\sup}})$ and $E(\varepsilon_t^{m\alpha_{\inf}})$ exist for some integer $m \geq 2$.

Taking the exponential transformation of (3.3), we obtain

$$X_t^* = \prod_{i=0}^{\infty} \varepsilon_{t-i}^{\alpha_i}, \tag{3.5}$$

which is the strictly and weakly stationary solution of X_t^*, if A1 and A2 hold. Under the same conditions, X_t^* has finite moments of any order as given in equation (9) of Karanasos (2008). By expanding (3.5) the locally stationary solution of X_t is given by:

$$X_t = \nu(\tau_t) \prod_{i=0}^{\infty} \varepsilon_{t-i}^{\alpha_i}. \tag{3.6}$$

Furthermore, if A1 and A2 hold, the corresponding moments of the linear process exist, because the scale function is bounded. If A1 holds, A2 is a necessary and sufficient condition for the existence of a stationary solution of Y_t^* with finite m-th moment, which is jointly determined by the properties of the conditional distribution as well as $\alpha_{\inf}$ and $\alpha_{\sup}$. Depending on the distribution choice, requirements on the fulfilment of A2 might lead to stronger restrictions on α_i than the one given in A1. Note that $\alpha_{\inf}$ can be negative, so the existence of the m-th moment of X_t^* might require the existence of negative moments for ε_t up to the order $m\alpha_{\inf}$.[8] For instance, $E(\varepsilon_t^\delta)$ only exist, for

- $\varepsilon_t \sim \exp(\lambda)$, when $\delta \geq -1$, so X_t^* is only weakly stationary, if $\alpha_{\inf} > -1/m$ and for
- $\varepsilon_t \sim$Rayleigh(σ), when $\delta \geq -2$, so X_t^* is only weakly stationary, if $\alpha_{\inf} > -2/m$.

The exponential distribution is a common choice for ACD models, as it was originally proposed by Engle and Russell (1998). The Rayleigh distribution is used, for example, by Pathmanathan et al. (2010) to compare different estimation methods for ACD models. For some other distributions, the moments of positive order only exist up to a certain order, which might require additional restrictions on $\alpha_{\sup}$. For instance,

- $\varepsilon_t \sim$ Fréchet (a, s, m), for $\delta < a$, so X_t^* is only weakly stationary, if $\alpha_{\sup} < a/m$;
- $\varepsilon_t \sim$ Gumbel (μ, β), for $\delta < 1/\beta$, so X_t^* is only weakly stationary, if $\alpha_{\sup} < 1/m\beta$ and
- $\varepsilon_t \sim$ Log-Logistic (α, β), for $\delta < \beta$, so X_t^* is only weakly stationary, if $\alpha_{\sup} < \beta/m$.

Zheng et al. (2016) propose the use of the Fréchet conditional distribution for capturing characteristics of block trades, such as heavy tails and extreme values in financial durations. Lindner and Meyer (2003) discuss the Gumbel distribution for the EGARCH and Log-ACD model and the log-logistic distribution is included here as a special case of the Generalized F distribution, discussed by Karanasos (2008). Further restrictions on both $\alpha_{\sup}$ and $\alpha_{\inf}$ are required for distributions with order limitations on both hand sides. For example,

- $\varepsilon_t \sim$ Lomax (λ, a), for $-1 < \delta < a$, so X_t^* is only weakly stationary if, $\alpha_{\inf} > -1/m$ and $\alpha_{\sup} < a/m$;
- $\varepsilon_t \sim$ Burr (λ, a, η), for $-a < \delta < a\eta$, so X_t^* is only weakly stationary, if $\alpha_{\inf} > -a/m$ and $\alpha_{\sup} < ma/\eta$ and
- $\varepsilon_t \sim$ F (d_1, d_2), for $-d_1/2 < \delta < d_2/2$, so X_t^* is only weakly stationary, if $\alpha_{\inf} > -d_1 m$ and $\alpha_{\sup} < d_2 m$.

[8] The following examples are on common distribution choices for short-memory ACD models, however, the discussion is general and not limited to a short-memory framework.

The Lomax distribution is included in the Generalized F distribution and a Burr-ACD model is introduced by Grammig and Maurer (2000). If the conditional distribution has finite moments of any order δ for $\delta \in (-\infty, \infty)$, the linear and log-linear process have finite moments of any order m for $m \in (0, \infty)$ and now A2 is automatically fulfilled. The use of different conditional distributions allows to model heavy tails at different levels in the data. Assuming that $E(X_t^{*2}) < \infty$ but $E(X_t^{*4}) = \infty$ we will say that X_t^* has clear heavy-tails. If $E(X_t^{*4}) < \infty$ but $E(X_t^{*8}) = \infty$, X_t^* is said to have light heavy-tails. A process with $E(X_t^{*8}) < \infty$ will be called a non-heavy tailed one in this chapter.

3.3 Correlation structure under log-normal assumption

If the conditional distribution is log-normal, moments of any order of X_t^* exist and the dependence structure of X_t^* is completely known. Detailed properties on the correlation structure under this strong assumption are first stated for the general linear process. Specified properties under a Gaussian Semi-Log-ACD model will be described briefly.

3.3.1 Results for the general linear process

Beran et al. (2015) propose to model the stochastic component of the log-data by a Gaussian FARIMA with long memory. In this chapter we will extend their results to a common power $(X_t^*)^m$, $m > 0$. The following lemma is a straightforward extension of Lemma 1 in Beran et al. (2015), which gives the closed form formula of the stationary solution of any m-th power of X_t^*.

Lemma 3.1 *Assume that ε_t is log-normally distributed, i.e. $\varepsilon_t \sim LN(0, \sigma_\varepsilon^2)$. Then for any $m > 0$, $(X_t^*)^m = \prod_{i=0}^{\infty} \eta_{t-i}^{ma_i}$ is a weakly and strictly stationary process with a $LN(0, m^2\sigma^2)$ marginal distribution, where $\sigma^2 = \sigma_\epsilon^2 \sum_{i=0}^{\infty} a_i^2$.*

Lemma 1 in Beran et al. (2015) corresponds to Lemma 3.1 with $m = 1$ here. The proof of Lemma 3.1 is therefore omitted. Since any power of a log-normal distribution is still log-normal, a straightforward extension of the dependence structure of x_t^* to $(X_t^*)^m$ for any $m > 0$ is possible. For a Gaussian log-linear process, the dependence structure of $(X_t^*)^m$, $m > 0$, is given by the following theorem.

Theorem 3.1 *Under the assumptions of Lemma 3.1 we have*

i)
$$E[(X_t^*)^m] = e^{m^2\sigma^2/2} \quad \text{and} \quad var[(X_t^*)^m] = e^{m^2\sigma^2}\left(e^{m^2\sigma^2} - 1\right),$$
where σ^2 *is as defined in Lemma 3.1.*

ii)
$$\gamma_m(k) = e^{m^2\sigma^2}\left(e^{m^2\sigma_\epsilon^2 \sum_{i=0}^{\infty} a_i a_{i+k}} - 1\right).$$

iii)
$$\rho_m(k) = \left(e^{m^2\sigma_\epsilon^2 \sum_{i=0}^{\infty} a_i a_{i+k}} - 1\right)\left(e^{m^2\sigma^2} - 1\right)^{-1}$$
$$= \left(e^{m^2\sigma^2\rho_{Y^*}(k)} - 1\right)\left(e^{m^2\sigma^2} - 1\right)^{-1}. \tag{3.7}$$

iv) The relationship between $\rho_m(k)$ *and* $\rho_{Y^*}(k)$ *for large* k *is given by*
$$\rho_m(k) \sim c_\rho(m)\rho_{Y^*}(k), \tag{3.8}$$
where $\sim$ *means that the ratio of both sides tends to 1, as* $k \to \infty$ *and* $0 < c_\rho(m) < 1$ *is a non-negative monotonically decreasing function of* m *as defined in the appendix with* $\lim_{m\to 0} c_\rho(m) = 1$ *and* $\lim_{m\to\infty} c_\rho(m) = 0$.

The proof of Theorem 3.1 is given in the appendix. Items *i)* to *iii)* provide simple closed form formulas for the dependence structure of $(X_t^*)^m$, in addition to the result in Lemma 3.1. This is in contrast to most volatility or duration models, where the correlation structure, conditions for the existence of high order moments and the marginal distribution are usually very complex or even unknown. Moreover, a particularly interesting finding is the result in (3.8), which provides an asymptotic relationship between $\rho_m(k)$ and $\rho_{Y^*}(k)$. The rate of decay of the autocorrelations of any power of X_t^* is the same as that for the underlying Gaussian process Y_t^*. The constant in the asymptotic formula, which is between 0 and 1, depends on m, though. The larger m, the smaller the constant. Note in particular that the fact $\lim_{m\to 0} c_\rho(m) = 1$ must hold, as $\lim_{m\to 0}[(X_t^*)^m - 1]/m \to \ln(X_t^*) = Y_t^*$.

3.3.2 Further properties in the short-memory case

The algorithm for estimating the scale function is based on an algorithm by Bühlmann (1996) which is developed for short-memory data. Thus, the following contents are described for short-memory data only, but the results on linear processes of the previous sections still hold. Also an extension of the following methods to long-memory data is possible but not subject of this chapter. A special case of the Log-ACD model that is discussed here and applied in section 3.6 is a linear ARMA model. Let Y_t^* follow a stationary and invertible ARMA(p, q) model (Allen et al., 2008):

$$\phi(\mathrm{B})Y_t^* = \psi(\mathrm{B})\eta_t, \tag{3.9}$$

where $\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$ and $\psi(z) = 1 + \psi_1 z + ... + \psi_q z^q$ are the characteristic polynomials of the AR and MA parts, respectively. It is assumed that they have no common factors and all of the roots of $\phi(z) = 0$ and $\psi(z) = 0$ lie outside the unit circle. The log-linear process, X_t^* and (3.9) define a Log-ACD model in a strict sense, whereas X_t in (3.1) and (3.9) define a semiparametric generalization of the Log-ACD model. The short-memory adapted Gaussian ARMA follows a MEM with log-normally distributed innovations, for which Allen et al. (2008) show that its log-transformation follows a linear ARMA. Consequently, the exponential transformation of this linear ARMA model can be written as a Log-ACD model. Based on the findings of Theorem 3.1 iv), the following corollary can be derived for the short-memory Gaussian Log-ACD model:

Corollary 3.1 *Under the assumptions of Theorem 3.1 and the further assumption that Y_t^* follows a stationary ARMA process, $\rho_m(k)$ for any $m > 0$ decay exponentially.*

The proof of Corollary 3.1 is straightforward and therefore omitted. This result indicates that now for any $m > 0$, the acf of $(X_t^*)^m$ are not only absolutely summable but also decay very quickly.

3.4 The two-stage estimation procedure

For estimating the scale function in (3.1) directly, the existence of the fourth order moment of X_t^* is usually required for developing a data-driven bandwidth selector. Furthermore, it involves a nonparametric regression with time-varying variance $\mathrm{var}\,(X_t) = \nu^2(\tau_t)\mathrm{var}\,(X_t^*)$. As the required assumption on the existence of the fourth order moment of Y_t is less of a problem for financial processes, we propose to estimate $\nu(\tau)$ equivalently via (3.2) and not directly via (3.1). The existence of $E[(X_t^*)^m]$ for all $m \in (-m_0, m_0)$ for $m_0 > 0$ implies the existence of the moment generating function of Y_t^* in the neighbourhood $(-m_0, m_0)$. Under the additional regularity assumption A1 on $\phi(\mathrm{B})$ and $\psi(\mathrm{B})$, Y_t^* has a strict stationary solution with finite moments of any order. Here, the methods are described for $m = 1$. The estimation of the scale function does not depend on the power of X_t^*, because after the log-transformation the power becomes a constant. In a first step $\mu(\tau)$ is estimated and the residuals $\hat{Y}_t^* = Y_t - \hat{\mu}(\tau_t)$ are obtained. In a second step, the parameters of the model can be estimated from the residuals. The following descriptions are for the very general case of estimating the v-th derivative of $\mu(\tau)$ to show that it also applies to other data than the one used here. The formulas can be extended to long-memory easily, however, for above mentioned reasons we limit this section to the short-memory case.

3.4.1 Local polynomial estimation of the trend

As opposed to the Semi-ACD and related models, we propose a local polynomial estimator for $\hat{\mu}$. Since the Log-ACD does not require parameter constraints to ensure non-negativity, the nonparametric trend estimation methods are not restricted to local linear regression, as in the case of the Semi-ACD model. Local polynomial is asymptotically a kernel estimator with automatic boundary correction (e.g. Fan and Gijbels, 1996 and Cheng et al., 1997) and eases the estimation of higher order kernel functions, as well as derivatives. Feng et al. (2018) apply local polynomial regression to estimate the trend and its derivatives for economic time series. Here we discuss this method for the log-trends of non-negative financial time series and extend the results of Feng et al. (2018) theoretically. We impose the following regularity conditions:

B1. The scale function $\mu(\tau)$ is strictly positive, bounded and $k = p + 1$ times continuously differentiable on [0,1].

B2. The kernel $K(u)$ is a symmetric density with compact support on [-1,1].

B3. The bandwidth b satisfies $b \to 0$ and $b\mathrm{T} \to \infty$ as $\mathrm{T} \to \infty$.

These conditions are necessary for the derivation of asymptotic results and ensure that the model can be estimated by the proposed Semi-Log-ACD model algorithm. Condition B1 is on the smoothness of the scale function. B1 and B3 are required to derive the order of magnitude of the bias of the estimated scale function. The assumption on the kernel function made in B2 is imposed for simplicity. The local polynomial estimator of the v-th derivative of μ is obtained by minimizing the weighted least squares equation:

$$Q = \sum_{t=1}^{T} \{Y_t - \sum_{j=0}^{p} \beta_j (\tau_t - \tau)^j\}^2 K\left(\frac{\tau_t - \tau}{b}\right), \tag{3.10}$$

where b is the bandwidth and K a second order kernel function on $[-1, 1]$. The estimate is obtained by $\hat{\mu}(\tau)^{(v)} = v!\hat{\alpha}_v$. The asymptotically equivalent kernel of $\hat{\mu}(\tau)$ is the kernel $K(u)$ itself at an interior point and a corresponding boundary kernel at a boundary point. For $p - v = odd$ the estimator has automatic boundary correction and the order of the bias is uniform, so that the local polynomial regression achieves the global optimal rates of convergence (Feng and Beran, 2013). Asymptotic properties of $\hat{\mu}(\tau)$ can be obtained by adapting known results in the literature on nonparametric regression with short-range dependent errors (e.g. see Hart, 1991 or Opsomer et al., 2001). For the kernel function define $R(K) = \int K^2(u)du$, $I(K) = \int u^k K(u)du$ and $I(\mu^k) = \int[\mu^k(\tau)]^2 d\tau$ and let c_f denote the value of the spectral density of Y_t^* at the origin.
The bias of $\hat{\mu}(\tau)$ is not affected by correlation in the errors, so it is the same as in nonparametric regression with i.i.d. errors (see e.g. Hart, 1991). The variance of $\hat{\mu}(\tau)$ is affected by the error correlation, so presuming error independence might lead to a too small (large) bandwidth that undersmoothes (oversmoothes) the regression function estimate (e.g. De Brabanter et al., 2011; Hermann et al., 1992; Hart, 1999). Under A1, A2 and B1 to B3 the bias is

$$B(\hat{\mu}(\tau)) = b^{(k-v)} \frac{\mu^{(k)}(\tau) I(K))}{k!}[1 + o(1)] \tag{3.11}$$

and the variance is given by

$$\text{var}\,(\hat{\mu}(\tau)) = \frac{2\pi c_f R(K)}{Tbb^{2v}}[1 + o(1)]. \tag{3.12}$$

The global optimal bandwidth that is to enter the weighted sum of least squares equation is the one that minimizes the dominant part of the MISE of $\hat{\mu}(\tau)$. It can be approximated for practical convenience by the asymptotic MISE:

$$\text{AMISE}(\hat{\mu}(\tau)) = b^{2(k-v)} \frac{I(\mu^k) I(K)^2}{[k!]^2} + \frac{2\pi c_f R(K)}{Tbb^{2v}} + o\left(\max(b^{2(k-v)}), o(\frac{1}{Tbb^{2v}})\right). \tag{3.13}$$

The AMISE can be calculated on the whole support [0,1], because the dependence of both the kernel and the bandwidth on τ in the bias and variance of $\hat{\mu}$ does not affect it. The contribution of the estimated values in the boundary area is asymptotically negligible and the asymptotically global optimal bandwidth is given by

$$b_A = \left(\frac{2v+1}{2(k-v)} \frac{2\pi c_f [k!]^2 R(K)}{I(K)^2 (K) I(\mu^k)} \right)^{1/(2k+1)} T^{-1/(2k+1)}. \tag{3.14}$$

If $2\pi c_f > \gamma(0)$ ($2\pi c_f < \gamma(0)$) the bandwidth becomes too small (too large), when the errors are correlated but assumed to be uncorrelated, whereas if they are correlated but in fact they are not, $2\pi c_f$ reduces to the residual variance (Hermann et al., 1992).

3.4.2 Parameter estimation from residuals

For estimating the unknown parameters, let $\theta = (\psi_1, ..., \psi_p; \phi_1, ..., \phi_q)'$ denote the unknown parameter vector and assume that $\hat{\mu}(\tau)$ is a consistent estimator of $\mu(\tau)$. Using the corresponding maximum likelihood estimators built in R or S-Plus, θ can be estimated from $\hat{Y}_t^*$ easily and will be denoted by $\hat{\theta}$. Under the conditional log-normal assumption, the estimation is carried out via approximate MLE, whereas if the conditional distribution is not log-normal, an approximate QMLE is used. If the conditional distribution is far from log-normal, θ can e.g. be estimated by the conditional least squares approach. As Y_t^* are not directly observable, the discussion of whether errors in the trend estimation affect the parametric estimation is of great importance (see Feng, 2004a). Letting $\tilde{\theta}$ denote the standard $\sqrt{T}$-consistent estimator of θ, obtained using Y_t^*, it can be shown that $\sqrt{T}(\hat{\theta} - \tilde{\theta}) = o_p(1)$, if $o(T^{-1/2}) < b < o(T^{-1/(4k)})$. The additional bias in $\hat{\theta}$ due to the bias of $\hat{\mu}$ is negligible, if $b < o(T^{-1/(4k)})$. That due to the variance of $\hat{\mu}$ is negligible, if $o(T^{-1/2}) < b$. The additional variance in $\hat{\theta}$ can always be considered negligible, if $\hat{\mu}$ is consistent (Beran and Feng, 2002b; Feng et al., 2018). As $b_A = O(T^{-1/(2k+1)})$, the asymptotically optimal bandwidth satisfies the condition for the $\sqrt{T}$-consistency and $\hat{\theta}$ is asymptotically not affected by the bias and variance of $\hat{\mu}(\tau)$.

3.5 Practical Implementation

We propose a data-driven IPI algorithm for estimating the scale function with automatic global bandwidth selection via a nonparametric regression approach with time series errors (e.g. see Beran and Feng, 2002a or Gasser et al., 1991). This method was also successfully extended to allow for local adaptivity (Herrmann et al., 1992; Brockmann et al., 1993 and Herrmann, 1997). The two unknowns in the formula of b_A in (3.14) are c_f and $\hat{I}(\mu)^k$. Of these two, only the variance term might be affected strongly by correlated errors (see Opsomer et al., 2001), whereas $I(\mu)^k$ can be estimated by established techniques for models with i.i.d. errors.

3.5.1 Variance factor estimation

Besides the limitation to local linear regression, the IPI developed for the Semi-ACD model also differs in the estimation of c_f. It is based on the residual autocovariances via a nonparametric estimator: $\tilde{c}_f = \frac{1}{2\pi} \sum_{|k|<K} \omega_k \hat{\gamma}(k)$, where $\omega_k = 1 - k/(K+1)$

are Bartlett-window weights (see e.g. Priestley, 1981) with $K = C_M T^{1/3}$. The window-width K is fixed manually by choosing a value of C_M. To increase the degree of automation of the IPI, we propose three methods in total to estimate c_f data-driven instead of fixing it beforehand by choosing a value for the maximal lag of the sum of residual autocovariances. Two of these methods are parametric by assuming AR(MA) models fitted to the obtained residuals, as described in model (3.9):

$$\hat{c}_{f,\text{ARMA}} = \left(\frac{1 + \hat{\psi}_1 + ... + \hat{\psi}_q}{1 - \hat{\phi}_1 - ... - \hat{\phi}_p} \right)^2 \sigma_\eta^2.$$

For an AR model the numerator reduces to 1. For consistency reasons the orders of the models are selected by BIC. To maintain the flexibility of the Semi-Log-ACD algorithm, we also propose a nonparametric estimation procedure for the variance factor, which does not require any parametric assumption on the residuals:

$$\hat{c}_{f,\text{NP}} = 1/2\pi \sum_{c=-C_K}^{C_K} \omega_c \hat{\gamma}(c).$$

We take on the proposal of Bühlmann (1996) to use an IPI procedure for automatically selecting the optimal window-width. It minimizes the asymptotic MSE at a fixed frequency of the lag-window estimator. In a first step, a global estimate $C_{K,\text{opt}}$ is obtained which is used in a second step to obtain a local estimate at the frequency $\xi = 0$. Let $\hat{f}(\xi)$ denote the lag-window estimator for the spectral density of Y_t^* at frequency ξ as given in (2) of Bühlmann (1996). The algorithm follows Bühlmann (1996) in most parts, except for the inflation factor being chosen smaller and, thus, more than a maximal number of four iterations for the global step (2.) is used:

1. Choose a starting value of the lag-bandwidth manually. Here $C_{K,0} = [T/2]$, where $[\bullet]$ denotes the integer part.

2. In the j-th iteration put $C'_{K,j} = [C_{K,j-1}/T^{2/21}]$. Using (2) of Bühlmann (1996) estimate $\int (f(\xi)^2) d\xi$ and $\int f^{(1)}(\xi) d\xi$, where $f^{(1)}(\xi)$ denotes the first order generalized derivative of $f(\xi)$. Insert the estimates into (5) of Bühlmann (1996) for the optimal window-width for a Bartlett window.

3. Increase j by one and repeatedly carry out Step 2 until convergence or until a maximal number of iterations is reached (here: 20). Set $\hat{C}_{K,G} = \hat{C}_{K,j}$.

4. Use $\hat{C}'_K = [C_{K,G}/T^{2/21}]$ to calculate $\int f^{(1)}(\xi) d\xi$. Insert it into (5) of Bühlmann (1996). Obtain $\hat{C}_K$ for estimating $f(0)$.

Feng et al. (2018) propose the use of a bandwidth correction factor following table 1 of Feng and Heiler (2009). This is to improve the quality of the estimation of c_f, as it ensures that the asymptotically optimal bandwidth for estimating the residual-based variance is larger than the one for estimating the scale function. The number of observations for HF financial data is usually much larger than in the economic time series setting of Feng et al. (2018), so we do not consider this adjustment necessary here.

3.5.2 The bandwidth selection algorithm

The main bandwidth selection IPI is adapted to (3.14) to estimate the scale function of equidistant log-data. The algorithm is as follows:

1. Choose a starting bandwidth, b_0, and put $j = 1$.

2. Refine b_0 via the IPI idea, ignoring correlation and scale change for J_1 iterations.

3. In the j-th iteration for $j > J_1$:

    a) Obtain $\hat{\mu}(\tau)$ via b_{j-1}. Calculate $\hat{Y}_t^* = Y_t - \hat{\mu}(\tau)$ and $\hat{\gamma}(k)$, estimate c_f.

    b) Obtain $\hat{I}(\mu^k)$ via an inflation method and factor of choice. Here we use the EIM of Beran and Feng (2002a) $b_{2j} = b_{j-1}^{\upsilon}$.

    c) Increase j by one and repeatedly carry out steps 3.a) to c) until convergence or a fixed maximal number of J iterations is reached. Set $\hat{b}_A = b_j$ or $\hat{b}_A = b_J$.

Convergence in this case means that a convergent output or a fixed point is achieved. The autocovariances $\hat{\gamma}(k)$ are calculated from the residuals, $\hat{Y}_t^*$, in the $(j-1)^{th}$ iteration. Gasser et al. (1991) propose a multiplicative inflation method, however the rate of convergence of the estimated bandwidth is much slower than the one of the bandwidth selected with an IPI using the EIM. This inflation method is chosen with an inflation factor of $\upsilon = 5/7$, because then the selected optimal bandwidth achieves the highest rate of convergence (see Beran and Feng, 2002a), which is also shown as one of the results of Theorem 3.2.
The main IPI algorithm is a fix-point search procedure. If the variance factor is chosen manually beforehand, the choice of the starting bandwidth does not affect the finally selected bandwidth, if chosen from a suitable range (Herrmann, 1997). If c_f is estimated, the choice of the starting bandwidth does affect the finally selected bandwidth in some cases, which was found and discussed in the practical evaluation of the adapted algorithm in section 3.6. Regardless of this practical issue, the quality of the finally selected bandwidth $\hat{b}_{\mathrm{A}}$ can be quantified by the following theorem:

Theorem 3.2 *Under assumptions A1, B1 through B3 and the additional assumption that $E(Y_t^{*4}) < \infty$, the following holds: For $\upsilon = 1/2$,*

$$\hat{b}_A = b_A[1 + O(T^{-1/(2k+1)}) + O_p(T^{-1/2}) + O_p(T^{-1/3})]. \tag{3.15}$$

For $\upsilon = (2k+1)/(2(k+2)+1)$,

$$\hat{b}_A = b_A[1 + O(T^{-2/(2(k+2)+1)}) + O_p(T^{-2k/(2(k+2)+1)}) + O_p(T^{-1/3})]. \tag{3.16}$$

For $\upsilon = (2k+1)/(2k+3)$,

$$\hat{b}_A = b_A[1 + O(T^{-2/(2k+3)}) + O_p(T^{-2/(2k+3)}) + O_p(T^{-1/3})]. \tag{3.17}$$

The proof of Theorem 3.2 is given in the appendix. For $k = 2$, the (relative) rates of convergence of the selected bandwidth are of the orders $O(T^{-1/5})$, $O_p(n^{-2/9})$ and $O_p(n^{-2/7})$. These rates of convergence are not very high but much higher than the $O_p(n^{-1/10})$ convergence rate for a cross-validation bandwidth selector. Compared to these terms, the $O_p(T^{-1/3})$ term, which is caused by the error in $\hat{c}_f$, is asymptotically negligible. The same holds for $k = 4$ and the (relative) rates of convergence of the selected bandwidth of the orders $O(T^{-1/9})$, $O_p(n^{-2/13})$ and $O_p(n^{-2/11})$. In summary, the error in the lag-window estimator does not affect the rate of convergence of the finally selected bandwidth. For $\upsilon = (2k+1)/(2k+3)$ the (relative) rate of convergence of the selected bandwidth is optimal. The variance and the bias converge at the same rate and the MSE of $\hat{I}[\mu^{(k)}]$ is minimized. It is known that this rate is the optimal rate of convergence for a local polynomial approach with independent or short-memory errors, when b_A is selected by a plug-in rule. If $\upsilon = 1/2$ is used, the variance of $\hat{I}[\mu^{(k)}]$ is minimized and does not depend on k, but its bias is clearly increased. Feng et al. (2016) show, though, that local linear regression with this inflation factor yields the best results for UHF data with large sample sizes. The other two inflation factors perform clearly worse in this setting. If the observation number is not as high as it is for UHF data this inflation factor is likely to be too strong and should not be used. Feng et al. (2018), for example, use $\upsilon = (2k+1)/(2(k+2)+1)$ and $\upsilon = (2k+1)/(2k+3)$ in an economic time series context, i.e. with usually much less observations than for UHF financial data. All of the inflation factors given in Theorem 3.2 are therefore of practical relevance and the choice of factor depends on the data at hand. In either case the estimation of the variance factor does not affect the convergence rate of the asymptotically optimal bandwidth. Let b_M denote the theoretically optimal bandwidth. A question that arises is on the relative convergence rate of $\hat{b}_A$ to b_M. Compared to the difference between $\hat{b}_A$ and b_A, the relative difference between b_A and b_M, which is of the order $(b_A - b_M)/b_M = O(b_M^2)$, is asymptotically negligible (Beran and Feng, 2002b). Thus we have

Corollary 3.2 *Under the same assumption as in Theorem 3.2 the relative convergence rates of* $(\hat{b}_A - b_M)/b_M$ *in any cases are the same as those given there.*

The proof of Corollary 3.2 is given in the appendix.

3.6 Application to real financial data

In order to evaluate the practical performance of previously described models and methods, they are applied to the daily average durations (MD), trading number (TrNo) data and volume (Vol) of the German companies Allianz SE (ALV), BMW AG (BMW), Deutsche Bank AG (DBK), Siemens AG (SIE), Thyssenkrupp AG (TKA) and Volkswagen AG (VW). The observation period is from January, 2, 2006, to September, 30, 2014. The ultra-high frequency raw data was retrieved from the Thomson Reuters Tick History Database and processed accordingly.
This section compares the performance of the Semi-Log-ACD modelling approach described in above sections to the Semi-ACD model. For either model, the scale function is estimated from the log-data, so the algorithm for and the results of the scale function estimation in both models are the same. For the Semi-ACD model, the original data is standardized by the retransformed scale function and an EACD(1, 1) model is fitted to the obtained residuals. Concerning the manual choice of some control parameters, we choose a bi-square kernel and the exponential inflation method of Beran and Feng (2002a, 2002b) with an inflation factor of 5/7 in a local linear regression. The time series are considered sufficiently large for this method to work best and a direct comparison with the performance of the Semi-ACD model and the Semi-Log-ACD model is possible.

3.6.1 Performance of the estimated variance factor

Figures A3.1 to A3.3 show the behaviour of the selected bandwidth with the variance factor estimated by the methods described above and different starting bandwidths. Being based on a fix-point search, the bandwidth selected by the IPI algorithm does usually not depend on the value of the starting bandwidth (see e.g. Hermann and Gasser, 1994 or Herrmann, 1997). The figures show, though, that if the variance factor is estimated, the bandwidths are selected differently in some cases depending on b_0. We consider all differences in the selected bandwidths negligibly small, if they are smaller than a convergence criterion of $1/T$. This holds for the majority of considered examples. For some of the data the selection of the bandwidth with $c_{f,\mathrm{ARMA}}$

is instable and the differences in the bandwidths are larger than the convergence criterion. This is the case for the trading number of BMW, the volume of DBK, the trading number of TKA and the volume of VW. The finally selected bandwidth by an algorithm containing $c_{f,\text{ARMA}}$ can also be illogically small or illogically large and fluctuate in value to value, depending on b_0. The bandwidth selections with an IPI containing $c_{f,\text{AR}}$ or $c_{f,\text{NP}}$ are about equally stable. The nonparametric variance factor estimation yields the smallest bandwidths in most cases, the ones based on AR and ARMA select larger bandwidths. The differences in the bandwidths selected by each method are very small, for the convergent bandwidth examples, though. Table A3.1 gives the bandwidths selected by the IPI with each of the variance factor estimation methods, $\hat{b}$ and obtained with a manually fixed $C_M = 4$, $\tilde{b}_{\text{opt}}$ for a starting bandwidth $b_0 = 0.10$. Additionally, the values for $\hat{c}_f$ and $\tilde{c}_f$ are given to compare the estimates with the manually fixed. The table shows that the bandwidths selected by the IPI with a manually obtained variance factor do not differ much from the bandwidths selected by the IPI with an estimated variance factor. Also the estimates of the variance factor are not much different to the manually fixed ones. Overall, the bandwidths are the smallest for the majority of data examples when the variance factor is fixed manually and could be adjusted by changing the value of C_M. However, considering the increased flexibility of automatically estimating the variance factor, we do not propose to include the manually fixed variance factor into the IPI. Also we do not propose the use of the ARMA based method to estimate c_f. Thus, our proposal for the practical application of here described methods is to include $\hat{c}_{f,\text{AR}}$ or $\hat{c}_{f,\text{NP}}$ into the IPI, so a general method that does not require a prior parametric model assumption is proposed, as well as an appropriate model assumption based alternative.

3.6.2 Final analysis

In order to discuss the overall performance of the Semi-Log ACD model the final analysis is carried out for the IPI with $\hat{c}_{f,\text{AR}}$ and $\hat{c}_{f,\text{NP}}$ and $b_0 = 0.10$, i.e. 20% of the observations are used in the first iteration to estimate $\mu(\tau)$. Please note, though, that this is not a general recommendation and was chosen here, because most bandwidths did converge for $b_0 \geq 0.10$. Figure A3.4 shows the original data of the trading number of TKA and VW, together with the retransformed trends estimated with the IPI containing $c_{f,\text{AR}}$ and $c_{f,\text{NP}}$. Also the estimated total mean durations following the Semi-Log-ACD model and the Semi-ACD model are given. Looking at the original data one can see that: 1. A nonparametric trend can be clearly found, so the use of the proposed semiparametric model extension is a reasonable choice. 2. The variation in the data is large when the trend is high, which supports the use of log-transformed data in order to stabilize the variance across time. 3. The underlying distributions of the data are clearly positively skewed, but only a few observations are close to zero. On the one hand, this finding also supports the choice of the log-transformation as it may reduce the skew. On the other hand, this restricts the distribution choices to those who do not have a peak at zero. The discussed special case with a conditional log-normal distribution meets this criterion. Also the distributions discussed in section 3.2.2 except for the Lomax and the exponential distribution do. Looking at the trends estimated via the IPI with the variance factor obtained by the AR and the NP method, the differences are small, especially at the boundaries. The numerical results of the bandwidths selected for estimating the trend of the log-data are given under the figures. They will not be discussed any further than indicating that they seem reasonably selected and the retransformed trends capture the movement of the original series well. For the examples considered here, the differences in the finally selected bandwidths by both methods are the largest for the trading number data of TKA and the smallest for the trading number data of VW. For VW, there is no visible difference in the trend estimations. For TKA minor deviations can be seen in the interior, where the trend estimated with $c_{f,\mathrm{NP}}$ is not as smooth as the one estimated with $c_{f,\mathrm{AR}}$. In order to quantify this visual impression the mean squared scaled differences (MSSD) between the two trend estimates are calculated and given under the figures. Overall, the differences are small and do not rule out the practical application of either method, so our proposal of $c_{f,\mathrm{AR}}$ and $c_{f,\mathrm{NP}}$ is supported by these empirical results. For the graphical final analysis the results for trends estimated by the NP method are shown. For the estimation of the total conditional means, it can be seen that the overall movement of the original data is caught very well in both, the estimations of the Semi-ACD and the Semi-Log-ACD model. The estimated values are at a smaller level than the original ones and the estimated series appear less noisy, especially for the total conditional durations of the Semi-Log-ACD model. Both modelling approaches perform well in practice, but following the ease of estimation and theoretical and practical flexibility of the Semi-Log-ACD model, we recommend this model for practical application.

3.7 Conclusion

We propose a semiparametric extension of the Log-ACD model to further increase its degree of flexibility. The model was discussed as a linear process and asymptotic properties were derived for a special case with a conditional log-normal distribution. For a selection of conditional distributions other than the log-normal one, conditions were derived under which they can serve to model different levels of heavy-tails in the data. We propose to estimate the scale function from the log-data regardless of whether a Semi-ACD or a Semi-Log-ACD model is fitted, as the fourth order moment requirement is less of a problem for Y_t^* than it is for X_t^*. Furthermore, the IPI for estimating the scale function of time series data with correlated errors is generalized to local polynomial regression. Also the estimation of the variance factor is automatized. We describe two parametric and one nonparametric method, where the parametric ones are AR and ARMA model based and latter adapts an already existing IPI. The methods were applied to three data-types of six firms and it is found that the starting bandwidth sometimes affects the finally selected bandwidth, if the variance factor is estimated. For the nonparametric and AR based variance factor estimation, the bandwidths converge, though, whereas the final bandwidth selection with the ARMA method is instable. Following the flexibility and ease of estimation of the Semi-Log-ACD model, we propose the use of the nonparametric variance factor estimation, as no prior assumptions on the model need to be made. The AR based method for estimating the variance factor is an adequate model assumption based alternative. The Semi-ACD and Semi-Log-ACD model yield similar results, but due to the advantages discussed, we propose the use of latter model. The trend estimation procedure is set up very generally, so it may also be applied after minor adaptation to different areas of research. Feng et al. (2018) were already cited to apply an IPI for a local polynomial estimation of the trend and its derivatives in macroeconomic time series together with a data-driven lag-window estimator for the variance factor. The Semi-(Log-)ACD model, introduced here is found to work very well in practice and the theoretical findings of this chapter should be of great value for the development of methods for the further analysis of non-negative financial data, as well as data of different research areas. An especially promising area for further future research could be its application to financial returns in a GARCH model framework. Using local polynomial regression with the methods described in this chapter for the estimation of the time-varying scale in log-returns could improve the Semi-GARCH model of Feng (2004a). Eventually this could improve the calculation of quantitative risk measures, such as the Expected Shortfall, and be a valuable contribution in this field of research.

A.3 Appendix of Chapter 3

Proof of Theorem 3.1. *i*) The mean and variance of $(X_t^*)^m$ follow directly from Lemma 3.1 and well-known properties of the log-normal distribution. The given formulae in this part can be obtained after a straightforward simplification. The variance is of course also a special case of $\gamma_m(k)$ in *ii*) with $k = 0$.

ii) The formula for the autocovariance is derived by subtracting the product of the expectations of lagged values from the expectation of the product of lagged observations. Note that $(X_t^*)^m$ is stationary, so $E[(X_t^*)^m]E[(X_{t+k}^*)^m] = \{E[(X_t^*)^m]\}^2 = e^{m^2\sigma^2}$. $E\left[(X_t^*)^m(X_{t+k}^*)^m\right]$ can be calculated as follows:

$$
\begin{aligned}
E[(X_t^*)^m(X_{t+k}^*)^m] &= E\left(\prod_{i=0}^{\infty}\eta_{t-i}^{ma_i}\prod_{i=0}^{\infty}\eta_{t-i+k}^{ma_i}\right)\\
&= E\left(\prod_{i=0}^{k-1}\eta_{t-i}^{ma_i}\prod_{i=0}^{\infty}\eta_{t-i+k}^{m(a_i+a_{i+k})}\right)\\
&= \prod_{i=0}^{k-1}E\left(\eta_{t-i}^{ma_i}\right)\prod_{i=0}^{\infty}E\left(\eta_{t-i+k}^{m(a_i+a_{i+k})}\right)\\
&= \prod_{i=0}^{k-1}e^{m^2a_i^2\sigma_\epsilon^2/2}\prod_{i=0}^{\infty}e^{m^2(a_i+a_{i+k})^2\sigma_\epsilon^2/2}\\
&= \prod_{i=0}^{\infty}e^{m^2a_i^2\sigma_\epsilon^2/2}\prod_{i=0}^{\infty}e^{m^2a_i^2\sigma_\epsilon^2/2}\prod_{i=0}^{\infty}e^{2m^2a_ia_{i+k}\sigma_\epsilon^2/2}\\
&= e^{m^2\sigma^2}e^{m^2\sigma_\epsilon^2\sum_{i=0}^{\infty}a_ia_{i+k}},
\end{aligned}
$$

as $\eta_t \sim LN(0, \sigma_\epsilon^2)$. The autocovariances are then calculated by

$$
\begin{aligned}
\gamma_m(k) &= e^{m^2\sigma^2}e^{m^2\sigma_\epsilon^2\sum_{i=0}^{\infty}a_ia_{i+k}} - e^{m^2\sigma^2}\\
&= e^{m^2\sigma^2}\left(e^{m^2\sigma_\epsilon^2\sum_{i=0}^{\infty}a_ia_{i+k}} - 1\right).
\end{aligned} \tag{A3.1}
$$

iii) To obtain the autocorrelations $\gamma_m(k)/var(X_t^*)^m$, i.e.

$$
\begin{aligned}
\rho_m(k) &= \left(e^{m^2\sigma_\epsilon^2\sum_{i=0}^{\infty}a_ia_{i+k}} - 1\right)\left(e^{m^2\sigma^2}-1\right)^{-1}\\
&= \left(e^{m^2\sigma^2\rho_{Y^*}(k)} - 1\right)\left(e^{m^2\sigma^2}-1\right)^{-1}.
\end{aligned} \tag{A3.2}
$$

iv) By means of Taylor expansions of both exponential functions on the right-hand side of (A3.2) we obtain

$$\rho_m(k) = \left\{\sum_{i=1}^{\infty}[\rho_{Y^*}(k)m^2\sigma^2]^i/i!\right\}\left\{\sum_{i=1}^{\infty}[m^2\sigma^2]^i/i!\right\}^{-1}. \tag{A3.3}$$

Now we will show that the first sum on the right hand side of (A3.3) is dominated by its first term $\rho_{Y^*}(k)m^2\sigma^2$. Note that $|\rho_{Y^*}(k)| \leq 1$ and $\lim_{k\to\infty}\rho_{Y^*}(k) \to 0$. We have

$$\begin{aligned}\sum_{i=2}^{\infty}|\rho_{Y^*}(k)m^2\sigma^2|^i/i! &\leq [\rho_{Y^*}(k)]^2\sum_{i=2}^{\infty}[m^2\sigma^2]^i/i! \\ &= [\rho_{Y^*}(k)]^2(e^{m^2\sigma^2} - 1 - m^2\sigma^2) \\ &= O\{[\rho_{Y^*}(k)]^2\} \\ &= o[\rho_{Y^*}(k)].\end{aligned}$$

Hence, we have

$$\rho_m(k) = c_\rho(m)\rho_{Y^*}[1 + o(1)], \tag{A3.4}$$

where

$$\begin{aligned} c_\rho(m) &= m^2\sigma^2\left\{\sum_{i=1}^{\infty}[m^2\sigma^2]^i/i!\right\}^{-1} \\ &= m^2\sigma^2[e^{m^2\sigma^2} - 1]. \end{aligned} \tag{A3.5}$$

It is clear that $0 < c_\rho(m) < 1$, for which we have

$$\begin{aligned} [c_\rho(m)]'_m &= 2m\sigma^2\left\{\sum_{i=1}^{\infty}[m^2\sigma^2]^i/i!\right\}^{-1} \\ &\quad -m^2\sigma^2\left\{\sum_{i=1}^{\infty}2im^{2i-1}\sigma^{2i}/i!\right\}\left\{\sum_{i=1}^{\infty}[m^2\sigma^2]^i/i!\right\}^{-2} \\ &= 2m\sigma^2\left\{\sum_{i=1}^{\infty}[m^2\sigma^2]^i/i!\right\}^{-1} \\ &\quad \times\left[1 - \left\{\sum_{i=1}^{\infty}i[m^2\sigma^2]^i/i!\right\}\left\{\sum_{i=1}^{\infty}[m^2\sigma^2]^i/i!\right\}^{-1}\right], \end{aligned}$$

which is negative for any $m > 0$. Thus $c_\rho(m)$ decreases monotonically. Furthermore, note that $\sum_{i=1}^{\infty}[m^2\sigma^2]^i/i! > m^2\sigma^2$. And, if $m \to 0$, we have

$$\begin{aligned}\sum_{i=1}^{\infty}[m^2\sigma^2]^i/i! &= m^2\sigma^2\sum_{i=0}^{\infty}[m^2\sigma^2]^i/(i+1)! \\ &< m^2\sigma^2 e^{m^2\sigma^2} \\ &= m^2\sigma^2[1+o(1)].\end{aligned}$$

Insert those results into $c_\rho(m)$ leads to $\lim_{m\to 0} c_\rho(m) = 1$. If $m \to \infty$, we have

$$m^2\sigma^2 = o\left(\sum_{i=1}^{\infty}[m^2\sigma^2]^i/i!\right).$$

Thence, $\lim_{m\to\infty} c_\rho(m) = 0$. Theorem 3.1 is proved. ◇

Proof of Theorem 3.2. Let C_A denote the constant of b_A and define $b_A = C_A \mathrm{T}^{-1/(2k+1)}$. We have $\hat{b}_A = \hat{C}_A \mathrm{T}^{-1/(2k+1)}$ and

$$(\hat{b}_A - b_A)/b_A = (\hat{C}_A - C_A)/C_A. \tag{A3.6}$$

Through Taylor expansion we obtain

$$\hat{C}_A - C_A \doteq O_p[\hat{I}(\mu^{(k)}) - I(\mu^{(k)})] + O(T^{-1/2}) + O(\hat{c}_f - c_f). \tag{A3.7}$$

The order $O(T^{-1/2})$ is due to the errors in $\hat{\theta}$. According to equation (3) in Bühlmann (1996), following the results in chapter 6.2 of Priestley (1981), the error in the lag-window estimator of c_f using the Bartlett-window and a bandwidth of the optimal order $K = O(\mathrm{T}^{1/3})$ is

$$\hat{c}_f - c_f = O_p(\mathrm{T}^{-1/3}). \tag{A3.8}$$

Both terms are considered negligible concerning the effect on the rate of convergence of $\hat{b}_A$, as they converge much faster than $O_p(\hat{I}(\mu^{(k)}) - I(\mu^{(k)}))$. The orders of magnitude of the dominant bias and variance terms in $\hat{I}(\mu^{(k)}) - I(\mu^{(k)})$ are given in (3.3) to (3.5) in Beran and Feng (2002a) for $k = 2$ or in Theorem 6.1 and Corollary 6.1 of Feng (2004b). Note that $\hat{I}(\mu^{(k)}) - I(\mu^{(k)})$ is at least of the order $O_p(b_{\mathrm{opt}}^2)$, where $b_{\mathrm{opt}} = O(T^{-2/(2k+3)})$ denotes the MISE minimizing bandwidth for estimating $I(\mu^{(k)})$. In either case, the estimation of the variance factor does not affect the rate of convergence of the finally selected bandwidth. The choice of inflation factor affects the rate of convergence. ◇

Proof of Corollary 3.2. The question to address is whether the difference be-

tween b_A and b_M is asymptotically negligible, so that b_A can be used to quantify the quality of $\hat{b}_A$, as done in Theorem 3.2. If the quality of the bandwidth is expressed by the difference between the selected bandwidth $\hat{b}_A$ and the theoretically optimal bandwidth b_M, the difference between b_A and b_M, the error in the model parameter estimation, as well as the error in estimating $I(\mu^{(k)})$ determine it. Here, also the error of estimating the variance factor via an IPI needs to be included.
Following (A.7) in Beran and Feng (2002b) let the rate of convergence be calculated as follows:

$$(\hat{b}_A - b_M)/b_M = (\hat{b}_A - b_A)/b_M + (b_A - b_M)/b_M. \tag{A3.9}$$

In order to discuss $b_A - b_M$ we include a third term into above equation.

$$\begin{aligned}(\hat{b}_A - b_M)/b_M &= (\hat{b}_A - b_A)/b_A + (\hat{b}_A - b_A)/b_M - (\hat{b}_A - b_A)/b_A + (b_A - b_M)/b_M \\ &= (\hat{b}_A - b_A)/b_A + \Big\{[(\hat{b}_A - b_A)/b_M] - [(\hat{b}_A - b_A)/b_A]\Big\} + (b_A - b_M)/b_M.\end{aligned} \tag{A3.10}$$

Let $T_1 = (\hat{b}_A - b_A)/b_A$ denote the first term of (A.10), which is at least of the order $O(T^{-2/(2k+3)})$ as shown in Theorem 3.2. Let $T_3 = (b_A - b_M)/b_M$ denote the third term of (A.10). Extending the proof of Proposition 1 of Beran and Feng (2002b) to a local polynomial short-memory case following Feng (2004b), we have

$$M_1 = O(b^{2k}) + O\{(Tb)^{-1}\},$$

where M_1 denotes the asymptotic MISE described in (3.13), which is obtained based on the k-th order approximation of the bias

$$B_1 = O(b^k) + o(b^k).$$

If $\mu^{(k+2)}$ is continuous the bias can be approximated by

$$B_2 = O(b^k) + O(b^{(k+2)}) + o(b^{(k+2)})$$

to obtain a more accurate approximation of the MISE, M, by

$$M_2 = O(b^{2k}) + O(b^{2(k+1)}) + O\{(Tb)^{-1}\}.$$

In the neighbourhood of h_M, the second term in the variance part is always negligible and

$$M(b_M) - M_1(b_M) \doteq M_2(b_M) - M_1(b_M) \doteq O(b_M^{2(k+1)}). \tag{A3.11}$$

Through Taylor expansion we obtain

$$b_A - b_M \doteq -[M_1'(b_M) - M'(b_M)]/M''(b_M) \tag{A3.12}$$
$$\doteq -[M_1'(b_M) - M_2'(b_M)]/M''(b_M). \tag{A3.13}$$

Note that $M_1'(b_M) - M_2'(b_M) \doteq O(b_M^{2k+1})$ and $M''(b_M) \doteq O(b_M^{2(k-1)})$ so

$$(b_A - b_M)/b_M \doteq O(b_M^2). \tag{A3.14}$$

Note that $b_M = O(T^{-1/(2k+1)})$, so $b_M^2 = O(T^{-2/(2k+1)})$. Recall that T_1 is at least of the order $O(T^{-2/(2k+3)})$, hence we have $b_M^2 = o(T_1)$. This holds for all cases considered in Theorem 2, since for $\upsilon = 1/2$ and $\upsilon = (2k+1)/(2(k+2)+1)$ the orders of the bias terms of $\hat{b}_A$ are larger than for $\upsilon = (2k+1)/(2k+3)$. The variance terms are negligible compared to the ones for estimating the variance factor and as explained, the $O_p(T^{-1/3})$ terms are asymptotically negligible compared to the bias terms. To discuss $T_2 = [(\hat{b}_A - b_A)/b_M] - [(\hat{b}_A - b_A)/b_A]$ it can be rewritten as

$$\begin{aligned} T_2 &= (\hat{b}_A - b_A)\left[\frac{1}{b_M} - \frac{1}{b_A}\right] \\ &= (\hat{b}_A - b_A)\left(\frac{b_A - b_M}{b_A b_M}\right) \\ &= (\hat{b}_A - b_A)/b_A \cdot (b_A - b_M)/b_M \\ &= T_1 T_3. \end{aligned} \tag{A3.15}$$

$T_2 = o(T_3)$ and $T_3 = o(T_1)$, so it is shown that the difference between b_A and b_M does not affect the rate of convergence and is, therefore, negligible for quantifying the quality of the finally selected bandwidth in Theorem 3.2. ◇

Table A3.1 – Bandwidth selection with $b_0 = 0.1$, different methods for obtaining c_f

		$\hat{b}_{\mathrm{AR}}$	$\hat{b}_{\mathrm{ARMA}}$	$\hat{b}_{\mathrm{NP}}$	$\hat{c}_{f,\mathrm{AR}}$	$\hat{c}_{f,\mathrm{ARMA}}$	$\hat{c}_{f,\mathrm{NP}}$	$\tilde{b}_{\mathrm{opt}}$	$\tilde{c}_f$
ALV									
	MD	0.1046	0.1073	0.0970	1.2387	1.3783	0.9446	0.1007	1.1859
	TrNo	0.1051	0.1074	0.0974	1.2591	1.3757	0.9546	0.1006	1.1911
	Vol	0.1298	0.1342	0.1239	1.2117	1.3912	0.9966	0.1205	1.1974
BMW									
	MD	0.1020	0.1029	0.0954	0.8861	0.9146	0.6921	0.1016	0.8905
	TrNo	0.1026	0.1034	0.0958	0.9088	0.9385	0.7063	0.1017	0.9010
	Vol	0.1107	0.1074	0.1056	0.9784	0.8576	0.8014	0.1049	1.0005
DBK									
	MD	0.1222	0.1154	0.1141	1.2707	0.9900	0.9463	0.1107	1.0439
	TrNo	0.1227	0.1160	0.1148	1.2760	0.9981	0.9586	0.1111	1.0530
	Vol	0.1170	0.1130	0.1125	1.3564	1.1372	1.1132	0.1075	1.2095
SIE									
	MD	0.1215	0.1147	0.1176	1.1636	0.9065	1.0099	0.1154	1.2326
	TrNo	0.1208	0.1144	0.1161	1.0868	0.8649	0.9170	0.1143	1.1505
	Vol	0.1344	0.1283	0.1289	1.0737	0.8844	0.9054	0.1270	1.1047
TKA									
	MD	0.1251	0.1258	0.1171	0.7743	0.7918	0.6043	0.1188	0.7089
	TrNo	0.1256	0.1242	0.1171	0.7869	0.7757	0.6039	0.1186	0.7050
	Vol	0.1280	0.1306	0.1233	1.0623	1.1692	0.8962	0.1258	1.1002
VW									
	MD	0.1180	0.1290	0.1172	2.3890	3.8192	2.3090	0.1145	2.4198
	TrNo	0.1176	0.1286	0.1166	2.3329	3.7027	2.2402	0.1144	2.3486
	Vol	0.1005	0.1182	0.1000	5.4786	10.7414	5.3383	0.0889	4.9629

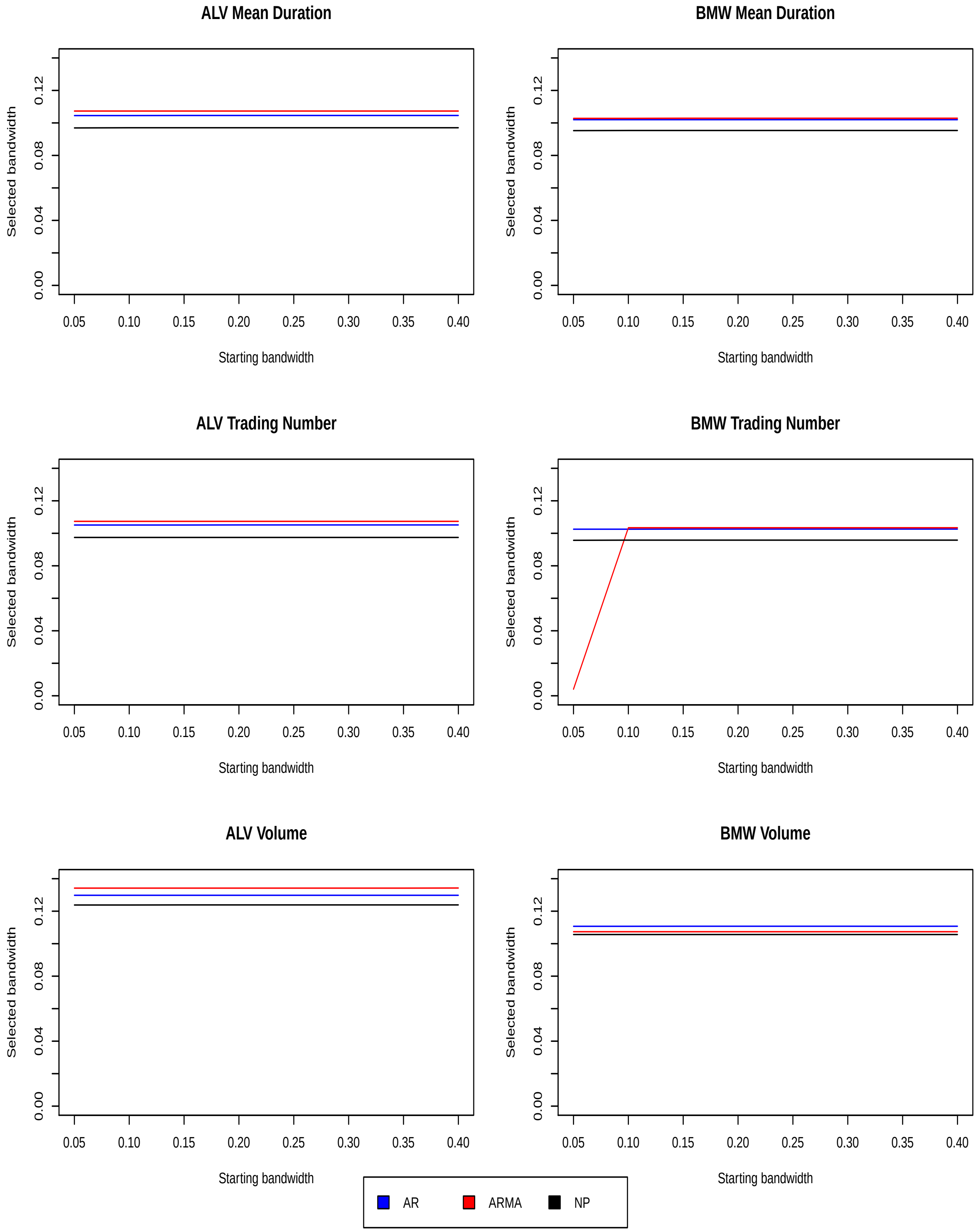

Figure A3.1 – Comparison of different methods with different starting bandwidths for ALV and BMW

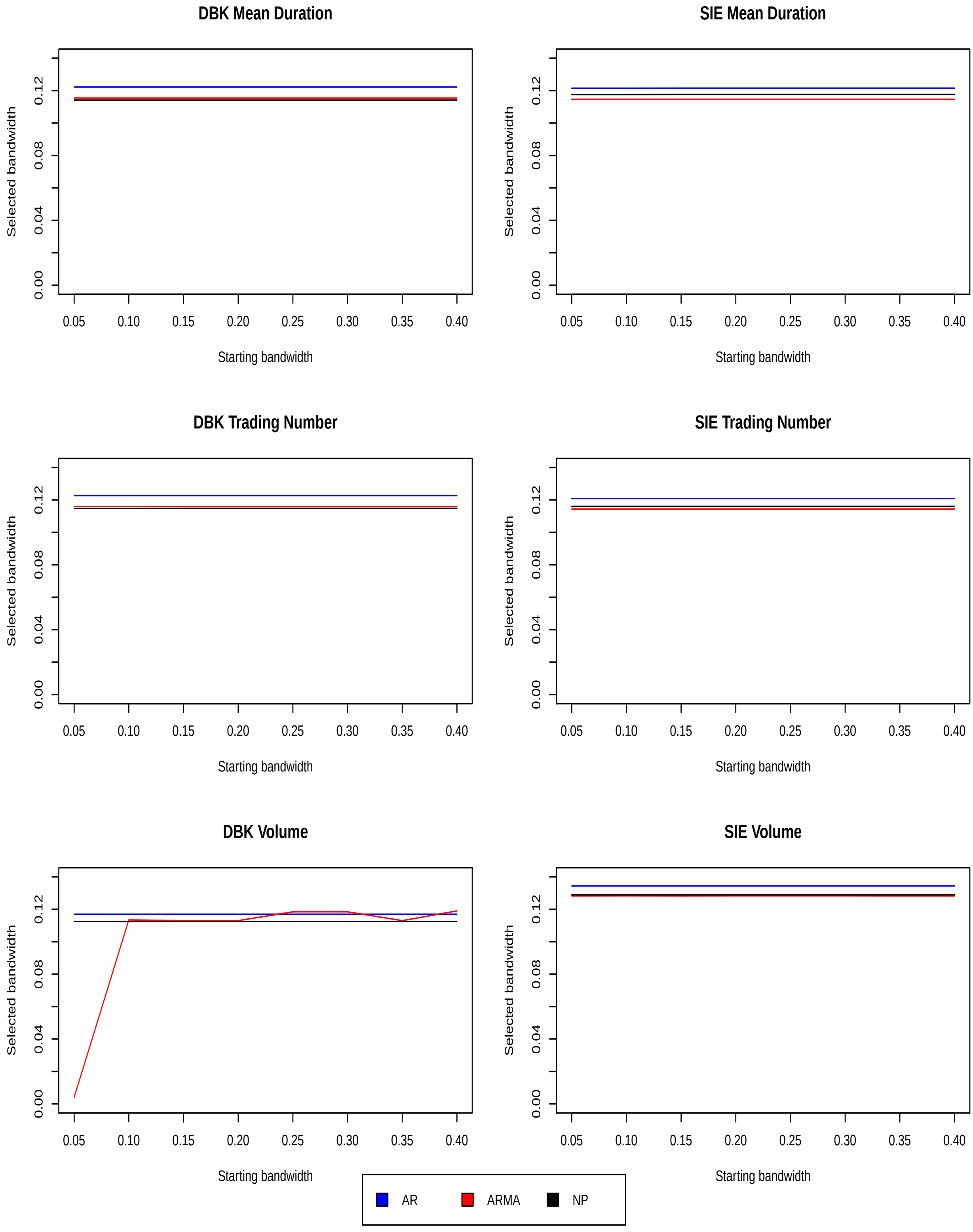

Figure A3.2 – Comparison of different methods with different starting bandwidths for DBK and SIE

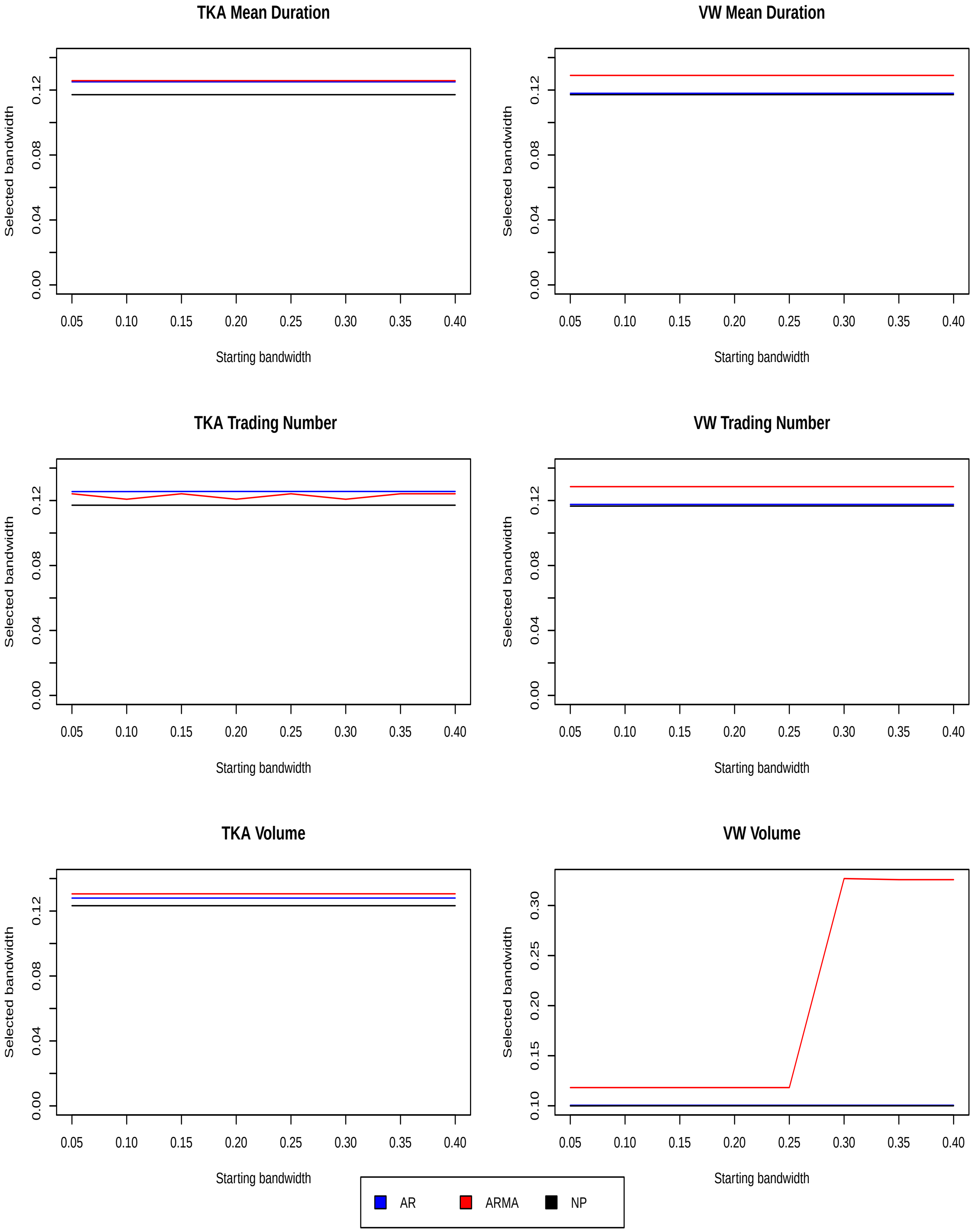

Figure A3.3 – Comparison of different methods with different starting bandwidths for TKA and VW

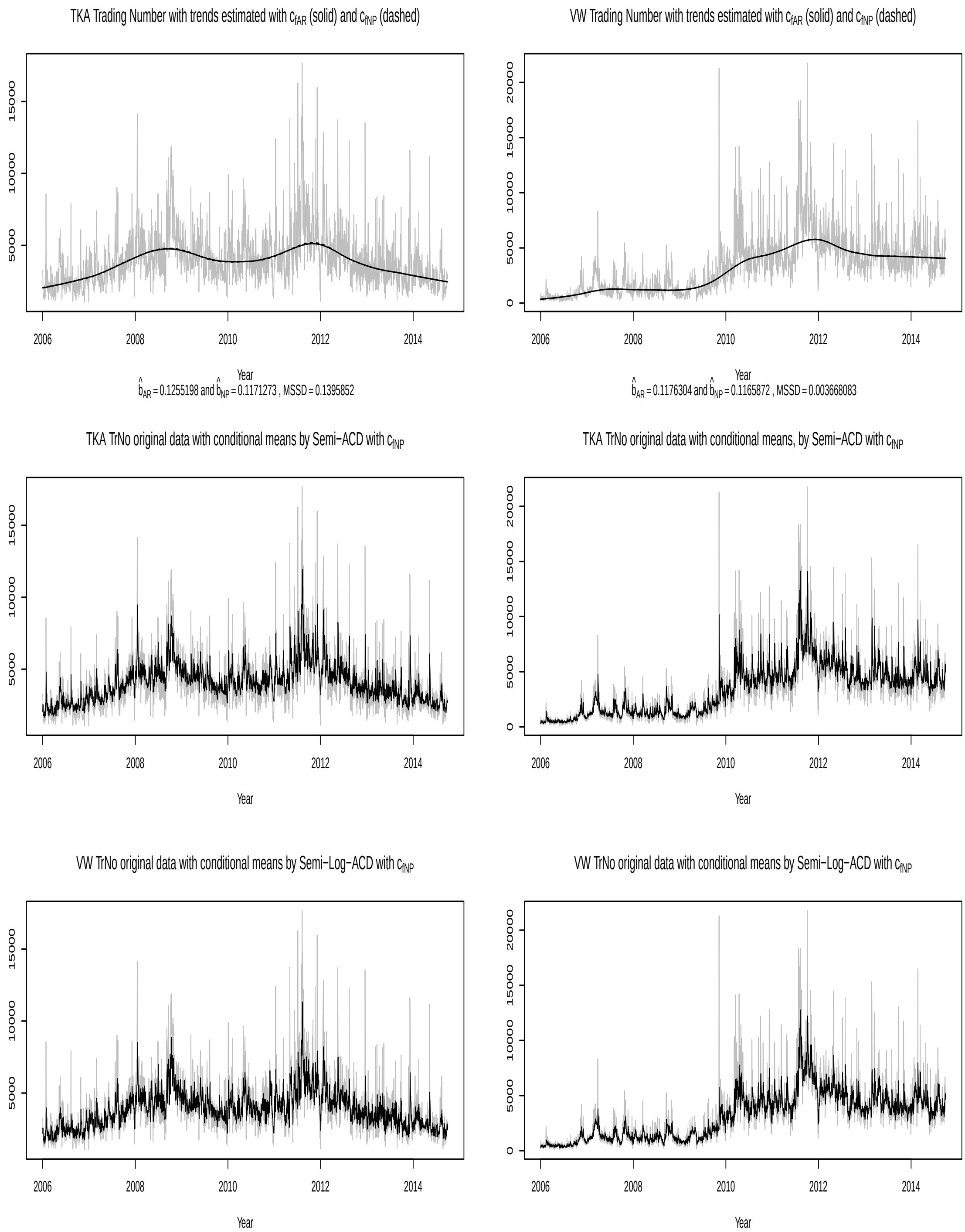

Figure A3.4 – TrNo TKA and VW with estimated trends and total conditional means estimated by Semi-(Log-)ACD

4 Forecasting non-negative financial processes using different parametric and semiparametric ACD-type models

4.1 Introduction

De Gooijer and Hyndman (2006) review '25 years of time series forecasting' until 2005. They consider heavy computation methods, realized volatilities (see Andersen et al., 2003) and transaction durations (see Engle and Russell, 1998) as important areas, which slowly began to be the subject of forecasting method developments. Their review includes a reference to Cogger (1988) calling for a departure from linear models with a too strict assumption of Gaussian i.i.d. errors. While the research on the forecasting of realized volatilities is very active, this is not the case for the forecasting of transaction durations. This chapter is part of the research on the further development of the ACD model, which already underwent several stages and is still subject to further development (see e.g. Pacurar, 2008; Saart et al., 2015 or Kaur Bhogal and Thekke Variyam, 2018).

In the following, different forecasting approaches are proposed based on the semiparametric extensions of the ACD model of Engle and Russell (1998) and the (first-type) Log-ACD model of Bauwens and Giot (2000) and Bauwens et al. (2008). Above mentioned issues are addressed by forecasting methods that do not require a prior parametric model assumption or that are restricted to a particular conditional distribution. ACD models are considered, as their properties and flexibility make them be of great value for the analysis of financial market data. With their semiparametric extension and generalization to multiplicative errors models (MEM; see Engle, 2002) more flexibility is gained. The general setup of the MEM for any sequence of non-negative random variables allows for its application to data, other than transaction durations. Different types of marked durations can be modelled, as well, where the durations not only measure the time elapsed, but the time until e.g. a minimum mid-price change is observed or a minimum volume is traded (see e.g. Giot, 2000 and Fernandes and Grammig, 2006). It can also be used to model (daily) average volume (Manganelli, 2005) and its semiparametric extensions include amongst others nonparametric methods for estimating the unknown model parameter vector (e.g. Drost and Werker, 2004 or Ranasinghe and Silvapulle, 2011) or are based on decomposing the data and applying nonparametric methods to at least one of the components (e.g. Engle and Russell, 1998; Brownlees et al., 2010 or Brownlees et al., 2012). Here, the data is decomposed into a deterministic and a stochastic part and estimated separately in order to account for intra-daily patterns (see McInish and Wood, 1992; Veredas et al., 2001; Bauwens et al., 2004 or Feng et al., 2016) or long-term behaviour (see Forstinger and Feng, 2018). The deterministic movement of the data is described by including a time-varying scale function into the model. It is proposed to be estimated from the log-data via local polynomial regression with automatic bandwidth and variance factor selection (Forstinger and Feng, 2018). After this step any parametric model can be fitted to the detrended data.

The trend is extracted in a first step and existing and new forecasting methods are applied to the obtained residuals. Eventually they are combined with the estimated and linearly extrapolated trend. For existing methods the Semi-Log-ACD model with a conditional log-normal distribution is used. For obtaining the point and 90% forecasting-intervals Kalman filter methods are applied. To overcome the issue that parametric forecasting interval methods perform poorly for non-Gaussian errors (see e.g. Stine, 1987 or Li and Maddala, 1996), another nonparametric bootstrap approach is proposed. Due to the correlation structure of the process the i.i.d. assumption of the observations for the bootstrap of Efron (1979) does not apply to time series data (e.g. Chatterjee, 1986 or Bühlmann, 2002). Therefore, the sieve bootstrap idea of Bühlmann (1997) and its extensions (Alonso et al., 2002; 2003; 2004) is adapted to the Semi-Log-ACD model. Analogously bootstrap for GARCH models is extended to the Semi-ACD model following Pascual et al. (2006) and Chen et al. (2011). Once the replicate data is generated, it is used for obtaining the point forecasts and the confidence intervals as the quantiles of the bootstrap distribution function of the forecasted bootstrap replicates in either case (see e.g. Pascual et al., 2006; Hwang and Shin, 2013 or Allende et al., 2015).

The remainder of this chapter is organized as follows: Section 4.2 gives an overview over the ACD and Log-ACD model and their semiparametric extensions. Section 4.3 describes the forecasting methods for the trend and the residuals. The semiparametric models, as well as their parametric counterparts are applied to three data types of six firms in section 4.4. The performance of the forecasts is discussed for training and validity sets with different forecast horizons. Section 4.5 summarizes the results and main findings.

4.2 The Semi-ACD and Semi-Log-ACD model

4.2.1 General setup

Let the observations of interest be denoted by $X_t \geq 0$ for $t = 1, ..., T$. The very basic setup of the ACD model by Engle and Russell (1998) follows the idea of a classical MEM (Engle, 2002) of decomposing X_t into a scale parameter, the conditional mean and a positively valued error term: $X_t = \nu\psi_t\varepsilon_t$, with $\nu > 0$. Different possibilities exist for the functional form of the conditional mean, $\psi_t > 0$, as well as for the distribution of the i.i.d. non-negative random variables, ε_t.[9] Here, the conditional mean function follows a linear parametrisation, as given in the original proposal of Engle and Russell (1998):

$$\psi_t = \omega + \sum_{j=0}^{p} \alpha_j X^*_{t-j} + \sum_{j=0}^{q} \beta_j \psi_{t-j}, \tag{4.1}$$

where $X^*_t = X_t/\nu$. To ensure non-negativity of ψ_t, $\omega > 0$, $\alpha_j \geq 0$ and $\beta_j \geq 0$ for $j = 1, ..., p$ and q. These constraints on the parameters can be dropped for the (first-type) Log-ACD model of Bauwens and Giot (2000) and Bauwens et al. (2008). Let $Y_t = log(X_t)$, $\mu = log(\nu)$, $\lambda_t = log(\psi_t)$ and $\eta_t = log(\varepsilon_t)$, so that $Y_t = \mu + \lambda_t + \eta_t$. In order to model the conditional and long-term dynamics simultaneously, the scale parameter is replaced with a nonparametric smooth scale function $\nu(\tau_t) > 0$ and $\mu(\tau_t)$, where $\tau_t = t/T$ denotes the rescaled time in either case. The estimation of the scale function and its removal from the data is important, because it is found that the ACD model is inconsistent, if a significant trend is not removed (see Feng et al., 2016). The semiparametric extension of the ACD model is obtained by

$$X_t = \nu(\tau_t)\psi_t\varepsilon_t \tag{4.2}$$

and (4.1). The Semi-Log-ACD model is given by

$$Y_t = \mu(\tau_t) + \lambda_t + \eta_t \tag{4.3}$$

and the corresponding conditional mean function $\lambda_t = \omega + \sum_{j=0}^{p} \alpha_j Y^*_{t-j} + \sum_{j=0}^{q} \beta_j \lambda_{t-j}$, where $Y^*_t = Y_t - \mu(\tau_t)$. The Log-ACD model and the Semi-ACD model are each more flexible than the baseline ACD model of Engle and Russell (1998). Applying semi-parametric methods to the Log-ACD model further increases the flexibility and eases

[9]Other specifications of the ACD model, such as mixture and component models, or regime-switching ACD models are not considered here (see chapter 6 of Hautsch, 2011 for an overview of further specifications).

the estimation of the trend, due to the additive structure of the model. Additionally, the local linear regression IPI of the Semi-ACD model of Feng (2014) and Feng et al. (2016) can be extended to local polynomial regression for the Semi-Log-ACD model.

4.2.2 Semiparametric model estimation

The Semi-ACD and the Semi-Log-ACD model are fitted in two steps, where first the scale function is estimated nonparametrically and removed from the observations. In a second step, the parametric models are fitted to the residuals (see Feng, 2014; Feng et al., 2016 and Forstinger and Feng, 2018). Different approaches to account for the deterministic trend are developed in the literature. Engle and Russell (1998), for example, propose cubic splines, but Feng et al. (2016) identify it as performing not as reliably good as the local linear regression of the Semi-ACD model. Dufour and Engle (2000a,b) propose linear splines and Brownlees and Gallo (2011) describe a shrinkage type estimator. Here, the scale function is proposed to be equivalently estimated via local polynomial regression of (4.3) for $\mu(\tau_t)$, as the existence of moments required for the development of the bandwidth selection algorithm is less of a problem for financial processes then (see Forstinger and Feng, 2018).
A local polynomial estimator for the log-scale function minimizes the weighted sum of least squares $Q = \sum_{t=1}^{T}\{Y_t - \sum_{j=1}^{p}\beta_j(\tau_t-\tau)^j\}^2 K\left(\frac{\tau_t-\tau}{b}\right)$. A weight kernel function enters this equation as $K(u)$, which is the asymptotically equivalent kernel of $\hat{\mu}(\tau_t)$ at an interior point and a boundary kernel at a boundary point. The estimator has automatic boundary correction and the local polynomial regression achieves the global optimal rates of convergence, if the difference between the orders, p and v of the polynomial regression and the derivative, respectively, is odd (see Fan et al., 1996; Cheng et al., 1997 and Feng and Beran, 2013). The bandwidth enters the weighted least squares equation as b and is estimated automatically via an IPI (see Gasser et al., 1991 for the original idea of the IPI; Hermann et al., 1992; Brockmann et al., 1993 and Hermann, 1997 for its extension to allow for local adaptivity and Forstinger and Feng, 2018 for its application to the Semi-Log-ACD model). Deriving the asymptotic properties of $\hat{\mu}^{(v)}(\tau_t) = v!\hat{\alpha}_v$ eventually leads to the formula for the global asymptotical optimal bandwidth, which minimizes the asymptotic mean integrated squared error. For $K(u)$ let $R(K) = \int K^2(u)du$, $I(K) = \int u^k K(u)du$ and $I(\mu)^k = \int[\mu^k(\tau)]^2 d\tau$. The variance factor c_f denotes the value of the spectral density of Y_t at the origin. The sum of the residual autocovariances of Y_t^* can be written as $2\pi c_f = \sum_{-\infty}^{\infty}\gamma(c)$, for which $\gamma(c) = \sigma^2\sum_{i=0}^{\infty}\alpha_i\alpha_{i+|c|}$:

$$b_A = \left(\frac{2v+1}{2(k-v)}\frac{2\pi c_f R(K)[k!]^2}{I(K)^2 I([\mu^{(k)}])}\right)^{1/(2k+1)} T^{-1/(2k+1)}. \quad (4.4)$$

The closed form expression of the residual autocovariances is derived from the MA(∞) representation of the standardized log-data, $Y_t^* = \sum_{i=0}^{\infty} \alpha_i \eta_{t-i}$. As one of the two unknowns in (4.4), $I(\mu^{(k)})$ can be estimated by already known methods for models with i.i.d. errors. The unknown variance factor c_f, however, might be affected by non i.i.d. errors (see Opsomer et al., 2001). A nonparametric estimator of c_f is proposed in this context by Forstinger and Feng (2018), together with a corresponding IPI following Bühlmann (1996). The estimator $\hat{c}_{f,NP} = 1/2\pi \sum_{c=-C_k}^{C_K} \omega_c \hat{\gamma}(c)$ is based on $2\pi c_f$ and Bartlett window weights ω_c. Instead of fixing the window-width manually by choosing a value for C_K an IPI selects the optimal window width by minimizing the asymptotic mean square error at a fixed frequency of the lag-window estimator (see Bühlmann, 1996 for the original idea and Forstinger and Feng, 2018 for the adapted implementation). Both IPI that apply to the trend estimation of the Semi-Log-ACD model are described in detail in Forstinger and Feng (2018), so another repeated description is abstained from here.

After the scale function estimation, the data of interest is standardized and the unknown model parameters can be estimated from the residuals. Let the unknown parameter vector of Y_t^* be denoted by $\theta = (\omega, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)'$. Under the assumption that $\hat{\mu}(\tau_t)$ is a consistent estimator of $\mu(\tau_t)$, $\hat{\theta}$ can be obtained via corresponding (Quasi-)Maximum Likelihood estimators based on the distribution of the innovations. Forstinger and Feng (2018) show, that the quality of the estimation of the parameters is not affected by the estimation of the scale function and the errors caused by the estimation of the variance factor are asymptotically negligible, as well.

4.3 Forecasting methods

The methods described in this section forecast the estimated trend and the residuals separately and eventually combine the forecasts. For all discussed semiparametric methods, we propose to forecast $\hat{\mu}(\tau_T)$ by means of linear extrapolation. For some forecast horizon $k = 1, 2, 3...$:

$$\hat{\mu}(\tau_{T+k}) = \hat{\mu}(\tau_T) + k\Delta\mu, \tag{4.5}$$

where $\Delta\mu = \hat{\mu}(\tau_T) - \hat{\mu}(\tau_{T-1})$. For the forecasting approach using the Semi-ACD model, $\hat{\mu}(\tau_t)$ is retransformed and forecasting methods apply to $\hat{X}_t^* = X_t/\exp(\hat{\mu}(\tau_t))$, whereas for the Semi-Log-ACD model forecasting methods retransformation is the final step: $\hat{X}_{T+k} = \exp(\hat{Y}_{T+k})$.

For forecasting the detrended data, we propose a total of three methods. One method uses the representation of Y_t^* through an ARMA(p, q) model (see Allen et al., 2008).

It puts itself forward for forecasting purposes, as already well studied methods and tools are readily available. Brockwell and Davis (2002) propose an approximately best linear predictor, based on the state space representation of the stationary ARMA process. In a long-memory framework Feng and Zhou (2015) apply this predictor for forecasting standardized log-data following a FARIMA process. Furthermore, two nonparametric bootstrap forecasting methods for the Semi-ACD and the Semi-Log-ACD model are proposed to avoid making assumptions about the distribution for the innovation. All methods described in the following also apply to the corresponding parametric models, i.e. the pre-step of removing the estimated scale function is not carried out and the forecasting methods apply to the (log-transformed) data directly.

4.3.1 ARMA(p, q) Kalman filter forecast

The Kalman filter forecast for the Semi-Log-ACD model under a conditional normal distribution and its implementation follow Harvey (1990), Harvey and McKenzie (1982) and Brockwell and Davis (1991, 2002). The focus of this chapter is on the bootstrap methods discussed in 4.3.2, so the methodological elements of the Kalman filter forecast are only outlined. It serves as a well-established forecasting technique to which the newly proposed methods are compared against later. For more details on assumptions and proofs, the reader is referred to aforementioned literature. Let Y_t^* follow a stationary and invertible ARMA model (see Allen et al., 2008):

$$\phi(B)Y_t^* = \psi(B)\eta_t, \tag{4.6}$$

where $\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$ is the characteristic polynomial of the AR-part and $\psi(z) = 1 - \psi_1 z + ... + \psi_q z^q$ is the characteristic polynomial of the MA-part. Per assumption they do not have common factors and all of the roots of $\phi(z) = 0$ and $\psi(z) = 0$ are outside the unit circle. The prediction method employed here is based on the state space representation of (4.6) and a Kalman filter for calculating the best linear mean square predictors.

1. Prediction of Y_{T+k}^* based on the ARMA(p, q) process given in (4.6)
   Based on the state space representation of the likely stationary ARMA process, $Y_t^* = G_t\tilde{X}_t + W_t$, where the state equation $\tilde{X}_{t+1} = F_t\tilde{X}_t + V_t$, the finite sample predictions are calculated by applying Kalman recursions. W_t and V_t are i.i.d. with zero means and known variances, $Var(V_t) = Q_t$ and $Var(W_t) = R_t$, as well as a known covariance matrix $Cov(V_t, W_t) = S_t$. G_t and F_t are the coefficient matrices. Following Brockwell and Davis (1991, 2002), the k-step

prediction, already including the correction step and Kalman gain is

$$\hat{\hat{X}}_{T+k} = F_{T+k-1}\tilde{X}_{T+k-1}, \tag{4.7}$$

which starts with $k = 1$:

$$\hat{\hat{X}}_{T+1} = F_T\hat{\hat{X}}_T + \Theta_T\Delta_T^{-1}(Y_T^* - \hat{Y}_T^*). \tag{4.8}$$

The covariance matrices of the prediction errors are $\Theta_n = F_T\Omega_T G_T' + S_T$ and $\Delta_T = G_T\Omega_T G_T' + R_T$ with the error covariance matrices $\Omega_T = E((\tilde{X}_T - \hat{\hat{X}}_T)(\tilde{X}_T - \hat{\hat{X}}_T)')$. The corresponding forecasting errors in terms of the covariance matrices are

$$\Delta_T^{(k)} = G_{T+k}\Omega_T^{(k)}G_{T+k}' + R_{T+k}, k = 1, 2, ... \tag{4.9}$$

with $\Omega_T^{(k)} = F_{T+k-1}\Omega_T^{(k-1)}F_{T+k-1}' + Q_{T+k-1}$. Accordingly, the observation predictions are given by:

$$\hat{Y}_{T+k}^* = G_{T+k}\hat{\hat{X}}_{T+k} \tag{4.10}$$

2. Prediction of $\hat{Y}_{T+k} = \hat{Y}_{T+k}^* + \hat{\mu}(\tau_{T+k})$.
3. Prediction of $\hat{X}_{T+k} = \exp(\hat{Y}_{T+k}) = \exp[\hat{Y}_{T+k}^* + \hat{\mu}(\tau_{T+k})]$.

The corresponding formulas of forecasting intervals are:

$$Y_{T+k} \in \left(\hat{\mu}(\tau_{T+k}) + \hat{Y}_{T+k}^* \pm q_{\frac{\alpha}{2}}\sqrt{V_{Y_{T+k}^*}}\right) \tag{4.11}$$

and

$$X_{T+k} \in \left(\exp[\hat{\mu}(\tau_{T+k}) + \hat{Y}_{T+k}^* \pm q_{\frac{\alpha}{2}}\sqrt{V_{Y_{T+k}^*}}]\right). \tag{4.12}$$

4.3.2 (Log-) ACD model bootstrap forecast

In the following, bootstrap methods are introduced for both the Semi-ACD and the Semi-Log-ACD model forecast. The sieve bootstrap idea of Bühlmann (1997) is extended. Unlike the classical bootstrap, it does not require i.i.d. observations, but takes into account the correlation structure of the process. It resamples the innovations obtained from the fitted model, instead of resampling the observations directly:

Stage I: Semiparametric fitting of the ACD model

1. Consider the Semi-ACD model as set up by (4.2) and (4.1) and obtain $\hat{\mu}(\tau_t)$ following section 4.2.2. Calculate $\hat{X}_t^* = X_t/\exp[\hat{\mu}(\tau_t)]$.

2. Fit an ACD(p, q) model, for which the orders are chosen via BIC and obtain $\hat{\psi}_t$ from the model estimation. Also obtain $\hat{\varepsilon}_t = \hat{X}_t^*/\hat{\psi}_t$ and $\hat{\theta}$.

Stage II: Simulation of m forecast series for a forecast horizon of k each

1. Define a matrix, X_B, with m rows and k columns.

2. In the first loop, let i run from 1 to m. For each i draw $\varepsilon_i^B = (\varepsilon_{i,1}^B, ..., \varepsilon_{i,k}^B)'$, which are sampled with replacement.

3. In the second loop, let i again run from 1 to m and j run from 1 to k. Calculate $\hat{\psi}_{i,T+j}^B = \hat{\omega} + \sum_{r=1}^{\hat{p}} \hat{\alpha}_r X_{i,T+j-r}^* + \sum_{s=1}^{\hat{q}} \hat{\beta}_s \psi_{i,T+j-s}$ and let $X_{i,T+j}^B = \hat{\psi}_{i,T+j}^B \varepsilon_{i,j}^B$. Set $X_{i,t}^* = \hat{X}_t^*$ for $t = 1, ..., p$ and $\psi_{i,l} = \hat{\psi}_l$ for $l = 1, ..., q$.

   For $t \geq p+1$ and $l \geq q+1$, set $X_{i,t}^* = X_{i,t}^B$ and $\psi_{i,l} = \hat{\psi}_{i,l}^B$, respectively.

4. Put $X_B[i,j] = X_{i,n+j}^B$.

Stage III: Calculation of the point and interval forecasts

1. Let the point forecasts be defined by $X_m^* = E(X_B[,1:k])$, the mean of the columns of X_B and define $m_{low} = [m * 0.025 + 0.5]$ and $m_{up} = [m * 0.975 + 0.5]$.

2. Let X_L^* and X_U^* be two vectors of length k.

3. Let i run from 1 to m. Let $X_B^{(i)}$ be the ordered vector of $X_B[i,1:k]$. Put $X_L^* = X_B^{(i)}[m_l]$ for the lower bootstrap-forecasting bound and $X_U^* = X_B^{(i)}[m_u]$ for the upper bootstrap-forecasting bounds.

The bootstrap forecast for the Semi-Log-ACD model overall follows the same stages and steps accordingly and is not given here.

4.4 Application to real financial data

For comparison and evaluation purposes, the described forecasting methods are applied to different types of non-negative financial data. The raw data was retrieved from the Thomson Reuters Tick History Database and processed accordingly. In particular, the daily average durations (MD), the daily trading volume (Vol) and the realized volatility (RV), defined as the sum of squared intraday log-returns, of six German DAX30 companies are analysed: Allianz SE (ALV), BMW AG (BMW), Deutsche Bank AG (DBK), Siemens AG (SIE), Thyssenkrupp AG (TKA) and Volkswagen AG (VW). The period of analysis starts in January 2006 and ends in September 2014. All data were standardized or centralized by their mean in a first step, to make sure that the level of each data is accurate by treating the means as global constants. In order to analyse the obtained results appropriately significance tests of the estimated scale functions and the normality assumption of the data are carried out. The 95% confidence intervals for the estimated log-scale functions were calculated to test whether they are statistically significantly different to the mean of each observed log-data series. To show that this insignificance is not a result of the log-transformation, the same test procedure was carried out for the original data. The results of these tests are that all of the estimated scale functions are statistically significantly different to the corresponding mean and, thus, the semiparametric methods are expected to be clearly better than the parametric methods. The visualization of the tests for the log-data are given in figures A4.1 to A4.3. It is apparent that for all data types used, clearly more than 5% of the mean of the log-data are outside of the calculated confidence interval bounds. The %-values of deviations are given in table B4.1. For testing the normality of the data, the Shapiro Wilk test (SW; see Shapiro and Wilk, 1965) is used and it is found that for all of the examples or their transformations the normal distribution hypothesis is rejected. The p-values of each SW test are given in table B4.1l.[10]

For the semiparametric methods, the scale function $\mu(\tau_t)$ is estimated via the IPI bandwidth selection algorithm, described in section 4.2.2. A local linear regression with a bi-square kernel, the exponential inflation method with an inflation factor of 5/7 and a starting bandwidth of $b_0 = 0.10$ is applied. For all examples forecast horizons of $k = 10$ to $k = 100$ are used and the point and $1 - \alpha\%$ interval forecasts are obtained for $\alpha = 10\%$. A total of six forecasting methods are applied, i.e. the bootstrap ACD model, the bootstrap Log-ACD model and the Kalman filters for the Log-ACD model with a conditional normal distribution. Same methods are applied semiparametrically. A training set of $n = T - 100$ observations and a validity set of k observations are used to assess the quality of the point and interval forecasts. To ensure that the forecasting quality can be adequately assessed, the point from which the forecasts start is the same for all k. The performance of the point and interval forecasts are evaluated by two criteria each.

[10]The SW test is carried out due to its expressive power, availability and common usage. A discussion on normality tests and drawbacks of the SW test are not included.

4.4.1 Quality of point forecasts

For assessing the quality of the point forecasts, the mean absolute scaled error (MASE) of Hyndman and Köhler (2006) is used:

$$MASE = mean\left(\left|\frac{e_t}{\frac{1}{n-1}\sum_{t=2}^{n}|X_t - X_{t-1}|}\right|\right), \tag{4.13}$$

where $e_t = X_t - \hat{X}_t$. As a second criterion the mean absolute error, standardized by the corresponding mean values of the training sets is calculated:

$$SMAE = \frac{\sum_{t=n-k+1}^{n}|e_t|}{\sum_{t=1}^{n-k}X_t}. \tag{4.14}$$

The MASE is scale-independent, less sensitive to outliers and does not require the observations to be sufficiently far from 0. Therefore, it is suitable for comparing different methods applied to different types of data, as is the case in this application section. Also it is easy to interpret, because the forecasting method of interest is evaluated against a one-step naïve forecast. If MASE >1 the one-step naïve forecast is better and v.v. (see e.g. Armstrong and Collopy, 1992 or Franses, 2016). The standardized MAE is included in the performance discussion to have a criterion without a reference method it is compared against. It also less sensitive to outliers than other commonly used scale-dependent measures and the standardization makes it comparable across the different data types and methods used (Hyndman and Köhler, 2006).

4.4.2 Quality of Forecasting Intervals

To take into account the quality of the forecasting intervals, the mean lengths of the forecasting intervals are proposed as an evaluation criterion. For the bootstrap methods the intervals are calculated with the quantiles of the bootstrap distribution function. For the ARMA based forecasts, the quantiles of the standard normal distribution are used:

$$LFI = mean\left(\frac{FI_{upper,k} - FI_{lower,k}}{\sum_{t=1}^{n-k}X_t}\right). \tag{4.15}$$

In order to make the results comparable among methods and data types, each length value is standardized by the corresponding mean value of the training set. Naturally, the shorter the interval, the more precise the interval estimation is assumed. In addition, a newly developed criterion is proposed, the "Error of points outside Forecasting Interval", EoPoFi. Let X_t be the validity set and $\text{Low}_{t|t-1}$ and $\text{Up}_{t|t-1}$

the lower and upper bounds of the forecasting interval obtained for t at $t-1$. Let $\text{PoFI}_{up} = \#(X_t > \text{Up}_{t|t-1})$ and $\text{PoFI}_{low} = \#(X_t < \text{Low}_{t|t-1})$. The EoPoFI is given by

$$\text{EoPoFI} = |(\text{PoFI}_{up} + \text{PoFI}_{low}) - k\alpha|. \tag{4.16}$$

The idea of using the points outside the forecasting intervals is related to the idea of the Peak over Threshold Approach, PoT, in Extreme Value Theory (see Leadbetter, 1991) and takes on the interval forecasts evaluation methods of Christoffersen (1998). Latter do not require a distributional assumption be made on the process of interest and therefore fit well in the bootstrap forecasting framework (see e.g. Reeves, 2005). The EoPoFI compares the theoretical number of points with the actual number of points outside the forecasting intervals and, thus, gives a good measure for discussing the intervals' quality in terms of meeting the theoretical deviations. The EoPoFI is scale-independent and can be used to compare different methods for different data types. Also it does not require a distribution assumption, so it is suitable for evaluating our bootstrap forecasting intervals, but also for assessing the quality of the forecasting intervals obtained under the normal distribution assumption. Furthermore, it supplements the evaluation of the length of the forecast intervals. A short interval in which the true values of the validity set exceed the limits excessively is certainly not to be described as precise for practical use. The two criteria for judging the interval forecasts should therefore not be interpreted separately.

4.4.3 Discussion of results

For discussing the forecast quality of the methods presented here, forecasts for ten different forecast horizons $k = 10, 20, ..., 100$ are calculated. The numerical averages of the evaluation criteria are summarized in table 4.1 for $k = 50$ and $k = 100$ and illustrated graphically in figure 4.1 over all k. Please note that for the graphical analysis, the EoPoFi values are standardized, i.e. the average number of points outside the forecasting interval is shown. Additionally, the detailed results of the MASE of all methods for all data examples and data types are given in table 4.2 for discussing the quality of point forecasts for $k = 50$ and $k = 100$. The results of all criteria for all forecast horizons are given in tables B4.2 to B4.21.
Figure 4.1 directly shows that the Semi-ARMA model consistently has the smallest average values in all evaluation criteria and thus the best forecast quality, on average. Conversely, the ACD bootstrap method has the worst forecast quality for the examples considered. The graphical results also show that the semiparametric methods are generally superior to the parametric ones. The MASE of the semiparametric forecasts even improves with increasing k. The results of the SMAE give the same

indication for the forecasting quality of point forecasts. For discussing the quality of the forecasting intervals, their mean length results show that they converge for the semiparametric methods, whereas those of the parametric methods become wider with increasing k. Despite the narrow intervals for the semiparametric methods only very few forecast values are outside the interval bounds. In contrast, no less predictive values lie outside the interval limits of the parametric methods, even though their forecasting intervals are wider. The graphical results are supported by the numerical results of table 4.1. Even though the results on only two forecast horizons are shown, the general tendency of the results can be clearly seen. On average, the Semi-ARMA model is the best forecasting method, the semiparametric methods perform about equally well and for the parametric methods the forecast quality of the ARMA model is consistently ranked fourth, the Log-ACD model bootstrap forecasts fifth and the ACD bootstrap model is ranked sixth for all criteria.

The detailed results for all forecast horizons indicate that the forecasting of MD is the most difficult and individually considered no better than the one-step naïve method. The values of the other criteria are also above average for MD forecasts. Despite the findings from the average results, for some single cases (e.g. MD BMW or MD VW, $k = 10$), the ACD bootstrap point forecasts are best. The bootstrap simulated observations follow a direction that is usually compensated for by the linear extrapolation of the nonparametric estimated trend or is not very pronounced due to the standardization of the data in advance. If the direction of the simulated data reflects the movement of the true observations well, the ACD bootstrap method is better than the semiparametric methods. Figure 4.2 shows the point forecast for the mean durations of BMW and DBK for $k = 50$ and $k = 100$ according to the semiparametric and the parametric ACD bootstrap model. In the DBK example, the ACD model bootstrap point forecast for $k = 50$ has the (slightly) smallest MASE and SMAE value overall, for $k = 100$ it is the Semi ACD bootstrap model. For BMW the ACD bootstrap model forecasts are the best in either case. In the BMW example, one can clearly see in both cases how the estimated trend determines the direction of the forecast values on the one hand, and on the other hand that these do not take the direction of the true observations. In contrast, the prediction of the ACD bootstrap model fits better due to its gradient, which is not changed by the declining scale. In the DBK example for $k = 50$ you can see that the slowly decreasing predictive values according to the ACD bootstrap method correspond more to the true observational values than those of the Semi-ACD bootstrap method. All in all, the difference is minimal (graphical and numerical). For $k = 100$ of the same example, however, you can see that on average the declining values of the ACD bootstrap method miss the true observation values in the long-run and the forecasts of the Semi-ACD bootstrap model fit better. The ACD bootstrap forecasts work either very well if the bootstrap prediction is able to map the data well, or very poorly. The Semi-ACD model bootstrap forecasts lie in between, so that they are better, on average, than those of the ACD model bootstrap. The results discussed above do not have to be revised, though, the Semi-ARMA method is the best and the ACD bootstrap method the worst, on average.

Table 4.1 – Average results of all criteria for k=50 and k=100

Criterion	mean	Semi-ACD BS		Semi-Log-ACD BS		SemiARMA		ARMA		ACD BS		Log-ACD BS	
		k=50	k=100	k=50	k=100	k=50	k=100	k=50	k=100	k=50	k=100	k=50	k=100
MASE		0.8366	0.7911	0.8514	0.8052	0.8206	0.7774	0.9650	0.9457	1.2710	1.3728	0.9975	1.0131
	MD	1.4292	1.3807	1.5833	1.4888	1.5659	1.4764	1.7117	1.6122	1.6598	1.6663	1.6888	1.5807
	RV	0.3187	0.2949	0.2238	0.2247	0.2101	0.2169	0.3870	0.4585	1.1011	1.4232	0.4943	0.6686
	Vol	0.7619	0.6976	0.7470	0.7021	0.6860	0.6389	0.7964	0.7665	1.0521	1.0289	0.8092	0.7901
	rank	2	2	3	3	1	1	4	4	6	6	5	5
SMAE		0.2515	0.2365	0.2506	0.2373	0.2390	0.2267	0.2874	0.2855	0.4116	0.4565	0.3023	0.3153
	MD	0.3612	0.3484	0.4024	0.3770	0.3964	0.3727	0.4322	0.4065	0.4194	0.4249	0.4265	0.3990
	RV	0.1327	0.1229	0.0934	0.0937	0.0877	0.0904	0.1603	0.1900	0.4548	0.5919	0.2050	0.2779
	Vol	0.2605	0.2384	0.2561	0.2412	0.2328	0.2169	0.2697	0.2599	0.3605	0.3526	0.2753	0.2690
	rank	3	2	2	3	1	1	4	4	6	6	5	5
length FI		0.9097	0.9086	0.8853	0.8952	0.8572	0.8579	1.0654	1.2339	1.4375	1.6605	1.1468	1.4041
	MD	1.1469	1.1540	1.0697	1.0745	1.1268	1.1332	1.1960	1.2963	1.4455	1.5829	1.1738	1.2937
	RV	0.5976	0.5863	0.5222	0.5185	0.4803	0.4742	0.8405	1.1434	1.5305	1.9449	1.0221	1.5272
	Vol	0.9847	0.9853	1.0641	1.0926	0.9645	0.9661	1.1597	1.2620	1.3366	1.4538	1.2445	1.3913
	rank	3	3	2	2	1	1	4	4	6	6	5	5
EoPoFi		3.3889	6.4444	4.2778	6.7222	3.0000	6.1667	5.3889	9.7222	10.6111	18.8333	7.6111	13.5556
	MD	3.1667	7.5000	5.5000	8.6667	3.6667	8.1667	4.3333	7.6667	5.1667	10.0000	4.5000	6.6667
	RV	2.6667	4.6667	3.1667	5.0000	1.8333	4.5000	5.6667	10.1667	17.3333	30.1667	11.1667	21.5000
	Vol	4.3333	7.1667	4.1667	6.5000	3.5000	5.8333	6.1667	11.3333	9.3333	16.3333	7.1667	12.5000
	rank	2	2	3	3	1	1	4	4	6	6	5	5
mean rank		2.50	2.25	2.50	2.75	1.00	1.00	4.00	4.00	6.00	6.00	5.00	5.00
mean diff		0.7344		0.5921		0.7602		1.0761		2.0253		1.4849	

Table 4.2 – MASE results for k=50 and k=100

Data	Type	Semi-ACD BS		Semi-Log-ACD BS		SemiARMA		ARMA		ACD BS		Log-ACD BS	
		k=50	k=100	k=50	k=100	k=50	k=100	k=50	k=100	k=50	k=100	k=50	k=100
ALV	MD	2.7910	2.7299	3.0808	2.9244	3.0781	2.9237	3.6880	3.5174	3.9883	3.8941	3.6189	3.4353
	RV	0.3574	0.3930	0.2190	0.2701	0.2096	0.2633	0.3217	0.4336	0.8985	1.1501	0.4761	0.7241
	Vol	0.5956	0.6732	0.5132	0.6014	0.5086	0.5979	0.7221	0.8201	0.9173	1.0800	0.7234	0.8337
BMW	MD	1.1983	1.1389	1.2815	1.1945	1.3102	1.2195	1.1842	1.0622	0.7626	0.9507	1.2100	1.0811
	RV	0.2856	0.2705	0.2234	0.2327	0.2108	0.2294	0.4657	0.5229	1.1979	1.4280	0.5774	0.7245
	Vol	0.5018	0.4954	0.4280	0.4420	0.4274	0.4423	0.5592	0.5668	0.7083	0.7454	0.5678	0.5794
DBK	MD	1.1911	1.2595	1.1594	1.2749	1.1588	1.2635	1.1657	1.3334	1.1576	1.3194	1.1668	1.3177
	RV	0.1926	0.1770	0.1777	0.1654	0.1863	0.1754	0.2191	0.2776	0.8866	1.9952	0.3122	0.5167
	Vol	1.4812	1.2091	1.5664	1.2829	1.5665	1.2848	1.5243	1.2418	1.3988	1.1620	1.4466	1.1950
SIE	MD	1.6146	1.4778	1.7929	1.6211	1.8045	1.6342	1.8894	1.6133	1.4862	1.3322	1.8177	1.5263
	RV	0.3211	0.3322	0.2173	0.2335	0.2149	0.2327	0.3703	0.4674	1.0359	1.3401	0.4946	0.6986
	Vol	0.3360	0.3251	0.2980	0.3012	0.2854	0.2961	0.4074	0.4087	0.4908	0.4972	0.3962	0.4007
TKA	MD	1.4347	1.3799	1.7121	1.5369	1.6327	1.4837	1.9573	1.8273	2.2318	1.9864	1.9444	1.7999
	RV	0.5404	0.4092	0.3395	0.2920	0.2928	0.2585	0.6859	0.7660	2.0773	1.9826	0.7862	0.9408
	Vol	0.6413	0.5698	0.5750	0.5074	0.5576	0.4936	0.8381	0.8773	1.6907	1.5946	0.8710	0.9297
VW	MD	0.3453	0.2984	0.4731	0.3809	0.4108	0.3337	0.3856	0.3200	0.3321	0.5148	0.3752	0.3239
	RV	0.2152	0.1875	0.1657	0.1546	0.1462	0.1421	0.2593	0.2833	0.5103	0.6431	0.3193	0.4069
	Vol	1.0157	0.9130	1.1014	1.0777	0.7703	0.7186	0.7271	0.6843	1.1066	1.0943	0.8505	0.8018
mean		0.8366	0.7911	0.8514	0.8052	0.8206	0.7774	0.9650	0.9457	1.2710	1.3728	0.9975	1.0131
mean w/o MD		0.5403	0.4963	0.4854	0.4634	0.4480	0.4279	0.5917	0.6125	1.0766	1.2261	0.6518	0.7293
mean MD		1.4292	1.3807	1.5833	1.4888	1.5659	1.4764	1.7117	1.6122	1.6598	1.6663	1.6888	1.5807
mean RV		0.3187	0.2949	0.2238	0.2247	0.2101	0.2169	0.3870	0.4585	1.1011	1.4232	0.4943	0.6686
mean Vol		0.7619	0.6976	0.7470	0.7021	0.6860	0.6389	0.7964	0.7665	1.0521	1.0289	0.8092	0.7901

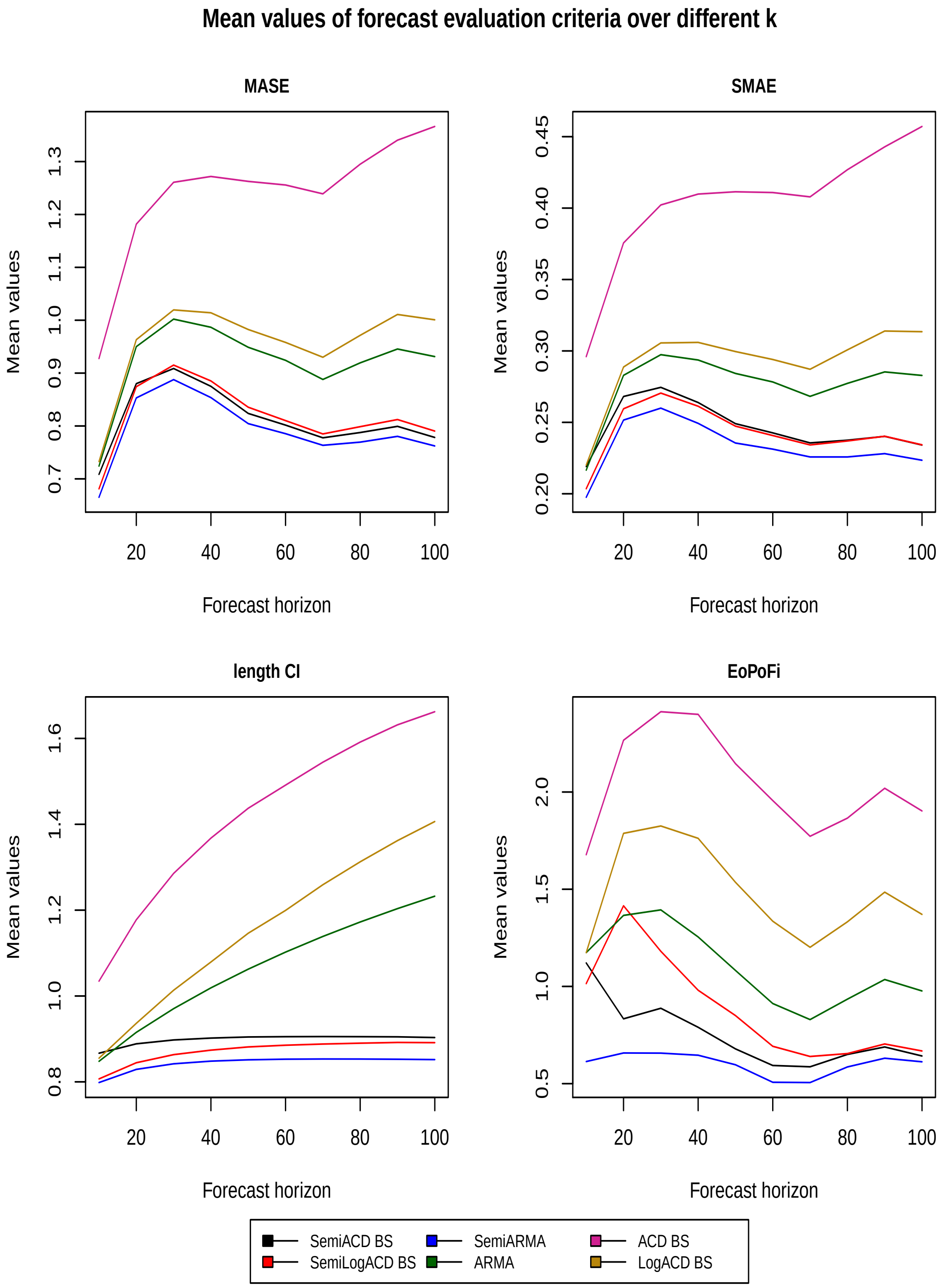

Figure 4.1 – Mean values of forecast evaluation criteria over k

4.5 Conclusion

The aim of this chapter was to develop methods that close a gap in the research on forecasting non-negative financial market data and trade durations, in particular. To this end, already known and new methods have been combined. Semiparametric extensions of the ACD and Log-ACD models are extended to a forecasting framework. The first step of equivalently estimating the scale function with a local polynomial regression is extended to forecast the estimated trend via linear extrapolation. The second step uses bootstrap methods on the residuals of the parametric model fitted to the residuals. Kalman filters are used as a model based alternative for the Semi-Log-ACD model with a conditional log-normal distribution. The semiparametric forecasting models, as well as their parametric counterparts are applied to different data-types of six firms. The performance of the point and 90% interval forecast performance was assessed for a training data set and 10 different forecast horizons with two criteria each. The EoPoFi was newly proposed in this chapter to evaluate the quality of interval forecasts, when used additionally to the mean length of forecasting intervals criterion, for example. The results show that the semiparametric methods clearly outperform the parametric model forecasts. Following the criteria used, the forecasting quality is better and even improves with an increasing forecast horizon. Furthermore, the lengths of the forecasting intervals converge and the number of validity observation points outside the intervals' boundaries does not increase. Thus, the forecasting of non-negative financial data with semiparametric (Log-) ACD models is shown to work well in practice. In particular, the Semi-Log-ACD model with a conditional normal distribution consistently performs the best. The ACD bootstrap method is consistently the worst, on average, but in detail it is shown to work either very well or very poorly. A closer examination of when the simulated bootstrap forecast captures the course of the actual data or not would be a subject that should be considered in future research. In this context, it could also be interesting to consider trend forecasts other than the linear extrapolation or to apply bootstrap techniques, other than the one proposed.

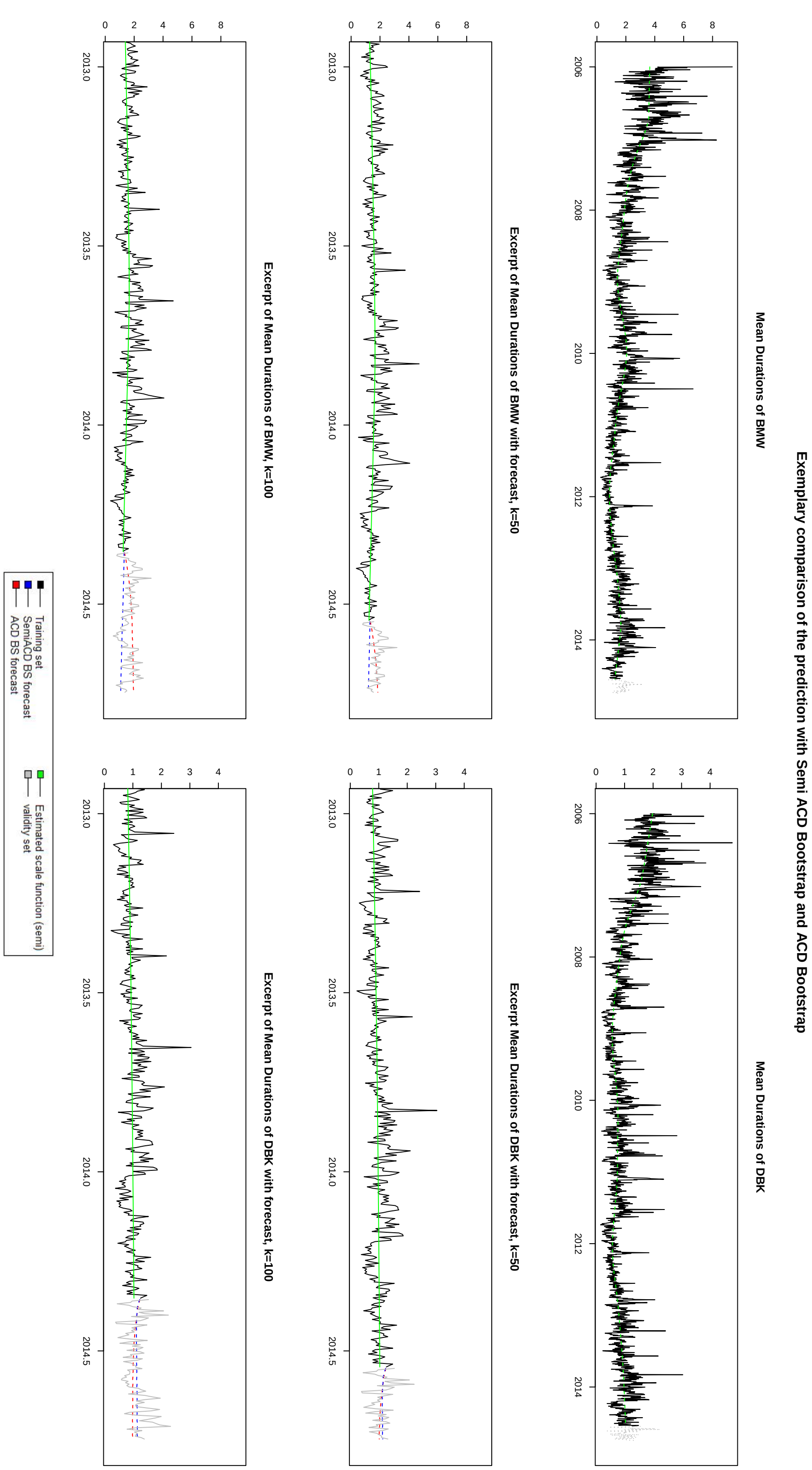

Figure 4.2 – Graphical Comparison of Semi-ACD Bootstrap and ACD Bootstrap forecasts

A.4 Appendix of Chapter 4

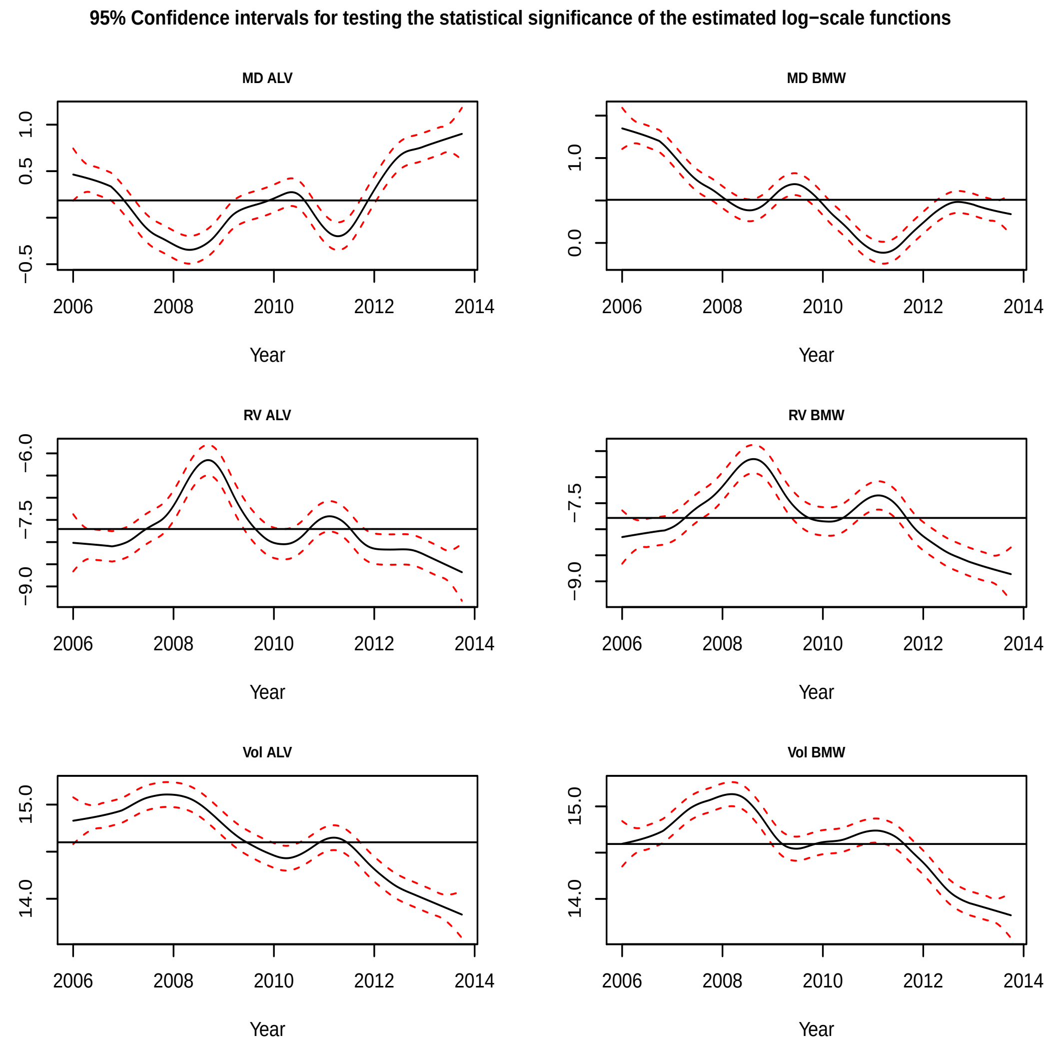

Figure A4.1 – 95% confidence intervals for testing statistical significance for all data type of ALV and BMW

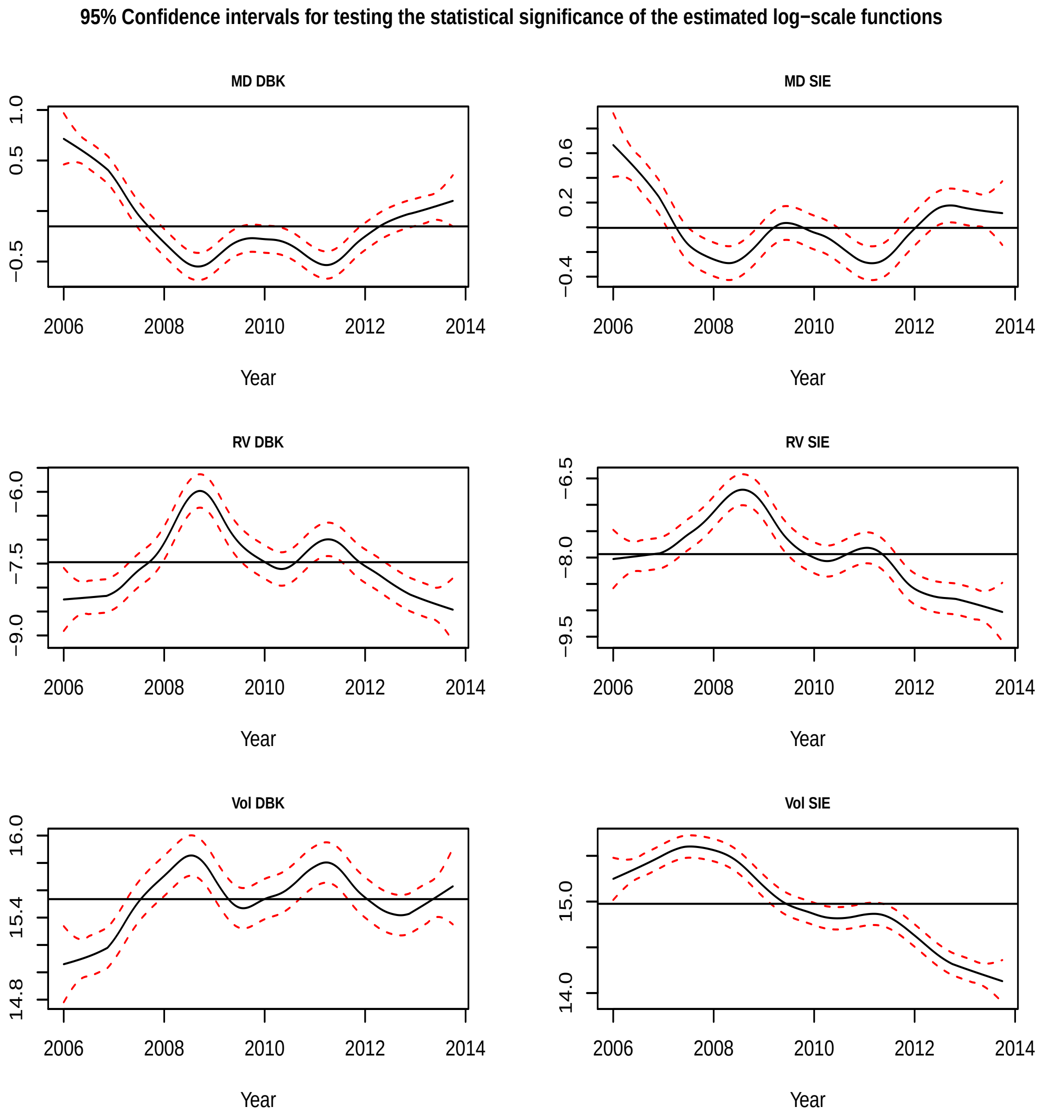

Figure A4.2 – 95% confidence intervals for testing statistical significance for all data types of DBK and SIE

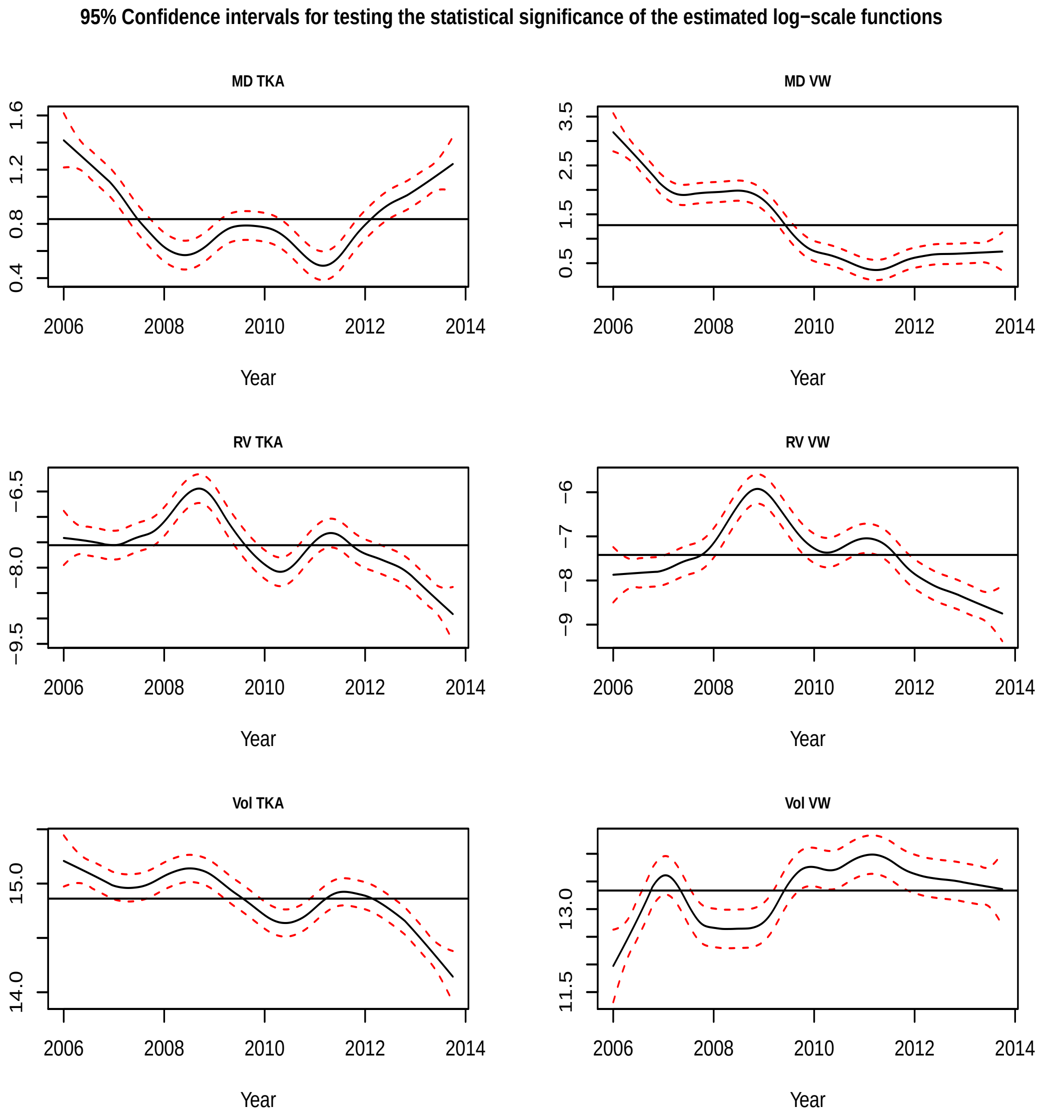

Figure A4.3 – 95% confidence intervals for testing statistical significance for all data types of TKA and VW

5 Forecasting Economic Growth Processes: Improved Forecast Accuracy by Combining Local Linear and Standard Measures[11]

5.1 Introduction

"A fundamental problem of economic forecasting is that many economic variables are inherently very difficult to forecast, and despite advances in data availability, theory, and computational power, we have not seen dramatic improvements in forecast accuracy over the past decades" (Stock and Watson, 2017, p. 70). Obviously, even forecasting of macroeconomic variables for advanced economies is a challenging task. Moreover, for developing economies the lack of data, low data frequency, high volatility, and often highly non-linear developments are even more of a severe problem and restrict forecasting for macroeconomic time series further. This data scarcity leads to complex forecasting problems which have not yet been sufficiently resolved. Nevertheless, forecasting macroeconomic variables for developing economies is increasingly important, not only for the countries themselves. Today, developing and emerging economies like China, India, or Brazil count for a large share of the world economy. Hence, developments of these countries have a major impact on global macroeconomic processes and reliable predictions of important macroeconomic variables in these countries are crucial for estimates of global processes and interactions.

Fildes and Stekler (2002) formulate four requirements for economic forecasting. The forecast should show the economy's growth direction, turning points and magnitudes as well as the time-length of movement persistence. Their survey summarizes different forecasting techniques separated into naïve, structural and time series forecast models. They conclude that time series forecasts from ARIMA, VAR and BVAR models sometimes outperform their structural alternatives. In contrast to Stock and Watson (2017), De Gooijer and Hyndman (2006) demonstrate refinements in time series' predictive accuracy by examining methodological improvements since the 1980s

[11] This chapter is available as a pre-print at Paderborn University and was written in joint work with Professor Dr. Yuanhua Feng, Professor Dr. Thomas Gries and Marlon Fritz (all Paderborn University). It is referred to as Fritz et al. (2018) or chapter 5 interchangeably in this thesis.

in their extensive survey about time series forecasting. Clements and Hendry (1998) demonstrate a detailed comparison of different forecasting models using different measures of forecast accuracy.

Stekler (2007) summarizes that forecasting models are used to make statements about the future under the assumption that the structure of the past process will not change in the future. Moreover, the forecasts need a starting point and thus are individually based on the underlying data. However, as Clements and Hendry (2002) argue, structural changes in deterministic trends produce serious forecast failures. Amongst others, Castle et al. (2010) and Castle et al. (2016) propose to consider structural breaks in order to avoid forecast failure. Therefore, Castle et al. (2016) apply different forecasting models, e.g. VAR and EqCM, which are demonstrated using UK GDP growth and unemployment during the 2008 crisis.

However, recent results show that there is no forecasting model which is preferred for all data sets. In other words, the accuracy of the forecasting method highly depends on the country and time horizon under consideration. Amongst others, Fildes and Stekler (2002) state that combining forecast methods improves accuracy and that the refinement is the most when different methods are used. Stock and Watson (2004) demonstrate the performance of forecast combinations of output growth for seven OECD countries. Surprisingly, they find the lowest mean squared forecast errors (MSFE) for the simple combination methods with equal weights (see also Hendry and Clements, 2004; Smith and Wallis, 2009). Furthermore, Hibon and Evgeniou (2005) argue that it is riskier to rely on a single forecast method, although the best individual forecast is not worse than the best combinations. However, in practice the best individual forecast method is unknown; hence combining forecasts significantly improves predictive performance.

Meese and Rogoff (1983) demonstrate that the random walk forecast for exchange rates is preferred over its structural alternatives (Meese-Rogoff Puzzle). Since then, there have been many analyses of the Meese-Rogoff Puzzle for monetary variables of different countries; however their results seem to remain valid. Marçal and Junior (2016) go a step further and combine VAR and VEC models with random walk models (with and without drift) for the Brazilian real. Calculating the MFSE, they conclude that for the short-term forecast accuracy for some combinations is equivalent to the random walk without drift. However, in general the random walk without drift is the most appropriate benchmark for the Brazilian real.

In this chapter, we propose the use of a nonparametric trend estimation approach in order to improve forecast accuracy. Feng et al. (2016), for example, introduce the semiparametric ACD model for analysing the duration between consecutive transactions on financial markets. Allen et al. (2008) show that the Log-ACD model is

equivalent to an ARMA model for log-data and Feng and Zhou (2015) propose an extension of the semiparametric Log-ACD model for forecasting purposes. We adapt these methods accordingly. Our method is, in accordance with Feng et al. (2018), a local linear trend estimation, where the bandwidth is estimated using a data-driven IPI algorithm for bandwidth selection. Furthermore, a slightly adjusted IPI algorithm is included to estimate the variance factor nonparametrically. This approach allows for a fully data-driven determination of the optimal bandwidth for the trend estimation. Consequently, the trend is adjusted smoothly to the underlying observations and enables a more appropriate approximation of the underlying long-term growth process since it is compatible with log-linear growth theories. A more accurate calculation of the variance factor further enhances the quality of the estimated prediction intervals. In addition, we adapt the results of Meese and Rogoff (1983) and propose the extension of the random walk model with drift by including the local linear trend as the drift. After the bandwidth is selected and the trend is estimated, we use these methods to forecast economic variables. Following Fildes and Stekler (2002) as well as Marçal and Junior (2016), the local linear estimation is combined with different random walk models to improve forecast accuracy. Therefore, and in order to comply with the requirements of De Gooijer and Hyndman (2006), we calculate point as well as interval predictions for each individual method and each possible combination. The results demonstrate that the combinations using the local linear estimation approach and the recently developed random walk with local linear drift are able to improve forecasting accuracy for advanced and developing economies. This inference is also reflected in the variance that is most stable in the combinations using our estimation method.
The rest of the chapter is structured as follows. Section 5.2 introduces the data, the IPI algorithm and the local linear estimation. Section 5.3 presents the forecasting methods, a new measure for assessing the forecasting quality and the construction of forecasting intervals. Section 5.4 compares the forecast accuracy of different models and their combinations. Section 5.5 concludes.

5.2 Data and semiparametric model

5.2.1 Data

To demonstrate the usefulness of our approach, we apply the data-driven IPI algorithm together with the forecasting methods to six different developing economies and two advanced economies. The data set includes (log) real GDP series, in constant 2011 national prices (in mil. 2011 US Dollars) for Brazil (1950-2014), China (1952-

2014), India (1950-2014), Mexico (1950-2014), Nigeria (1950-2014), South Africa (1950-2014) as well as the United States (1950-2014) and the United Kingdom (1950-2014). The series are extracted from the Penn World Table Version 9.0 of Feenstra et al. (2015). India and Nigeria represent the lower middle-income countries, whereas Brazil, China, Mexico and South Africa are examples of upper middle-income countries.

5.2.2 Semiparametric model

This section presents a data-driven local linear estimation approach for macroeconomic variables, also called the semiparametric regression model, the local linear regression (LLR) or model (5.1) in this chapter. In accordance with Beran and Feng (2002a), as well as Feng et al. (2018), a time series Y_t with time $t = 1, ..., T$ is decomposed into

$$Y_t = m(x_t) + Z_t, \tag{5.1}$$

where $x_t = t/n$ denotes the rescaled time, $m(x_t)$ is some smooth trend function and Z_t denotes a stationary process. For simplicity it is further assumed that Z_t follows an ARMA model. We extend the approach of Beran and Feng (2002a) and apply it to macroeconomic time series with short-range dependence. Thus, in accordance with Feng et al. (2018), the ν-th derivative of $m(x_t)$, defined as $m^{(\nu)}(x)$, is estimated by minimizing the locally weighted least squares:

$$Q = \sum_{t=1}^{n} \left\{ y_t - \sum_{j=0}^{p} \beta_j (x_t - x)^j \right\}^2 W\left(\frac{x_t - x}{h}\right), \tag{5.2}$$

where W and h are the weight function and relative bandwidth, respectively. To be consistent with log linear growth theories we use $p = 1$. The obtained trend estimates are $\hat{m}^{(\nu)}(x) = \nu! \hat{\beta}_\nu$. We propose to choose the asymptotical global optimal bandwidth h_A, for estimating $m^{(\nu)}$ on [0,1], by minimizing the asymptotic mean integrated squared error (AMISE):

$$AMISE(h) = h^{2(k-\nu)} \frac{I[m^{(k)}]\beta^2}{[k!]^2} + \frac{2\pi c_f R(K)}{nh^{2\nu+1}}. \tag{5.3}$$

The bandwidth is then calculated using

$$h_A = \left(\frac{2\nu + 1}{2(k - \nu)} \frac{2\pi c_f [k!]^2 R(K)}{I[m^{(k)}]\beta^2_{(\nu,k)}} \right)^{\frac{1}{2k+1}} n^{\frac{-1}{2k+1}}, \tag{5.4}$$

where $k = p+1$, $I[m^{(k)}] = \int [m^{(k)}(x)]^2 dx$, $\beta_{(\nu,k)} = \int_{-1}^{1} u^k K(u) du$, and $R(K) = \int K^2(u) du$. $c_f = f(0)$, the variance factor, is estimated in a data-driven manner using a slightly adjusted IPI algorithm, where $f(\lambda)$ denotes the spectral density of a stationary error process. To determine the bandwidth, the key point is to estimate the variance factor c_f correctly. Therefore, we introduce another IPI algorithm for estimating c_f using a lag-window estimator with Bartlett-Window weights.

5.3 Proposed forecasting approaches and their combinations

In the following, we propose different forecasting approaches based on the semiparametric regression model (5.1) and the well-known random walk model. For the latter three alternatives, i.e., random walks with a constant, a linear drift, and a local linear drift are considered. In particular, in line with Fildes and Stekler (2002) and Timmermann (2006), we propose the use of some combinations of those single forecasting approaches. We assess the practical performance of those single forecasting methods and combinations using a well-known criterion.

5.3.1 Point prediction based on the semiparametric model

Point prediction based on a semiparametric regression model is usually conducted in two steps. Consider first the data-driven local linear regression estimate $\hat{m}_{LL}(x_t)$ obtained following the algorithm described by Feng et al. (2018). Assume that we would like to carry out forecasts for k steps. A typical practice in the literature is to use a simple linear extrapolation of the fitted trend (see, e.g. Beran and Ocker, 1999). The point forecasts $\hat{m}_{LL}(x_{n+j})$ for forecasting horizons $j = 1, ..., k$ are then given by:

$$\hat{m}_{LL}(x_{n+j}) = \hat{m}_{LL}(x_n) + j * \Delta m, \tag{5.5}$$

where $\Delta m = \hat{m}_{LL}(x_n) - \hat{m}_{LL}(x_{n-1})$ denotes the estimated increment of the trend at the end of the time series.

Secondly, the approximately best linear prediction for the ARMA part is calculated. Under the assumptions of model (5.1) Z_t follows a stationary ARMA(p, q) model. Denote this model by $\phi(B)Z_t = \psi(B)\zeta_{1,t}$, where $\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p$ and $\psi(B) = \psi_1 B + ... + \psi_q B^q$ are the characteristic polynomials of the AR and MA parts, respectively, which have no common factor and all roots are outside the unit circle, and $\zeta_{1,t}$ are i.i.d. $N(0, \sigma_1^2)$ random variables. Brockwell and Davis (2002) propose an approximately best linear predictor based on the state-space representation of

Z_t via an algorithm developed by Harvey and McKenzie (1982). In this chapter we propose calculating the (approximately) best linear predictions $\hat{Z}_{n+j}$, $j = 1, ..., k$ based on the residuals $\hat{Z}_t = Y_t - \hat{m}(x_t)$. It is well known that the additional errors in those best linear predictions caused by the estimation errors in the trend function are asymptotically negligible (Feng and Zhou, 2015). The total point forecasts are then given by $\hat{Y}_{1,n+j} = \hat{Z}_{n+j} + \hat{m}_{LL}(x_{n+j})$.

5.3.2 Some random walk models

Since the time series considered in this chapter are not very long, which is typical for developing economies, the practical performance of many other approaches developed based on asymptotic theory may be limited. Furthermore, it is well-known that the forecasting quality can be clearly improved by combining different forecasting methods (see e.g. Stock and Watson, 2004). For this purpose, we describe two variants of the (parametric) random walk models and introduce a semiparametric extension of this model. The commonly used random walk model is the one with a constant drift (called RCS) defined by

$$Y_t = Y_{t-1} + c + \zeta_{2,t}, \tag{5.6}$$

where $\zeta_{2,t}$ are assumed to be i.i.d. $N(0, \sigma_2^2)$ random variables and a random walk with a linear drift (called RSL):

$$Y_t = Y_{t-1} + a + bx_t + \zeta_{3,t}, \tag{5.7}$$

where $\zeta_{3,t}$ are i.i.d. $N(0, \sigma_3^2)$ random variables. See e.g. models (2.1) and (2.2) in Dickey and Fuller (1979) with $\rho = 1$. Estimation and forecasting using both of the RCS and RSL models are straightforward. The point forecasts using the RCS are

$$\hat{Y}_{2,n+j} = Y_n + j\hat{c}, j = 1, ...k. \tag{5.8}$$

Those obtained by means of the RSL are

$$\hat{Y}_{3,n+j} = Y_n + j\hat{a} + \hat{b}\sum_{i=1}^{j}(n+i), j = 1, ..., k. \tag{5.9}$$

Following the SEMIFAR model (Beran and Feng, 2002b), the RSL can be further extended to

$$Y_t = Y_{t-1} + \delta(x_t) + \zeta_{4,t}, \tag{5.10}$$

where $\delta(x)$ is a nonparametric smooth drift function and it is again assumed that $\zeta_{4,t}$ are i.i.d. $N(0, \sigma_4^2)$ random variables. We propose to estimate $\delta(x)$ by the local linear regression. This model can be referred to as a random walk with a local linear drift (RLL). For practical purposes one can simply apply the data-driven algorithm in this chapter to the differences of the observed time series. Now, an adjusted algorithm for this data volume should be employed because a larger bandwidth for estimating the drift function is more preferable. Point predictions of the drift function $\hat{\delta}(x_{n+j}),\ j = 1, ..., k$ can again be obtained by linear extrapolation. This results in the following point predictions:

$$\hat{Y}_{4,n+j} = Y_n + \sum_{i=1}^{j} \hat{\delta}(x_{n+i}), j = 1, ..., k. \tag{5.11}$$

Although the three random walk-models are closely related, we see that forecasting quality can be improved clearly by combining them. The use of suitable combinations of forecasting methods is well studied in the literature (see, e.g., Section 11 of De Gooijer and Hyndman, 2006; Timmermann, 2006 and references therein). Although some sophisticated combining methods, such as the optimal combination of Bates and Granger (1969) exist in the literature, it is often found that the simple average of all candidates is more robust and preferable (Elliott, 2011; Qian et al., 2015). In this chapter we hence only consider the use of those simple combinations. Based on the above four single approaches we can define a total of 15 combinations including the individual methods. Table 5.1 shows the composition of the combinations and their corresponding denotations for the following performance evaluation. To assess the forecasting quality of those combinations over the different data examples considered, we propose using the mean absolute scaled error (MASE) introduced by Hyndman and Köhler (2006), which is defined by

$$MASE = mean(|q_t|), t = n + 1, ..., n + k,$$

Table 5.1 – Composition of Combinations

LLR	Local Linear Regression Method										
RCS	Random Walk with a Constant Drift										
RSL	Random Walk with a Linear Drift										
RLL	Random Walk with a Local Linear Drift										
Combination	$C2_1$	$C2_2$	$C2_3$	$C2_4$	$C2_5$	$C2_6$	$C3_1$	$C3_2$	$C3_3$	$C3_4$	C4
Combined Models	LLR, RCS	LLR, RSL	LLR, RLL	RCS, RSL	RCS, RLL	RSL, RLL	LLR, RCS, RSL	LLR, RCS, RLL	LLR, RSL, RLL	RCS, RSL, RLL	LLR, RCS, RSL, RLL

where $q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|}$ denote the scaled errors with forecast errors $e_t = Y_t - \hat{Y}_t$. The MASE eliminates many drawbacks of commonly used accuracy measures because it is scale-independent and can thus be used to compare forecasts across different data sets of different scales. It is also less sensitive to outliers and does not require the observations to be sufficiently far away from 0. Furthermore, as it is evaluated by a benchmark value of 1, it is straightforward to interpret. If the MASE < 1, the forecasting method of interest, on average, is better than a one-step naïve forecast (see Armstrong and Collopy, 1992; Fildes and Stekler, 2002; Hyndman and Köhler, 2006; Franses, 2016). The MASE is also an established tool for evaluating the practical performance of methods in the GDP (growth) forecasting literature (see, e.g., Šindelář, 2017; Gerunov, 2016). Based on the root mean square error (RMSE) and following Hyndman and Köhler (2006), we further calculated the root mean square scaled error (RMSSE), $RMSSE = \sqrt{mean}(\tilde{q}_t)$, for $\tilde{q}_t = \frac{e_t^2}{\frac{1}{n-1}\sum_{i=2}^{n}(Y_t - Y_{i-1})^2}$. Assessment by means of the RMSSE is quite similar to that using the MASE. Therefore, the detailed discussion of the RMSSE results given in table B5.1 is omitted in the next section.

5.3.3 Individual forecasts, prediction intervals and densities

Prediction intervals, more exact prediction densities, can help practitioners to understand the accuracy and limitations of point forecasts. Studies on these topics have become much more common over the past 35 years. A brief review on those topics is given in Section 12 of De Gooijer and Hyndman (2006) and Section 6 of Timmermann (2006). In the following we discuss this problem under the normal distribution assumption. It is assumed that all of the above mentioned four models hold at least approximately for the observed time series. For instance, if Y_t follows the RCS model, then the RSL and RLL approaches are both true. It can also be shown that the LLR approach is now a good approximation. Furthermore, it is assumed that the cross correlation between the errors of all models under consideration is stationary with the correlation matrix R. The empirical study in the next section shows that results obtained in this way work quite well in practice. A detailed theoretical study on those assumptions is beyond the aim of the current chapter.

Let the $MA(\infty)$ representation of Z_t be given by $Z_t = \sum_{i=0}^{\infty} \alpha_i \zeta_{1,t-i}$. Then an individual forecast following this model is given by $Y_{1,n+j} = \hat{Y}_{1,n+j} + \sum_{i=0}^{j-1} \alpha_i \zeta_{1,n+i}$ with the approximate variance $var(Y_{1,n+j}) = \sigma_1^2 \sum_{i=0}^{j-1} \alpha_i^2$. For simplicity, the error in $\hat{m}_{LL}(x_{n+j})$ of the order $O_p(n^{-2/5})$ is omitted. Individual forecasts under the three random walk models are given be $Y_{i,n+j} = \hat{Y}_{j,n+j} + \sum_{j-1}^{j} \zeta_{i,n+l}$ with the (approximate) variances $var(Y_{i,n+j}) = j\sigma_i^2$, $i = 2, 3$ and 4, and $j = 1, ..., k$, respectively. The combinations of forecasts mentioned above are also applied to those four individual forecasts. For the j-step prediction by a given combination, let $S_{comb,j}$ denote the vector of the standard errors of individual forecasts and R_{comb} be the matrix of the cross correlations between the errors in different models, which does not depend on the observation time. Then the (approximate) variance for the forecast of an individual observation is given by

$$var(Y_{comb,n+j}) = S^T_{comb,j} R_{comb,j} S_{comb,j}. \tag{5.12}$$

Assume that the point forecasts are asymptotically unbiased. We then determine the forecasting densities of a combination. Forecasting intervals for this combination at given confidence level can be easily calculated. It is well known that the variance of a combined forecasting method is usually smaller than that of a single forecast, in particular if some of the used approaches are independent or even negatively correlated to the others. We observe that this is true for the proposed approaches in this chapter. In this case, the bias can also be reduced through a suitable combination. These facts are discussed in detail in the next section using the selected data examples.

5.4 Application to the selected examples

For the empirical analysis and the evaluation of the validity of the methods proposed in sections 5.2 and 5.3, predictive models were developed based on a training set of the first 60 (58 for China) observations. The remaining five observations were used as the validity set to evaluate the model's performance (see e.g. Picard and Berk, 1990). The forecast horizon is set at five years. Based on the errors between the true values of the data and the estimated forecast values obtained by each of the proposed methods, table A5.1 shows the MASE values of all combinations for all considered data examples. In addition, the mean values and the standard deviations of the MASE for each combination are given. As explained previously, the RMSSE are calculated too yet do not yield differing information over and above that delivered compared to the MASE. Hence, they are omitted in the following discussion, but the values are given in table B5.1 to numerically support the overall findings on the MASE. The results following both criteria show that with all proposed methods and their combinations, a good forecast can be achieved on average, but that C4 clearly outperforms all other approaches. Each of the four proposed models thus helps to improve forecasting quality.

This is apparent when looking at the mean values and standard deviations, because in most cases both of them are clearly lower for the combination methods than for the single methods. LLR shows a mediocre performance, as the forecast is distorted by structural breaks at the end of the training set. Therefore, the forecasts for Brazil, South Africa, the USA and the UK are not good. Mexico is an exception, because after the break Log GDP increased again at a rate which met the estimated trend. For the previously mentioned examples, Log GDP did not increase as rapidly after the break, so the local linear trend for these countries was clearly overestimated. For the remaining examples with no structural break at the end of the training set, the forecasting performance is good. RLL faces similar problems. It performs better than the LLR for the log data for Brazil and India, but worse for the remaining examples and especially badly for the advanced economy examples. RSL can capture a regular movement in the data. In our examples, this is the case for India, the USA and the UK. The considered criteria are very small for these examples, whereas for the examples showing an unstable and/or non-linear movement, the forecasts with the RSL are worse and for some examples even >1. Nigeria is the exception here, where the linear drift captures the overall movement despite the instability. Due to the very good and very bad forecasting performance of RSL, the mean MASE is the smallest of all single-method mean MASE values. However, it is also the most unstable method of all with a standard deviation that is nearly as large as the mean value. RCS works worst for India and the USA, well for Mexico and well for all remaining examples. The results demonstrate that the forecast accuracy of the different models depends heavily on the country under consideration.

Combining these methods is a logical way to improve the forecasting accuracy and stability, as flaws of one methods can be compensated for by combining another. Also, as mentioned in the previous section, the variance can be reduced. Table 5.2 is strong indication of this intuition. It shows the variances, covariances and correlations for China and the USA on the left and right side of the table, respectively. The variances are shown in the diagonals, the covariances above the diagonals, and the correlations below the diagonals. The variance between the residuals obtained

Table 5.2 – Var, (diagonal), Covar (upper part) and Corr (lower part) of the residuals of China (left) and USA (right)

	$\hat{Y}^*_{RCS}$	$\hat{Y}^*_{RSL}$	$\hat{Y}^*_{RLL}$	$\hat{Y}^*_{LLR}$		$\hat{Y}^*_{RCS}$	$\hat{Y}^*_{RSL}$	$\hat{Y}^*_{RLL}$	$\hat{Y}^*_{LLR}$
$\hat{Y}^*_{RCS}$	0.4974	0.4678	0.4645	-0.1307		0.0494	0.0464	0.0436	-0.0180
$\hat{Y}^*_{RSL}$	0.9729	0.4648	0.4611	-0.1264		0.9615	0.0471	0.0444	-0.0174
$\hat{Y}^*_{RLL}$	0.9706	0.9966	0.4605	-0.1245		0.9418	0.9825	0.0433	-0.0160
$\hat{Y}^*_{LLR}$	-0.7862	-0.7865	-0.7785	0.0555		-0.5752	-0.5694	-0.5459	0.0198

by RCS is the largest and decreases for the residuals obtained by RSL and RLL. It is the smallest for the residuals obtained by LLR; they are negatively (sometimes even strongly negatively) correlated with the residuals obtained by the Random Walk models. The residuals obtained by RCS, RSL and RLL are all highly positively correlated, as expected. The effects are displayed in table A5.2, which shows the first values of the (approximate) standard deviation for the forecast of an individual observation, following (5.12) The results for other forecasting horizons are omitted to save space, because the order of those quantities does not depend on the length of the forecasting horizon. These values show only little differences for the random walk methods and the resulting combinations, but they decrease when LLR is used alone or added to RCS and/or RSL and/or RLL. Not surprisingly, the methods including LLR have the smallest standard deviation for calculating the forecasting intervals. Thus, the LLR endogenously estimates data-driven the local linear trend at each point in time, and in particular LLR takes care of boundary points, this method offers the best identification of the current trend and in combination systematically improves the forecast. Thus, this method provides good results for both developing and advanced economies. Developing economies are well known for non-log-linear growth processes and show both accelerating and decelerating phases. Therefore, the flexibility of the LLR method allows us to systematically adjust to these phases and hence improve the accuracy of trend identification. Consequently, the LLR contributes the most when the underlying growth process is clearly non-linear. This can be seen from Nigeria in figure A5.3, where different forecast starting points, beginning in 1996, with their respective prediction intervals demonstrate the adjustment process of C4. Obviously, the combination performs well during normal periods and adjusts quickly (after 2 years) to the new normal after massive changes due to its flexibility. This could be seen after 2000 where the IMF confirms a 1 billion Dollar loan for Nigeria that is followed by a boom in economic activities.

However, even if advanced economies are closer to a log-linear trend stationary growth process, growth episodes still differ. In advanced economies we also observe phases of more or less rapid growth. Thus, identifying such varying phases helps to improve trend forecasts for this country group as well. Figure A5.3 shows the same adjustment process for the UK and their prediction intervals. Evidently, a massive change in the growth process during the 2008 financial crisis occurs and, as demonstrated in table A5.1 and A5.2, every individual forecasting method cannot depict this distortion even after 2010. Notwithstanding, C4 is able to capture those movements after 2 years and hence also improves forecast accuracy of advanced economies.

Following the previous discussion of the results, we propose the combination including all four methods, i.e. C4, as the best approach. Figure A5.1 shows the point forecasts together with the 95% prediction intervals obtained by this combination for all countries. As a comparison, the same results for the second-best method, $C3_2$, are displayed in figure A5.2. In general, the results for both methods work very well, as the figures show. Almost all true observations lie within the forecasting intervals and are very close to the point forecasts. The one clear difference is South Africa, where some validity observations lie outside the forecasting intervals for $C3_2$, but not for C4 (see Figure A5.1e and A5.2e).

5.5 Conclusion

This chapter introduces the local linear forecasting approach for macroeconomic variables with a data-driven IPI algorithm for bandwidth selection. The random walk with drift is combined with the local linear trend estimation in order to improve forecast accuracy. This innovative forecast approach is applied to six arbitrarily selected developing and two advanced economies. It is shown that the best forecasts for all countries are obtained by combining our suggested endogenously determined local linear regression (LLR) with RCS, RSL and RLL (C4). Furthermore, the forecasts using the proposed local linear method are most stable. Thus, the local linear estimation and the random walk with a local linear trend are able to detect variations in the trend process that are typical for developing countries. Therefore, using this LLR method more precisely identifies current trend processes in the underlying data and in turn, significantly improves forecasts for advanced and developing economies.

However, there is potential for further methodological research. It would be interesting to carry out detailed theoretical studies on the relationship between the different single forecasting methods, the search of possible optimal combinations of those methods, the inclusion of further candidates of single forecasting methods, and a simulation study to confirm the theoretical and practical performance of the proposals in the current chapter.

A.6 Appendix of Chapter 5

Table A5.1 – MASE of all methods for all countries with mean and standard deviations

	Brazil	India	Mexico	China	S. Africa	Nigeria	USA	UK	mean	s.e.
LLR	0.803	0.546	0.439	0.378	1.519	0.495	0.940	1.122	0.780	0.397
RCS	0.670	1.486	0.369	0.790	0.547	0.768	0.997	0.673	0.787	0.336
RSL	1.512	0.235	1.625	0.434	0.834	0.216	0.059	0.206	0.640	0.619
RLL	0.534	0.449	1.307	0.709	1.019	0.825	1.603	1.939	1.048	0.529
$C2_1$	0.530	0.609	0.394	0.206	1.022	0.631	0.968	0.897	0.657	0.288
$C2_2$	1.158	0.382	0.650	0.406	0.436	0.355	0.471	0.656	0.564	0.266
$C2_3$	0.659	0.498	0.491	0.543	1.258	0.282	0.331	0.420	0.560	0.306
$C2_4$	0.576	0.838	0.669	0.178	0.183	0.492	0.496	0.431	0.483	0.225
$C2_5$	0.494	0.652	0.509	0.068	0.783	0.204	0.303	0.633	0.456	0.244
$C2_6$	0.936	0.333	1.466	0.571	0.186	0.382	0.803	0.875	0.694	0.414
$C3_1$	0.651	0.469	0.337	0.052	0.444	0.493	0.642	0.661	0.469	0.205
$C3_2$	0.532	0.346	0.233	0.113	1.021	0.156	0.147	0.223	0.346	0.304
$C3_3$	0.892	0.405	0.869	0.507	0.585	0.207	0.222	0.269	0.495	0.273
$C3_4$	0.562	0.498	0.881	0.133	0.274	0.155	0.203	0.359	0.383	0.254
$C4$	0.618	0.307	0.580	0.187	0.566	0.171	0.121	0.219	0.346	0.208

Table A5.2 – First (approximate) standard deviations for calculating the forecasting intervals

	Brazil	India	Mexico	China	S. Africa	Nigeria	USA	UK	mean	s.e.
LLR	0.026	0.016	0.026	0.024	0.014	0.049	0.014	0.012	0.03	0.016
RCS	0.038	0.031	0.037	0.071	0.024	0.065	0.022	0.021	0.065	0.036
RSL	0.035	0.029	0.034	0.068	0.022	0.065	0.022	0.021	0.062	0.035
RLL	0.032	0.029	0.033	0.068	0.019	0.063	0.021	0.020	0.060	0.035
$C2_1$	0.021	0.011	0.020	0.027	0.013	0.039	0.009	0.009	0.031	0.018
$C2_2$	0.019	0.010	0.019	0.026	0.012	0.039	0.009	0.009	0.029	0.018
$C2_3$	0.018	0.010	0.018	0.026	0.011	0.039	0.009	0.009	0.029	0.018
$C2_4$	0.036	0.030	0.035	0.069	0.022	0.065	0.022	0.021	0.063	0.035
$C2_5$	0.034	0.029	0.034	0.069	0.021	0.064	0.021	0.021	0.061	0.035
$C2_6$	0.033	0.029	0.033	0.068	0.020	0.064	0.021	0.021	0.060	0.035
$C3_1$	0.024	0.016	0.023	0.040	0.015	0.045	0.012	0.012	0.040	0.023
$C3_2$	0.023	0.016	0.022	0.040	0.014	0.044	0.012	0.012	0.039	0.023
$C3_3$	0.022	0.016	0.022	0.039	0.014	0.044	0.012	0.012	0.038	0.023
$C3_4$	0.034	0.029	0.034	0.068	0.021	0.064	0.021	0.021	0.061	0.035
$C4$	0.025	0.019	0.024	0.047	0.016	0.049	0.014	0.014	0.044	0.026

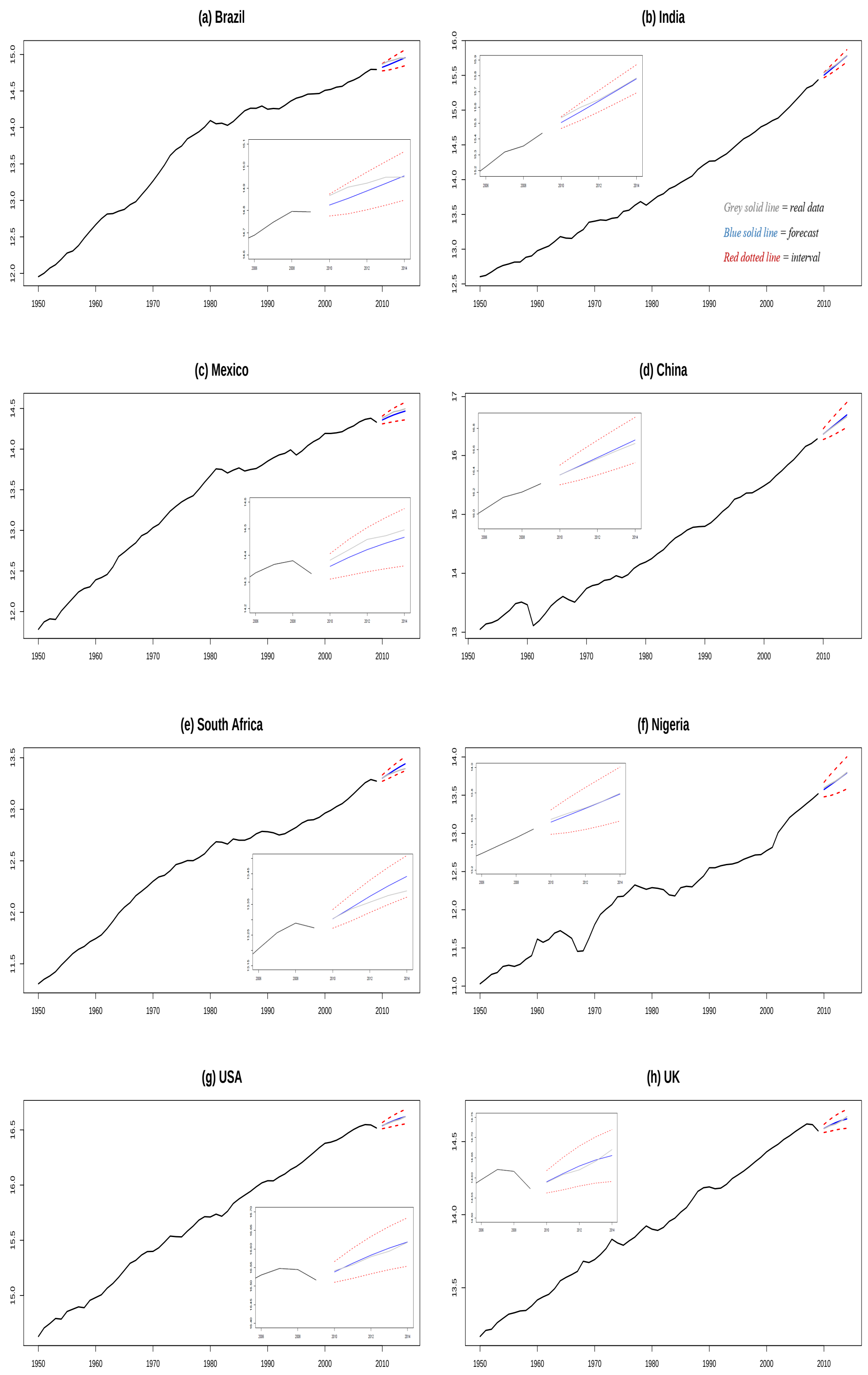

Figure A5.1 – Point and 95% Interval Forecast with C4 for all countries: green dotted = data, blue solid = forecast, red=interval

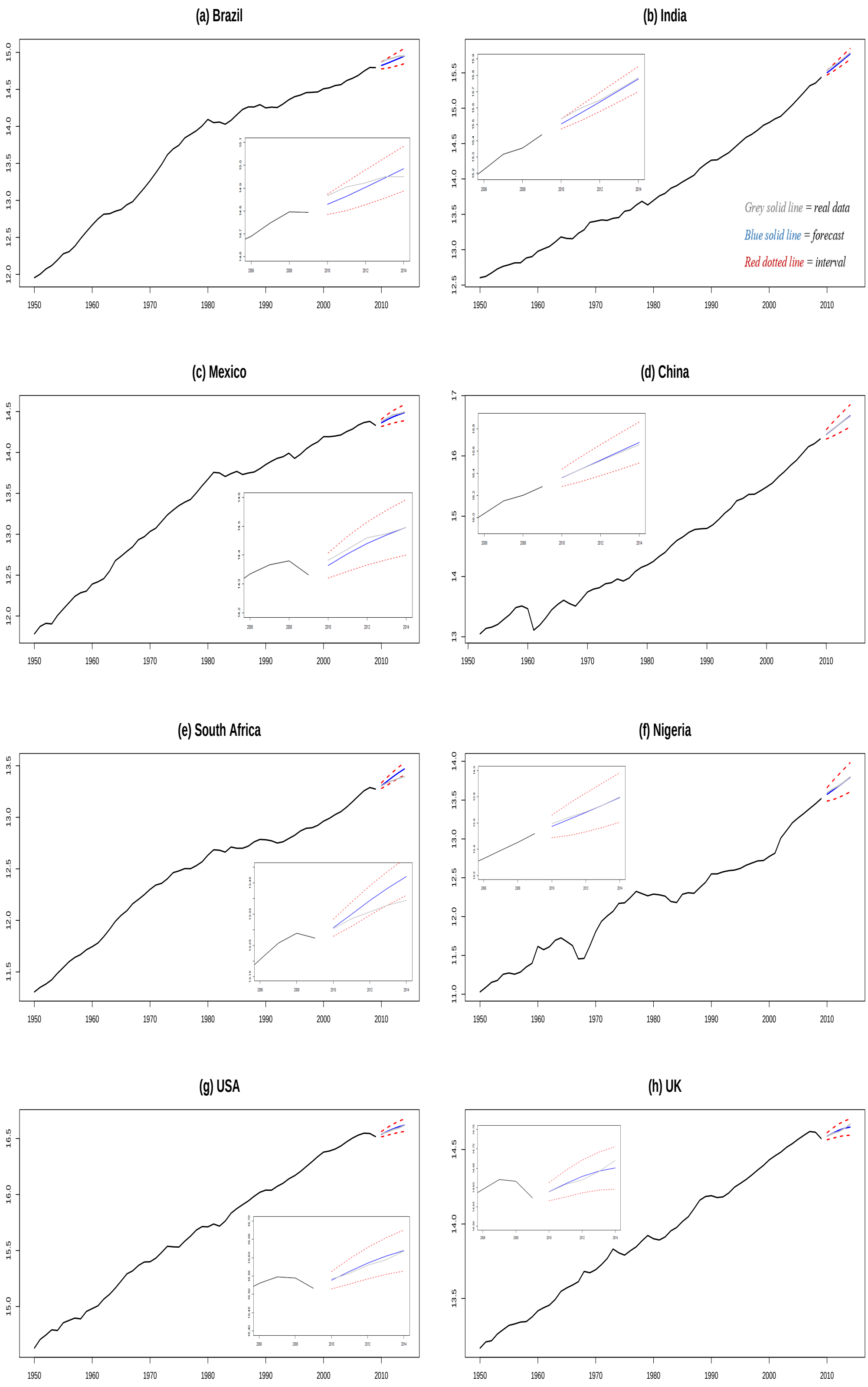

Figure A5.2 – Point and 95% Interval Forecast with $C3_2$ for all countries

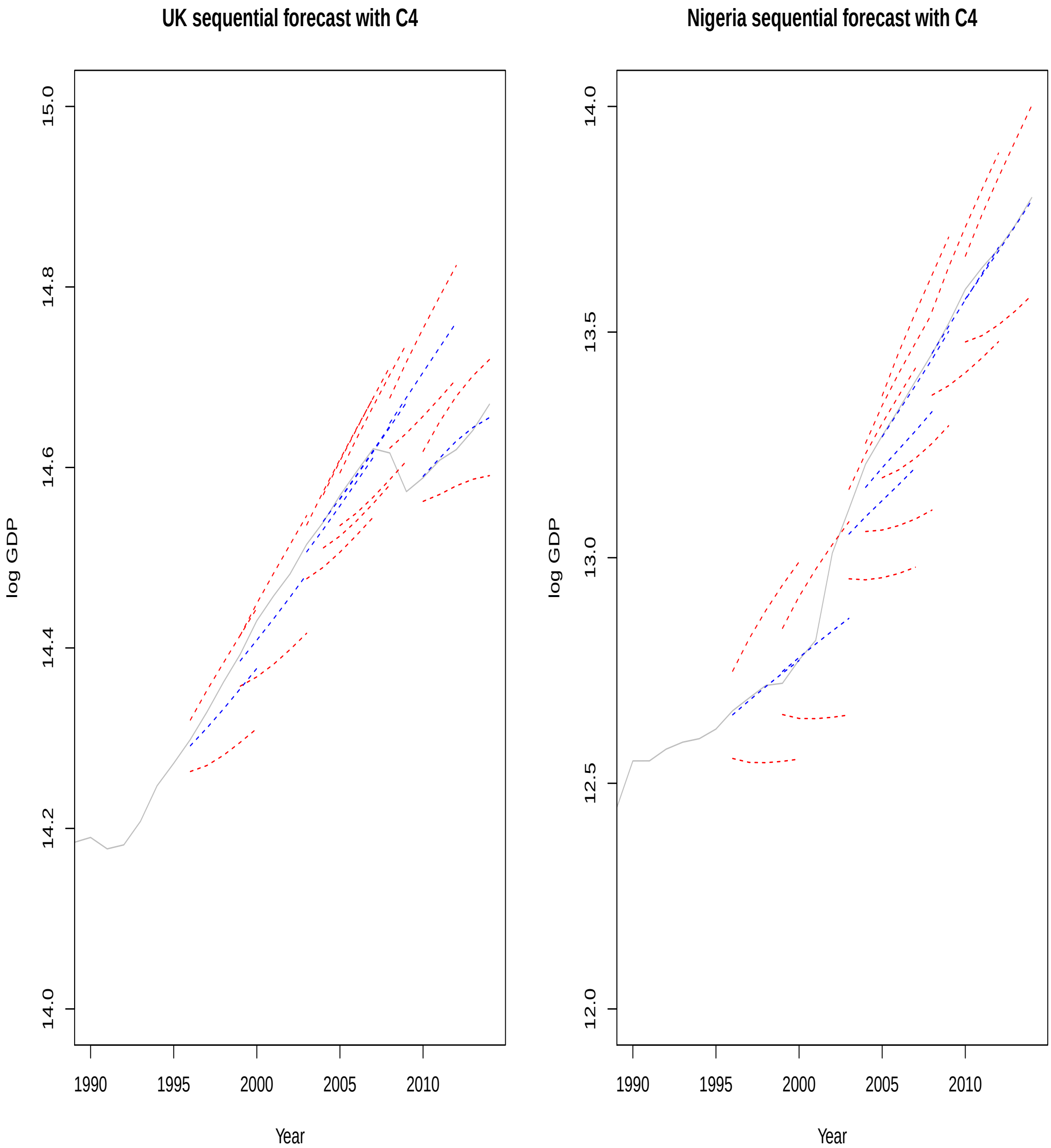

Figure A5.3 – UK and Nigeria point and 95% Interval Forecast with C4 for different starting values

6 Future Research Questions

The results of the analyses carried out in the previous chapters of this work show that the proposed methods overall work reliably well. Also it was shown that the introduced methods are applicable to a variety of data types in a financial and a non-financial setting. Transforming the data to obtain a model, which satisfies the targeted distribution assumption or developing distribution-free procedures are common methods, when distribution assumptions do not hold (see Graybill, 1976 p.213 in Sakia, 1992). It is therefore reasonable to pursue possible further improvement of these methods. In the following first ideas on possible future research are described. They are not mutually exclusive and the quality of a combination of proposals should be further examined after successful implementation.

A promising idea presented in section 6.1 is the application of the methods described in chapter 3 in a GARCH model framework. Estimating the time-varying scale from log-transformed returns with local polynomial regression is evaluated in a simulation study and compared to the Semi-GARCH modelling approach of Feng (2004a) via real financial data examples. The proposal given in section 6.2 is on additionally transforming the log-data to further normalize it. By doing so (log-) normal distribution assumption based methods are applicable without the possible problem of misspecification. The results of chapter 4, in particular, suggest that the model with a conditional normal distribution assumption performs the best for forecasting non-negative financial data. However, the SW tests carried out in advance also show that the normal distribution assumption is rejected for the log-data. First numerical results are in clear favour of this new transformation proposal. For the forecast of the GDP of developing countries section 6.3 extends the empirical analysis of chapter 5 by neural network methods. Section 6.4 briefly summarizes further ideas without first empirical results. Section 6.4.1 proposes the inclusion of a local bandwidth factor into the IPI. Estimating functions are suggested in section 6.4.2 as an alternative to (Q)MLE and section 6.4.3 suggests improving the forecast of non-negative financial data by a block bootstrap.

6.1 A Semi-Log-GARCH model extension

The often emphasized increased flexibility of the methods presented in the previous chapters suggests their application to other types of financial data or another model framework. Especially improving the estimation of the volatility of financial returns with GARCH models should be of interest with Basel IV replacing Basel III (BCBS, 2017). The idea of fitting GARCH models semiparametrically is already described in the literature, as mentioned in chapter 1. An approach closely related to the ones described throughout this work is the one of Feng (2004a) on simultaneously modelling the conditional heteroskedasticity and scale change with a Semi-GARCH model. It uses a kernel estimator of the scale function. Analogously to the reasoning behind the development of the Semi-Log-ACD model, the Semi-Log-GARCH model has some technical advantages over the Semi-GARCH and GARCH models. The Semi-GARCH model improves the GARCH model by accounting for a possibly time-varying conditional variance. Nelson (1991) introduces the Exponential GARCH (EGARCH) model, which is more flexible than the GARCH model, because it does not require non-negativity constraints on the model parameters. Same holds for the Log-GARCH model described by Geweke (1986), Pantula (1986) and Milhøj (1987). The approach taken here combines the constraint free and the nonparametric scale functions ideas and extends the semiparametric methods of Feng (2004a) to the local polynomial regression methods described in chapter 3. The proposed model is called Semi-Log-GARCH model, but the choice of GARCH model to be fitted after standardization is not limited to the Log-GARCH model. Peitz (2016) applies a Semi-EGARCH model using kernel regression and compares the actual deviations of the Value at Risk obtained by different models to the theoretical ones. He finds the Semi-EGARCH model to perform best, so it is expected that the results further improve, when local polynomial instead of kernel regression is used.

6.1.1 Description of the Semi-Log-GARCH model

Following Bollerslev (1986) let r_t^* for $t = 1, ..., T$ denote the log-returns of a price series, h_t their conditional variance and ε_t i.i.d. $N(0,1)$ random variables. Let the GARCH model be defined by

$$r_t^* = \sqrt{h_t}\varepsilon_t, \qquad r_t|F_{t-1}\ N(0, h_t) \tag{6.1}$$

and

$$h_t = \omega \sum_{j=1}^{p} \alpha_j r_{t-j}^{*2} + \sum_{k=1}^{q} \beta_k h_{t-k}. \tag{6.2}$$

To ensure non-negativity of returns, restrictions $\omega > 0$ and $\alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q \geq 0$ are imposed on the parameters. It is further assumed that $\sum_{j=1}^{p} \alpha_j + \sum_{k=1}^{q} \beta_k < 1$ to ensure that a unique strictly stationary solution of the GARCH model exists (see Feng, 2004a). The parameter constraints can be dropped for the EGARCH model of Nelson (1991) described by (6.1) and:

$$\log(h_t) = \omega + \sum_{j=1}^{p}(\alpha_j \varepsilon_{t-j} + \gamma(|\varepsilon_t| - E(\varepsilon_t)) + \sum_{k=1}^{q} \beta_k \log(h_{t-k}). \tag{6.3}$$

The size effect of shocks on h_t is expressed by $\gamma(|\varepsilon_t| - E(\varepsilon_t))$ here. The Log-GARCH model is described by Geweke (1986), Pantula (1986) and Milhøj (1987) and compared to the EGARCH model by Francq et al. (2013). Amongst other things, they find that concerning estimation, the Log-GARCH model is more tractable than the EGARCH model. The Log-GARCH model is given by (6.1) and

$$\log(h_t) = \omega + \sum_{i=1}^{p} \alpha_i \log(r_{t-i}^{*2}) + \sum_{j=1}^{q} \log(h_{t-j}). \tag{6.4}$$

For the Semi-GARCH model of Feng (2004a) a time-varying smooth scale function $\sigma(\tau_t)$ is introduced multiplicatively into (6.1), where $\tau_t = t/T$ denotes the rescaled time:

$$r_t = \sqrt{h_t} \varepsilon_t \sigma(\tau_t). \tag{6.5}$$

The Semi-GARCH model is defined by (6.5) and (6.2) and aims to account for a possible time-varying unconditional variance. Let $\nu(\tau) = \sigma^2(\tau)$ denote the local variance of a time series model $Y_t - \mu = \sigma(\tau_t) r_t^*$, the model can be rewritten as a nonparametric regression model. It is given by

$$(Y_t - \mu)^2 = \nu(\tau_t) + \nu(\tau_t)\zeta_t, \tag{6.6}$$

where ζ_t are zero mean stationary time series errors. The semiparametric estimation procedure following Feng (2004a) is the kernel estimation of $\nu(\tau)$ in a first step. The model parameters are estimated from the standardized residuals $\hat{r}_t = (y_t - \bar{y}_t)/\hat{\sigma}(\tau_t)$ in a second step. For the Semi-Log-GARCH model, the scale function is estimated from $log(r_t) = R_t$ and a stationary additive model with a trend

$$(Y_t - \mu)^2 = \nu(\tau_t) + R_t \tag{6.7}$$

is obtained. The scale $\nu(\tau)$ does not enter the nonparametric regression model twice now, i.e. there is no more heteroskedasticity in the model to account for. The model fitting procedure is similar to the one of the Semi-Log-ACD model. The scale is estimated from the log-transformed returns, where local polynomial regression methods apply. In a second step, the scale is removed from the transformed data, any parametric GARCH model of choice is fitted to the residuals and the eventually combined estimated scale and model estimates are retransformed accordingly.

6.1.2 First empirical results

The generality of the IPI described in chapter 3 allows it to be applied directly to estimate the scale function of log-transformed return series. To get a first impression of the performance of the estimators a rudimentary simulation study is carried out in the following. A trend is simulated based on the square root of the average of two squared scale functions estimated from a Semi-Log-GARCH model for SAP data from January 1997 to April 2018. The daily data was obtained from Yahoo Finance and not processed further. One scale function is estimated via local linear and the other one with a local cubic regression. For both estimations the variance factor of the IPI is estimated nonparametrically and a bandwidth correction factor as proposed by Feng and Heiler (2009) is applied to increase the bandwidth for estimating the variance. For both estimations, the EIM is applied with a naive inflation factor, i.e. $\upsilon = 5/9$ and starting bandwidth $b_0 = 0.075$ for the local linear regression and $\upsilon = 9/13$ and $b_0 = 0.15$ for the local cubic. The combination of trends to be used for the simulation study is standardized to obtain the standardized SAP returns, to which a GARCH(1,1) model is fitted. The parameter estimates $\omega = 0.1$, $\alpha = 0.13$ and $\beta = 0.77$ are used to simulate a return series with 5365 observations by the *garchSpec* function of the *fGarch* package of R. The simulated returns and the simulated trend function are combined multiplicatively and Semi-Log-GARCH models are fitted to the log-transformed absolute centralized returns.
A local linear and a local cubic regression, each with a nonparametric variance factor estimation, a bandwidth correction factor and the EIM with a naive inflation factor are applied. Additionally, we apply the same methods, but with optimal inflation factors, i.e. $\upsilon = 5/7$ and $\upsilon = 9/11$ for the local linear and the third order local polynomial regression, respectively. The simulation is carried out with 1000 replications. Figure A6.1 shows the data simulated without trend, the data simulated with trend and the simulated scale function, together with its local linear and third order local polynomial estimations of the first replication. A visual assessment of the performance of the trend estimation shows that all methods applied capture the trend well and that the regressions each perform similarly for different inflation factors. Compared to the data simulated without a trend, the returns standardized by either method capture the return dynamics well and appear stationary without

lacking prominent features of the simulated stationary data.
The numerical evaluation of the Semi-Log-GARCH model performance is done via a discussion on the bandwidth selection, the scale function estimation and the model parameter estimation. The quality of the bandwidth selection is discussed with the means, standard deviations and mean squared errors of the finally selected bandwidths of each method in each replication. The basis of comparison is a numerical derivation of the true optimal bandwidth, following Beran et al. (2009) by bandwidths h_{ASE}. Latter are obtained by minimizing the ASE of the simulated trend and the local polynomial trend estimates with i different bandwidths. For each of the j replications of the simulation the $\text{ASE}_{i,j}$ and corresponding $h_{\text{ASE},j}$ are calculated. The optimal bandwidth is obtained by minimizing the means of ASE_i. Let $\hat{h}_{M,LL}$ and $\hat{h}_{M,LC}$ denote the approximated optimal bandwidths for the local linear and the local cubic regression. Here, $\hat{h}_{M,LL} = 0.095$ and $\hat{h}_{M,LC} = 0.2$. Let LL_A and LL_B denote the local linear estimations with $\upsilon = 5/7$ and $5/9$, respectively and accordingly LC_A and LC_B the local cubic estimations with $\upsilon = 9/11$ and $\upsilon = 9/13$. The finally selected bandwidths of each replication are compared against these values and the results are given in table 6.1 and figure 6.1.

Table 6.1 – Selected bandwidth statistics for the simulation study

	$E(\hat{b}_A)$	$sd(\hat{b}_A)$	$\text{MSE}(\hat{h}_M)^o$
LL_A	0.0933	0.0062	0.0413
LL_B	0.0973	0.0048	0.0288
LC_A	0.1800	0.0256	1.0574
LC_B	0.1983	0.0171	0.2943

^{o} The MSE values obtained were multiplied by 1000.

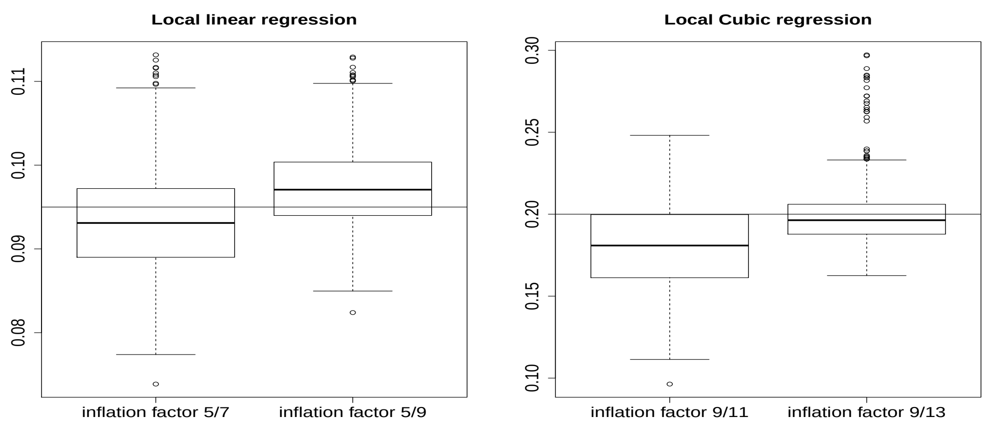

Figure 6.1 – Semi-Log-GARCH simulation box-plots of bandwidth selection

Overall the bandwidth selection works well and the means of the finally selected

bandwidths of each replication with either method do not differ much from the numerically derived optimal bandwidths. Also the MSE are small. The methods using the naive inflation factors 5/9 and 9/13 have smaller MSE and standard deviations than the ones with optimal inflation factors. The distributions of the bandwidth estimation are displayed graphically in figure 6.1 in the form of box-plots and support previous findings. The bandwidth selection of either method applied works well and the bias and variance decrease for the methods using naive inflation factors, compared to the ones using optimal inflation factors.
To evaluate the performance of the scale function estimation a new criterion is proposed here to assess the gain of estimating the scale function as compared to fitting a standard GARCH model. It is calculated to obtain the empirical efficiency of the Semi-Log-GARCH model. Let this criterion be denoted by Empirical reduction of squared errors (RSE). It can be calculated by means of the MSE or the RMSE. Under the assumption that the scale function is not time-varying but constant,

$$MSE_C = E[(G_t - E(G_t))^2],$$

where G_t denotes the simulated trend and $RMSE_C = \sqrt{E[(G_t - E(G_t))^2]}$. For each replication of the simulation, j, the MSE is calculated for the scale function estimated by the four different methods, $M = LL_A, LL_B, LC_A, LC_B$.

$$MSE_{j,M} = E[(G_t - \hat{G}_{t,j,M})^2]$$

and correspondingly $RMSE_{j,M} = \sqrt{E[(G_t - \hat{G}_{t,j,M})^2]}$. To determine the degree of improvement by accounting for a time-varying scale function, instead of assuming it to be constant the RSE is calculated as

$$RSE_{j,M} = \left(1 - \frac{(R)MSE_{j,M}}{(R)MSE_C}\right) \cdot 100\%. \tag{6.8}$$

The quality of the parameter estimation is assessed with the relative efficiencies described in chapter 2. Table 6.2 shows the means of each criterion over the 1000 replications. Concerning the quality of the scale function estimation, the RSE allows to interpret the results with regards to a GARCH model, for which no trend was estimated and eliminated before fitting. When the RSE is calculated based on MSE it can be seen that for all methods about 97% of the squared estimation error can be reduced with the nonparametric estimation of the scale function. When the RSE is RMSE based it is about 83%. The local cubic regression with a naive inflation factor performs best concerning both RSE criteria and the relative efficiencies of all parameters. Furthermore, these results all show that the assumption of a constant

Table 6.2 – Estimation quality criteria and finally selected bandwidths of 1000 replications in a simulation study

	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	RSE_{MSE}	RSE_{RMSE}
LL, $\upsilon = 5/7$	67.080	101.780	89.340	96.923	83.023
LL, $\upsilon = 5/9$	71.490	101.960	91.770	96.937	83.080
LC, $\upsilon = 9/11$	66.900	101.660	88.600	97.021	83.325
LC, $\upsilon = 9/13$	74.250	101.600	92.820	97.234	83.947
param. GARCH	3.050	33.100	4.860	*0.012*	*1.078*

Remark: The italic values are the mean values of $(R)MSE_C \cdot 100$.

scale function is a clear misspecification. The results on the relative efficiencies of the parameter estimates with the Semi-Log-GARCH model are similar to the ones obtained in chapter 2. The estimation quality of the scale parameter ω is the worst, the highest relative efficiencies are for the short-term dependence parameter estimation of α, and the ones for $\hat{\beta}$ are also high. In general, the naive inflation factors gives better results than the optimal ones. The behaviour of the relative efficiencies should be evaluated further for an increased sample size and for different simulated data cases. Of course, a general more in-depth analysis of more results is necessary to evaluate these methods meaningfully. In particular, it is crucial to include at least the Semi-GARCH model approach to compare results. The Semi-GARCH algorithm is not yet available in a form that allows it to be compared to the one of the Semi-Log-GARCH model in a simulation study. However, the results obtained so far promote further research on this topic in a timely manner. Fitting a GARCH model parametrically is a clear misspecification and the computational methods developed for the Semi-Log-ACD model work well for the Semi-Log-GARCH model.

The methods applied in the simulation study are also used for two real financial data examples in the following. The data of ALV and BMW from January 1997 to April 2018 is used, which was retrieved from Yahoo Finance. The results are compared graphically to the ones obtained by the Semi-GARCH model with an IPI described in (30) of Feng (2004a). Figure A6.2 shows the price and return series for ALV and BMW, a graphical comparison of the scale function estimations, as well as the returns standardized by the two shown local polynomial estimations. The figures of the returns show that the unconditional variance is not constant but changing over time. The comparisons of the estimated scale functions show that the local linear regression seems to underfit the data and the kernel regression seems to overfit it. The local cubic regression seems to estimate the trend of the series well. The performance at the boundaries would need special consideration, though. The standardized returns seem to be stationary after standardization with either method. However, as stated before, the computational implementation of the Semi-GARCH model is not

yet equal to the Semi-Log-GARCH model in options and a direct comparison at this point is not just.
The Semi-Log-GARCH model can be interesting especially in the field of quantitative risk management. It is more flexible than the Semi-GARCH model, because after standardization, any GARCH model of choice can be fitted. Also, an algorithm developed for the Semi-Log-ACD model can be applied directly to the log-transformed returns and hence the technical implementation is already available. The simulated and real financial examples show, that the Semi-Log-GARCH model performs well. A comparison study similar to the one of Peitz (2016) could deliver deeper insight into the performance of this model proposal compared to other models used in this field of research. Already well-established GARCH models should be discussed as reference models nonetheless. A further discussion of this model should be a valuable contribution to the research on GARCH models and quantitative risk management.

6.2 The Log-sinh-arcsinh-transformation

As stated before, the results of chapter 4 suggest that the method with a normal distribution assumption performs the best, even though this assumption is rejected by a SW normality test. Figure 6.2 shows the histograms of the log-residuals for the Mean Durations (MD), Realized Volatility (RV) and Volume (Vol) of ALV, together with their densities and normal density curves. The corresponding QQ normal plots are given in the second column. Figure A6.3 in the appendix shows the histograms and QQ normal plots for the log-data. The densities look about normal for MD and Vol, but the QQ normal plots indicate heavy-tails. Figure 6.2 shows that the normal distribution assumption for the detrended log-data is not completely absurd, but cannot be supported either. The curves are all skewed, leptokurtic and the QQ normal plots indicate heavy tails of varying degrees. The approach presented in this section is different to the methods described in the previous chapters of this work. Instead of developing new or improving existing methods that do not require a conditional normal distribution, it aims to further refine normality of the data through a transformation. If this idea proves successful it can improve the accuracy of results obtained by methods that assume a conditional normal distribution.

6.2.1 Description of the Log-SAS-transformation

The idea of further normalizing the data already exists in the literature. Hyperbolic power transformations are appropriate in transforming skewed distributions to nor-

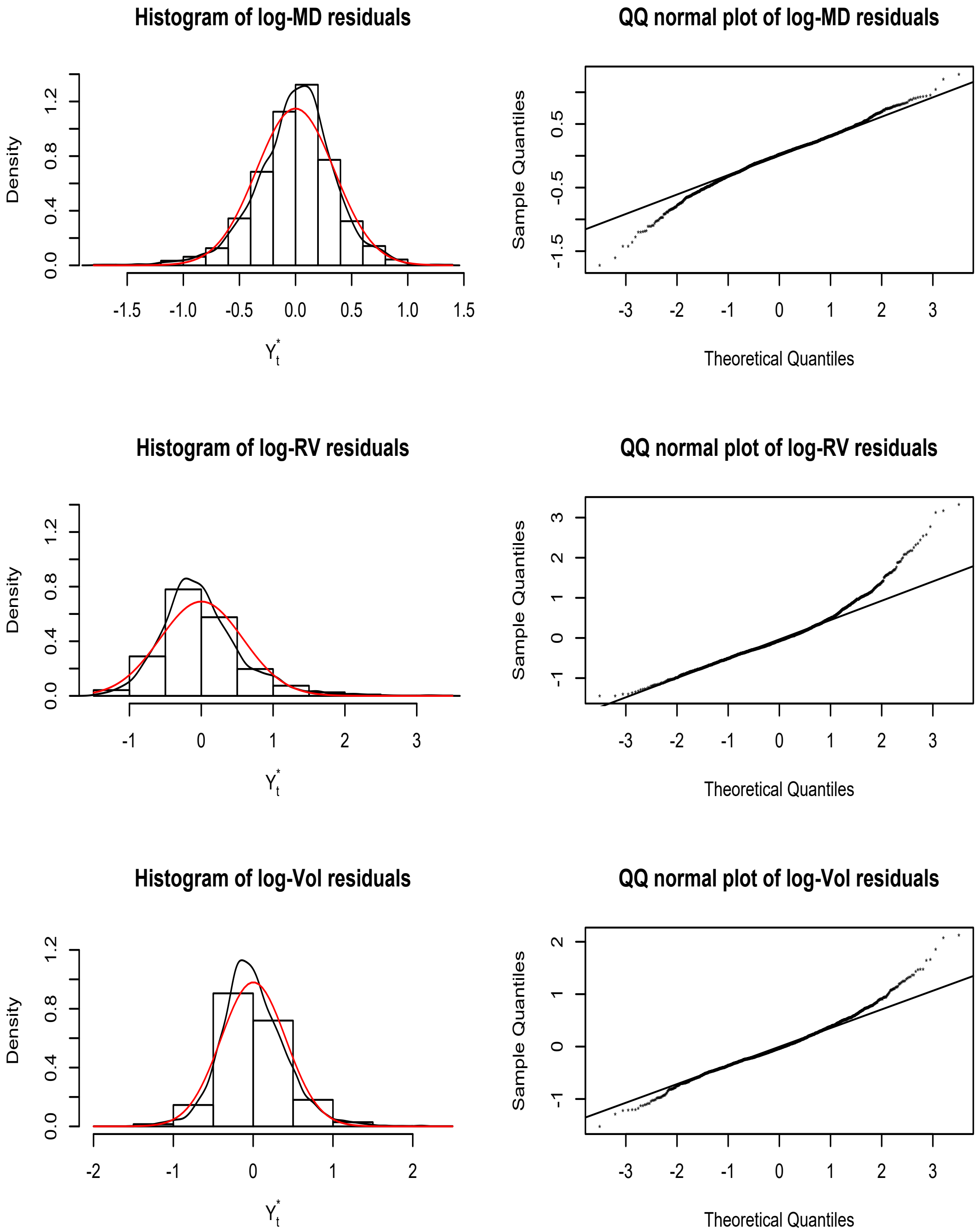

Figure 6.2 – Histograms and QQ normal plots of ALV log-residuals

mal. Since log-data is considered, the use of e.g. the Box-Cox transformation (Box and Cox, 1964) is ruled out, because it is only defined on the positive real line. Also modifications of this transformation, such as the shifted power transformation or the modulus transformation of John and Draper (1980) do not apply to $(-\infty, \infty)$. A promising transformation is the sinh-arcsinh (SAS) transformation, proposed by Jones and Pewsey (2009) and generalized by Feng (2018). It applies to the whole range, allows for symmetric and skewed densities and also for heavy and light-tail distributions. Furthermore, many properties are explored for the resulting family of distributions and moments of any order exist (see Rubio et al., 2016). Following Feng (2018), the sinh-arcsinh transformation is applied to the log-data for improving and further refining the normality of it. This idea is not yet discussed in the ACD model literature. Let it be denoted by log-sinh-arcsinh (log-SAS-) transformation and the model used to obtain first results later in this section is called Log-SAS-ACD model. Let $Y_t = \log(X_t)$, as defined in (3.2). The log-SAS-transformation is defined by assuming that Y_t has a SAS-marginal distribution for:

$$Z_t = S_{\varepsilon,\delta}(Y_{t,\varepsilon,\delta}) = \sinh[\varepsilon + \delta \sinh^{-1}(Y_{t,\varepsilon,\delta})]. \tag{6.9}$$

The parameters $\varepsilon \in \mathbb{R}$ and $\delta > 0$ in (6.9) control for the skewness and tailweight, respectively. $Y_{t,\varepsilon,\delta}$ is a random variable associated with the normal density $f_{\varepsilon,\delta}$, which is given by:

$$f_{\varepsilon,\delta}(y) = \frac{1}{\sqrt{2\pi}} \frac{\delta C_{\varepsilon,\delta}(y)}{\sqrt{1+y^2}} exp\left(-\frac{1}{2} S^2_{\varepsilon,\delta}(y)\right), \tag{6.10}$$

where $C_{\varepsilon,\delta}(y) = \cosh(\varepsilon + \delta \sinh^{-1}(y))$. The transformation in (6.9) is designed for the canonical case, where the location and scale parameters, σ and μ, are removed. Using $\sigma^{-1} f_{\varepsilon,\delta}(\frac{y-\mu}{\sigma})$ allows to reinstate those parameters, yielding a four-parameter distribution. Adjusting the values of those four parameters in the transformation of the data adjusts the location, scale, skewness and kurtosis of the density. In the following, log-data is analysed with and without the SAS-transformation. The parameters for skewness, ε, and kurtosis, δ, are adjusted and the ones for location and scale are not considered. Let $\hat{Y}_t^* = Y_t - \hat{\mu}_t$ denote the detrended log-data of interest. The standardized data is given by $\mathring{Y}_t^* = \frac{\hat{Y}_t^* - \bar{\hat{Y}}_t^*}{sd(\hat{Y}_t^*)}$. The transformation in (6.9) is applied to $\mathring{Y}_t^*$ via an algorithm that finds the combination of ε and δ which minimizes the Jarque-Bera (JB, see Jarque and Bera, 1980) test-statistic or maximizes the p-value of the SW test. Let the data obtained by either algorithm be denoted by $\tilde{Y}_{t,\text{JB}}^*$ and $\tilde{Y}_{t,\text{SW}}^*$.

6.2.2 First empirical results

To obtain first results, MD, RV and Vol of ALV are log-SAS-transformed. The combinations of parameters to be used for the SAS-transformation are $\varepsilon_{\min} = -0.4$ to $\varepsilon_{\max} = 1.4$ and $\delta_{\min} = 0.01$ to $\delta_{\max} = 2$, each with a grid of 0.01. The trend is estimated via a local linear regression, with the EIM inflation method, inflation factor $\upsilon = 5/7$ and a bi-square kernel. For estimating the variance factor, the nonparametric method is used with a starting bandwidth $b_0 = 0.10$, significance level $\alpha = 10\%$ and a bi-square kernel (methods described in chapter 3). The left column of figure 6.3 shows the histograms of $\tilde{Y}^*_{t,\mathrm{SW}}$, together with the density and normal density curve and the right column shows the histograms of $\tilde{Y}^*_{t,\mathrm{JB}}$, together with the density and normal density curve. QQ normal plots of all data are shown in figure 6.4. The figures clearly show the improvement towards normality after the log-data has been SAS-transformed with parameters selected by either algorithm. Skewness and tails are corrected towards normality, which is especially apparent for the RV data example, and also the tails are clearly lightened. The SW p-values and JB log-statistics, given under each histogram, show that for the log-SAS-transformed data the Null is clearly not rejected in each case for either method. Even though the JB test is based on a function of the sample skewness and kurtosis and the SW test uses two variance estimators, they select similar values for δ and ε. However, this is just a finding for the data considered here. A more in-depth discussion of the parameter selection of both methods naturally requires further data examples and possibly an examination of further goodness-of-fit tests additional to theoretically derived properties.
Generally, an in-depth theoretical discussion of this transformation needs to be included in the intended future research. In particular, the effect of the choice of δ on the tailweights needs proper investigation. Jones and Pewsey (2009) state that for $\delta < 1$ the tails are heavier than for the normal distribution and lighter if $\delta > 1$. Based on the tail behaviour formula given in (5) of Jones and Pewsey (2009), we propose to only consider $\delta > 0.5$:

$$f_{\varepsilon,\delta}(|y|) \approx exp(-sgn(y)\varepsilon)|y|^{\delta-1} exp(-e^{-sgn(y)2\varepsilon}|y|^{2\delta}). \tag{6.11}$$

If $\delta = 1$ the tails are as the ones of the normal distribution. The tails are heavier (lighter) than the ones of the normal distribution for $\delta < 1 (\delta > 1)$. If $\delta < 0.5$ the moments of the original data do not exist after retransformation. Here, $0.5 < \delta < 1$ was selected via both algorithms for all data examples, even though $\delta < 0.5$ was a possible choice given by the range of values for the algorithm. For manually fixing δ, $\delta < 0.5$ is suggested to be excluded and $\delta = 0.5$ is not recommended either as the results might be instable. The performance of the algorithm with respect to the choice of δ needs to be studied further to gain more insight, just as a discussion of the properties of the SAS distribution and transformation. To highlight the importance of an adequate parameter selection for the log-SAS-transformation, the contour plots are given in figure 6.5 to show how the values of the test statistics of SW and JB change as a function of both parameters. Please note that the statistics were logarithmized for display reasons. Here it can easily be seen, that the statistics quickly take on values that ultimately lead to a rejection of the Null for slight changes in the parameters. Also the range of parameter values that leads to the JB statistic minimizing or SW p-value maximizing combination is very small. This is illustrated by figure A6.4, which shows close-ups of the cores of the contour plots. The excerpts also show, that the grid chosen between the different starting and end values for the parameters can further refine the combinations. Therefore, the parameter selection and the development of adequate methods should be of central practical importance. The idea of the log-SAS-transformation requires a much deeper theoretical discussion and a practical assessment that goes beyond the examples shown here. Nevertheless, the results obtained here are very promising and advocate an extensive study of it in the near future.

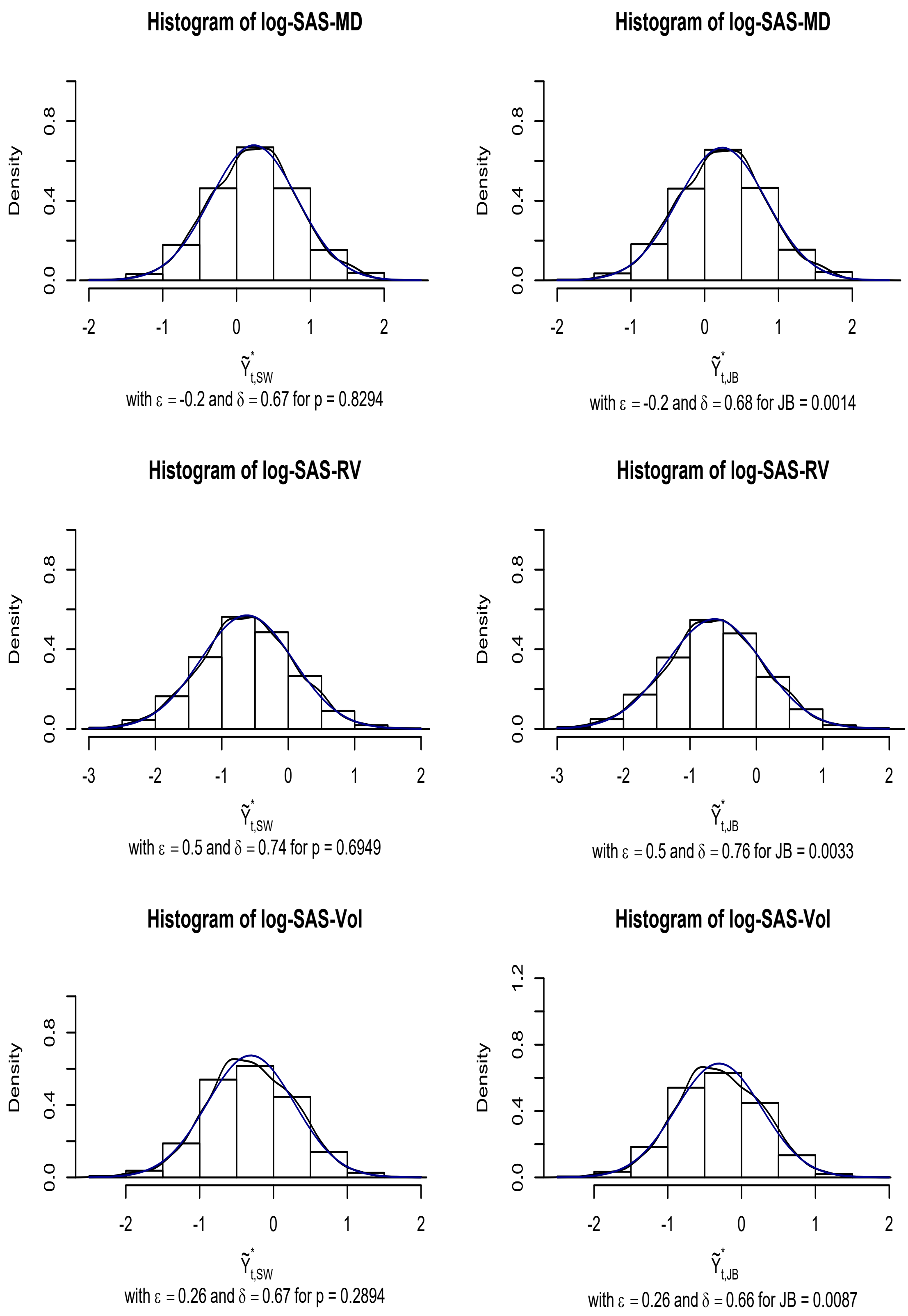

Figure 6.3 – Histograms of log-SAS-transformed data

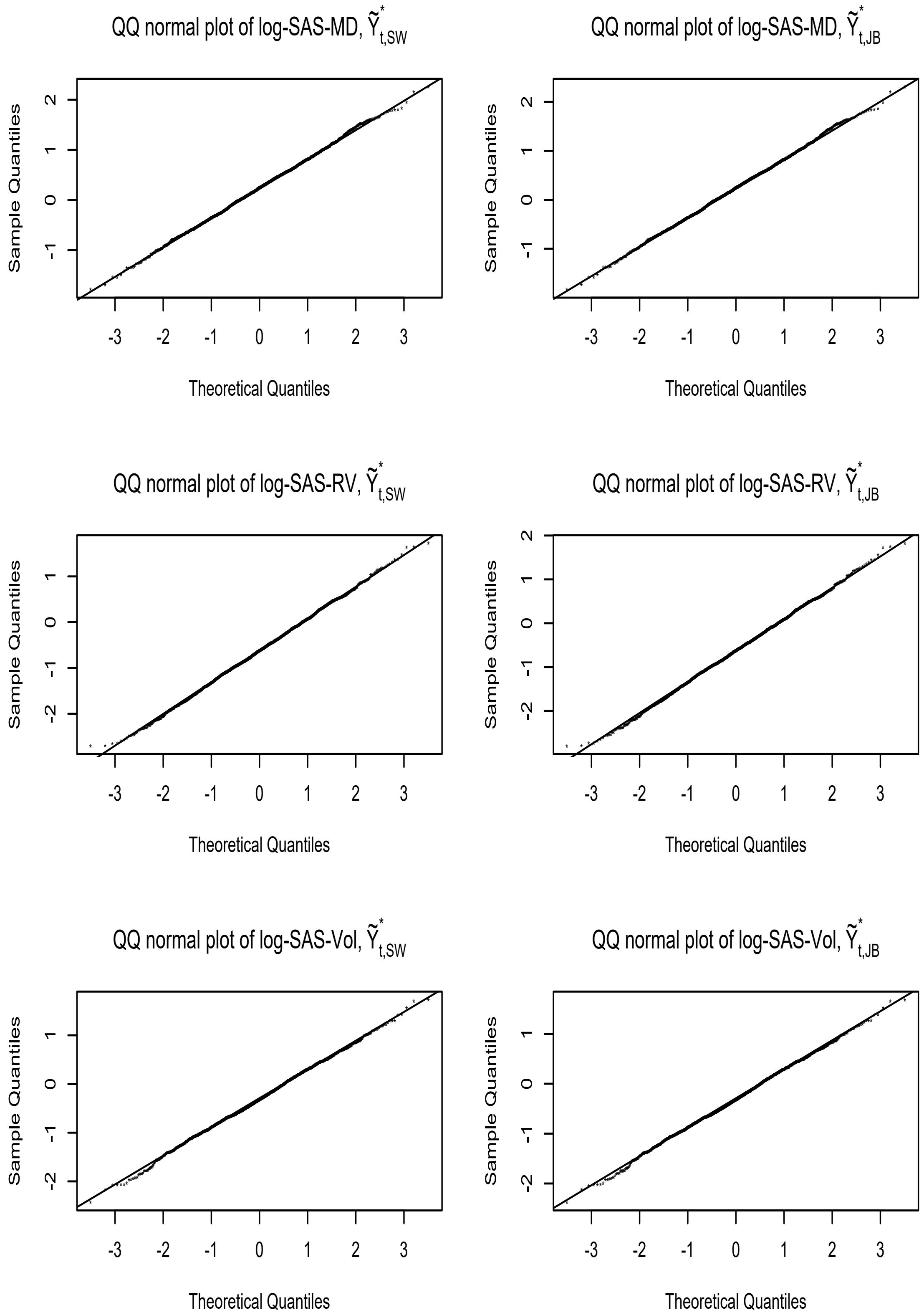

Figure 6.4 – QQ-Normal-Plots of log-SAS-transformed data

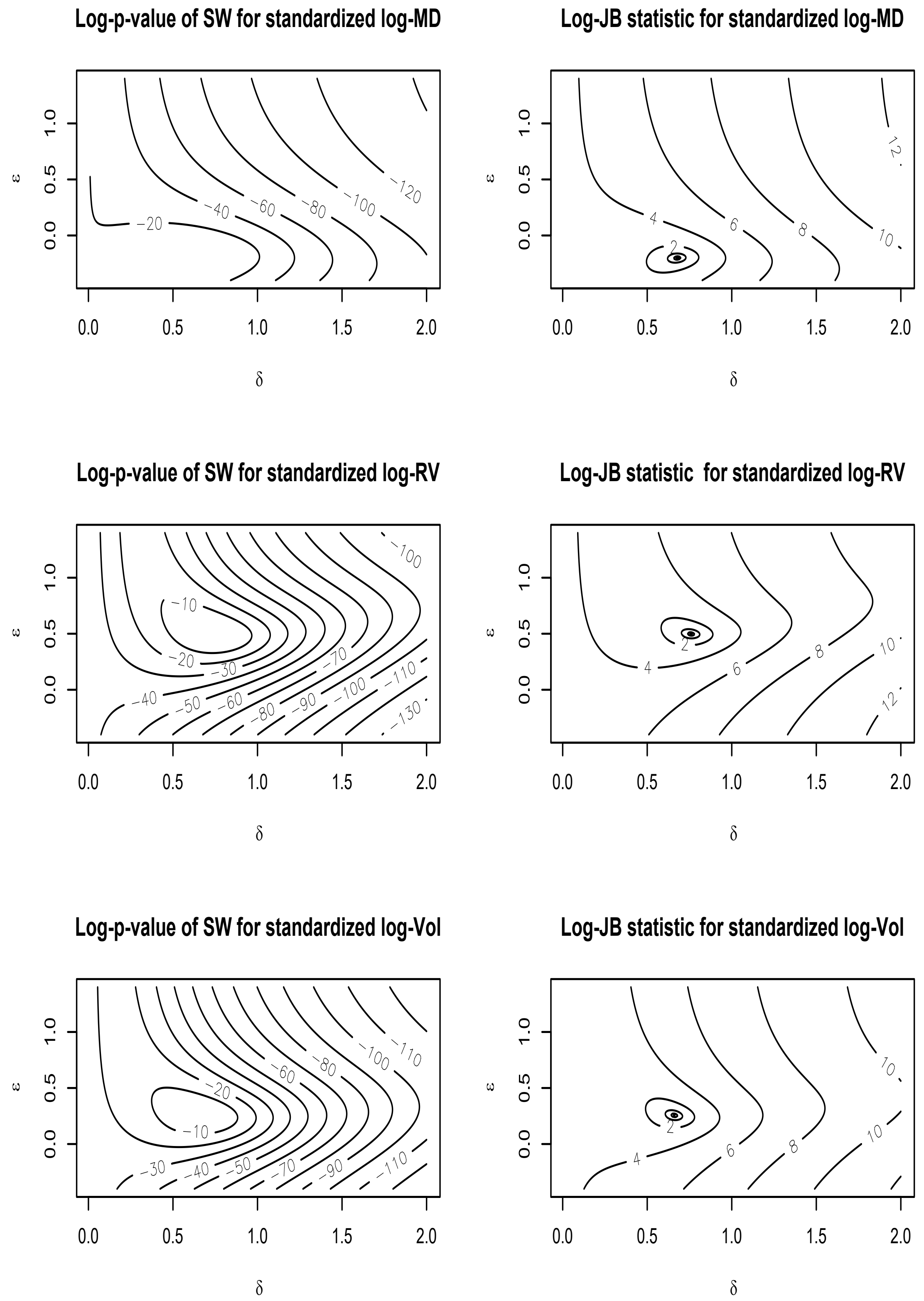

Figure 6.5 – Contour plots of goodness-of-fit test statistics with different δ and ε

6.3 Neural Network GDP forecasts

Fritz et al. (2018) show that the combination of different forecasting methods and especially the inclusion of local linear methods delivers good forecasts for the log-GDP data of developing and advanced countries. Nevertheless, the search for further single forecasting methods is considered interesting to possibly improve the forecast quality. For this purpose, Neural Network (NN) methods are described briefly in the following and applied to the data examples used in Fritz et al. (2018). Also it is examined, if the NN method proposed can itself be improved by local linear regression. NN models are able to capture complex non-linear relationships and, amongst other features, strict assumptions on the error terms are not required (White, 1989; Sena and Nagwani, 2016). Examining whether local linear regression methods can improve its performance should, therefore, be interesting, as well. Admittedly, there are different results and opinions in the literature regarding the performance of NN methods, despite their rapid growth in prominence (Crone et al., 2011 and references therein). Crone et al. (2011) extend the M3 competition of Makridakis and Hibon (2000) towards NN and computational intelligence methods. For the submissions they find, amongst others, that it is no longer generally true that simple methods outperform complicated ones. Even though they cannot identify one main finding or “best practices” (Crone et al., 2011, p. 657), they still find the results on the NN and computational intelligence forecasting encouraging for future research.
For the analysis intended here, autoregressive Neural Network models of orders p and k are considered. The order p for the NNAR model denotes the number of lagged inputs used and k the number of nodes in the hidden layer (see Hyndman and Athanasopoulos, 2014). Dismissing the hidden layer would yield an AR(p) model. However, unlike the AR model, the NNAR model does not require parameter restrictions to ensure stationarity. Following Dietz (2012) the NNAR is defined by:

$$X_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{k} \psi \left(\gamma_{0j} + \sum_{i=1}^{p} \gamma_{ij} X_{t-i} \right) \beta_j + \epsilon_t, \qquad (6.12)$$

where X_t denotes the value of a time series at a point in time $t = 1, ..., n$ and ω, α and β are the NNAR model parameters. The first part of (6.12) denotes the autoregressive part, i.e. i weighted lagged values are used as input. The activation function $\psi(\bullet)$ consists of weight parameters γ_{ij}, for which j describes the number of nodes in the hidden layer and a bias, $\gamma_{0,j}$. Before entering the output layer, each node in the hidden layer is weighted by β_j and ϵ_t are i.i.d. random variables with zero mean and constant variance.
The NN methods are applied in the empirical application framework of Fritz et al.

(2018) to show the effect of including them into combinations examined previously. Since the literature disagrees on whether to remove the trend from the data before applying NN methods or not (see e.g. Zhang and Kline, 2007; Ahmed et al, 2010 or Shmueli and Lichtendahl, 2016) they are applied with and without trend here. The forecasting methods described in Fritz et al. (2018) are supplemented by the NN method without trend removal (NNAR) and with trend removal by local linear regression (NNARLL). The forecasting intervals cannot be derived analytically for the NN methods, however, future sample paths of the model can be simulated iteratively to give a good idea of the forecast distribution (Hyndman, 2017). Including the two NN proposals gives 63 single and combination methods. Table A6.1 gives an overview over the methods and their denotation. Table A6.2 gives the mean MASE, as well as the mean SMAE, the mean length of the forecasting intervals and the mean EoPoFi for the 5-year forecasts of the log-GDP data of the countries considered in Fritz et al. (2018). The combination of LLR, RCS, RSL and RLL with a mean MASE of 0.346 is identified best in the previous log-GDP forecast analysis, so this is the benchmark value to which the results of this section are compared against. The results of the MASE in table A6.2 indicate that none of the new single or combination methods are better than the combination identified in Fritz et al. (2018). Except for the single RWLL and NNAR the forecasts are still better than the one-step naïve, though, but $C2_{11}$ and $C2_{13}$, which both contain NNAR are close to 1. First mentioned combination is worsened, as compared to the MASE of the single method RWL, whereas second mentioned combination is improved as compared to the MASE of the single method RWLL. The effect of including NN methods is overall inconclusive. The values of each criterion for each country are given graphically in figures A6.5 to A6.8. Figure A6.5 shows, that combining NN with a local linear regression improves the forecasts in most of the examined cases. Also, the MASE decreases and becomes more stable, the more methods are combined. Overall the superiority of a method over the one-step naïve seems to be the result of combinations in general. A positive or negative influence of a particular method on the forecasts cannot be detected. The SMAE given in figure A6.6 does not indicate anything different than the MASE concerning the analysis sought here. The more methods are combined, the smaller and less fluctuant is the SMAE. The mean lengths of the 95% forecasting intervals given in figure A6.7 show that they increase, if LLR is not included in the combination of methods. Also the forecast intervals are narrower for LLR and NNARLL as single methods than for the other single methods. For discussing the EoPoFi via figure A6.8 it must be noted first, that the theoretical deviation is 0.25 here and cannot be met by actual deviations. For most examples the number of actual deviations from either the upper or the lower bound is 0 or in less cases > 0, the more methods

are combined. The EoPoFi for China and Nigeria are the smallest overall, but their mean lengths of forecasting intervals are the largest, which explains why there are no deviations from the interval bounds. For $C4_1$ the forecasting intervals are not too large and no observations of the validity set lie outside the bounds.
The inclusion of NNAR and NNARLL methods for forecasting GDP growth of developing countries did not give conclusive results. The criteria show, that the forecasts are in most cases more stable the more methods are combined. A method that consistently improves the forecast quality if added to another is not identified, though. On average, the forecasts of $C4_1$ are still the best. Overall, the idea of including further prediction methods in more combinations is not invalidated by these results, just the choice of NN methods did not give clear indication of improvement. Also whether the trend should be removed via local liner regression before applying NN methods is without clear result. Following the results of Crone et al (2011), other NN or computational intelligence approaches may be used in future research on this topic.

6.4 Miscellaneous research topics

Three further ideas on future research are briefly described in the following. The suggestions made derive from findings obtained in previous chapters of this work and are supported by results of literature on the here discussed or similar topics. For the following ideas no numerical results are presented.

6.4.1 Local bandwidth factor for IPI improvement

The estimation of the scale function with a global asymptotically optimal bandwidth yields good results for the modelling of intraday data, daily average data, as well as the forecasting of such. The results of the RASE and their discussion in section 2.4.3 give numerical evidence of the good practical performance of the proposed trend estimation procedure. The results of applying the more automatized IPI to real financial non-negative data in each of the main chapters support this finding. In order to further optimize the introduced algorithm and to also prepare for its possible application to data with a more complex underlying deterministic structure, a local bandwidth selection approach could be included in the IPI (see e.g. Fan et al., 1996). Fan and Gijbels (1992) propose a local linear smoother with a variable bandwidth $b_n/\alpha(X_j)$, where $\alpha(\bullet)$ is a non-negative function and obtained via an IPI algorithm. Fan and Gijbels (1995) let the bandwidth vary with location via assessment of the conditional MSE. Brockmann et al. (1993) extend the global IPI (Gasser et al., 1991) by local steps for a local bandwidth estimator. Herrmann (1997) adds an additional final step to obtain a local plug-in bandwidth to the Brockmann et al. (1993) IPI and Staniswalis (1989) proposes a two-step local bandwidth estimation procedure based on the finite sample expression for the MSE. In many of these proposals, the local asymptotical bandwidth, $b_A(\tau)$, is time-varying and can be obtained by minimizing the dominant part of the MSE of $\hat{m}(t)$ (see Brockmann et al., 1993). In accordance with the denotation used for the formula of the asymptotical optimal bandwidth given in (2.10), let

$$b_A(\tau) = \left(\frac{R(K)2S}{I^2(K)} \frac{\mu^2(t)}{([\mu]'')^2} \right)^{1/5} N^{-1/5}. \tag{6.13}$$

The selection of a local bandwidth at each point of the observation would be very computer-intensive for the (U)HF data considered, though. A local bandwidth factor could be suggested as a solution. A global factor to generally increase or decrease would not serve the issue well. Either it would worsen the estimation at the boundary points, which are found to be estimated well already or it would not solve the problem of estimating more complex trends, as the bandwidth would still be a global one. For kernel density estimation, the use of a local bandwidth factor to obtain a variable bandwidth is proposed by Abramson (1982), Demir and Toktamiş (2010) or Aljuhani and Al Turk (2014). The idea to pursue in the local linear regression context is based on the proposals of Feng (unpublished) and Feng and Beran (unpublished) on a robust estimation of the volatility trend in semiparametric GARCH models. Let the local bandwidth selector be defined as $b_A(\tau) = C(\tau)b_A$, where b_A is the asymptotically global optimal bandwidth as given for example in (2.10). The local bandwidth factor is given by

$$C(\tau) = \left(\frac{\mu^2(\tau)}{I(\mu^2(\tau))} \frac{I([\mu(\tau)'']^2)}{[\mu(\tau)'']^2} \right)^{1/5}, \tag{6.14}$$

with $I(\mu^2(\tau)) = \int \mu^2(\tau)d\tau$. It is obtained by

a) Obtain $[\check{\mu}(\tau)]''$, the estimate of $[\mu(\tau)]''$ with $\hat{b}_d = b_A^{5/9}$.

b) Obtain $[\bar{\mu}(\tau)]''$, a smoother of $[\check{\mu}(\tau)]''$, with $\hat{b}_A$.

c) Calculate $C(\tau)$ by inserting $\hat{\mu^2}(\tau)$ and $[\bar{\mu}(\tau)]''$ into (6.14).

This idea could first be tested on whether it improves the trend estimation in the simulation carried out in chapter 2 or the results of real financial non-negative data examples used throughout this work. If the results are clearly improved compared to the ones of chapter 2, it could also be of interest to extend this idea to a local polynomial regression for log-data.

6.4.2 Model parameter estimation

Throughout the course of this work, the focus was not on estimating the model parameters, but rather on the steps before and after. The methods for estimating the scale function and for predicting future values of the data were aimed to be improved. The discussion of relative efficiencies in section 2.4.4 showed that the ML methods implemented in the R-packages used work sufficiently well. In order to further improve the semiparametric fitting of the ACD models discussed, estimating the model parameters semiparametrically could be an approach to consider for future research. In particular, using the estimating function (EF) of Godambe (1960) instead of (Q)ML or GMM estimators could be a viable and valuable method. The EF is suggested by the research literature, especially when the true underlying density is unknown (see e.g. Hansen, 1982, Grammig and Wellner, 2002, Hallin and La Vecchia, 2017 or see Bera and Bilias, 2002 and Bera et al., 2006 for a detailed description of the development of the EF and a summary of other estimation approaches). Improving the accuracy of model parameter estimation adds to the overall performance of the proposed semiparametric ACD models.

EF have already been show to work well for the estimation of ACD model parameters. Allen et al. (2013a, b) develop the EF estimation procedure for the Weibull ACD model and Exponential and Box-Cox ACD Model, respectively. They each design a simulation study to compare EF and QML methods. Another comparison of EF and ML methods for ACD models is done e.g. by Pathmanathan et al. (2010), who extend the work of Peiris et al. (2007) by comparing EF and ML estimation methods for ACD model parameters with different error distributions. Ng and Peiris (2013) use a simulation study to evaluate an EF approach based on a Generalized Gamma distribution and Ng et al. (2014) compare a semiparametric parameter estimation method for Log-ACD models based on the theory of EF with corresponding (Q)MLE via a simulation study, as well. The results of applying EF methods to ACD models can be summarized as follows: EF and ML methods both deliver comparable results, but EF is faster and more efficient, if the true distribution is unknown. Thus, the EF is superior to (Q)ML regarding practical evaluation and application without prior knowledge on the error distribution.

The inclusion of an EF based parameter estimation method into the semiparametric ACD models would be new to the literature. If it is proven to perform better than the (Q)MLE it should be an important contribution to the research on Semi-ACD models. Following the already existing literature, a simulation study should be the most useful in discussing and evaluating the performance of the EF model parameter estimation.

6.4.3 Block-bootstrap for forecasting non-negative financial data

Concerning the forecast it is apparent from the results obtained in chapter 4, that the semiparametric bootstrap methods performed mostly well, however not as good as the semiparametric Log-ACD model with the conditional normal distribution assumption. Besides improving the normality of the data an alternative approach could be the improvement of the distribution assumption free bootstrap method. The bootstrap applied in chapter 4 is model-based and in particular based on autoregressive processes with i.i.d. innovations. Even though the forecasting results do not suggest this assumption to be apparently false, the data could be autocorrelated, which may lead to a poor(er) performance of the applied model based bootstrap (Singh, 1981). A possible solution to suggest is the use of block bootstrap to guarantee stationarity for short-range dependent data ('stationary bootstrap', Politis and Romano, 1994), and to follow the works of Paparoditis and Politits (2002), Dowla et al. (2003) and Dowla et al. (2013) on a modification of the block bootstrap for locally stationary processes. Here, the blocks are resampled in a way that blocks can only be replaced by other blocks, if their starting points are close to each other. The advantage of applying the block bootstrap idea is that it can also be applied to non i.i.d. data and no model needs to be specified to obtain the residuals. The crucial point of the block bootstrap, though, is the choice of block sizes. When applying the block bootstrap for locally stationary data, the choice of an additional parameter is required, namely the parameter which determines what is considered 'close to each other' (see Gonçalves and Politis, 2011). The block-length selection is discussed in the literature. Bühlmann and Künsch (1999), for example, propose a fully data-driven block-length selection procedure via a linearised extension (Hampel et al., 1986) of the Brockmann et al. (1993) IPI to spectral estimation (Bühlmann, 1996) and to use the inverse of the selected bandwidth as the optimal block-length. Politis and White (2004)[12] also give an automatic block-selection method based on spectral estimation via flat-top lag windows. Another selection procedure is required to determine the optimal block size. Whether already existing procedures should be applied or the development of a new procedure is required would have to be investigated after the theoretical discussion of block-bootstrap properties. An empirical application would not only have to determine whether the bootstrap forecasts are improved but more importantly whether the forecasts are better than the ones of the Semi-Log-ACD model under a conditional normal distribution.

[12]Patton et al. (2009) give a correction on their 2004 article, as Nordmann (2009) found an error in the calculations of Lahiri (1999), an article the 2004 idea was substantially based on. The original idea is unchanged, though, and the results after the corrections are even better.

A.5 Appendix of Chapter 6

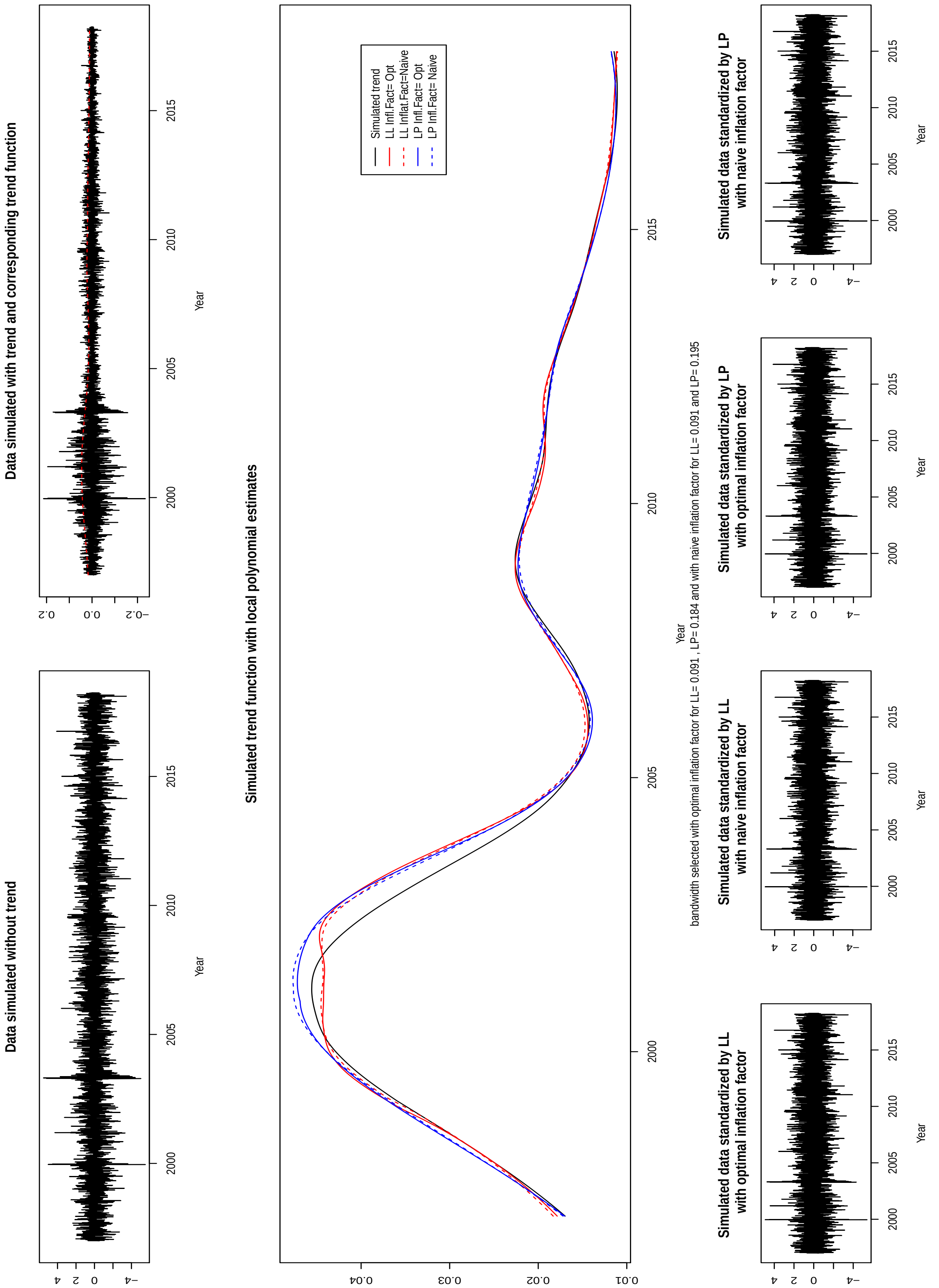

Figure A6.1 – Semi-(Log-)GARCH model first replication simulation

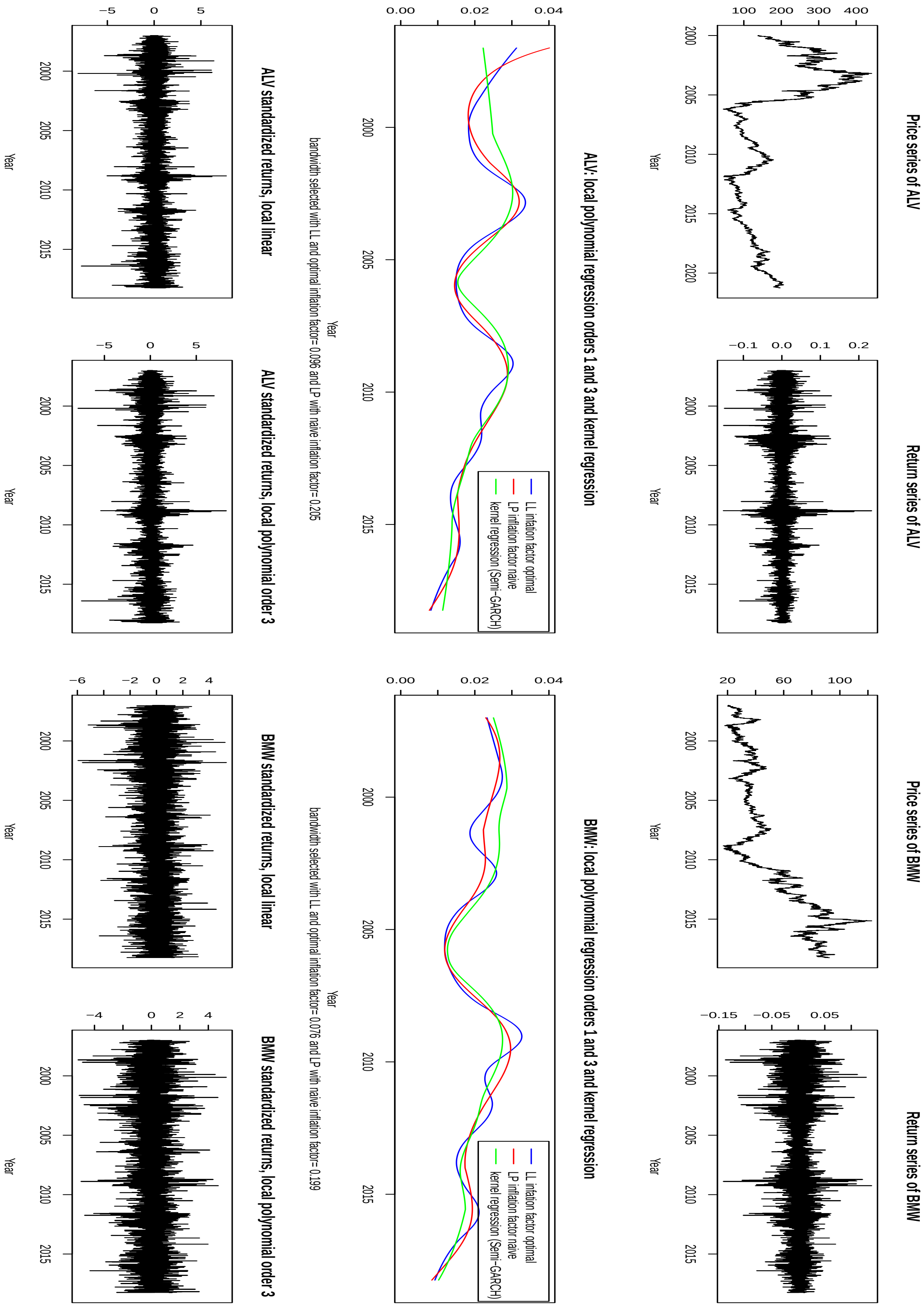

Figure A6.2 – Semi-(Log-)GARCH model fitting to returns of ALV and BMW

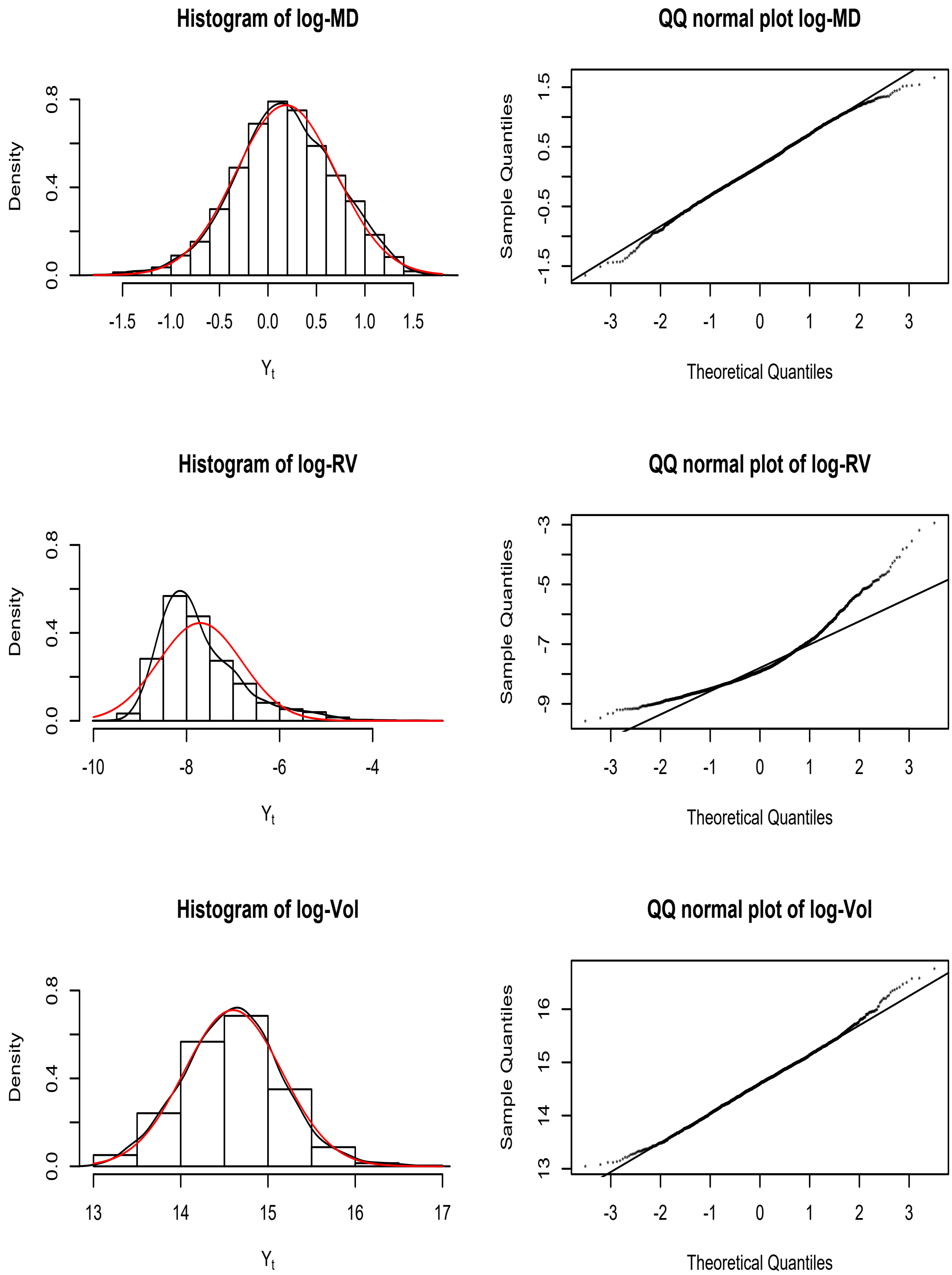

Figure A6.3 – Histograms and QQ normal plots of ALV log-data

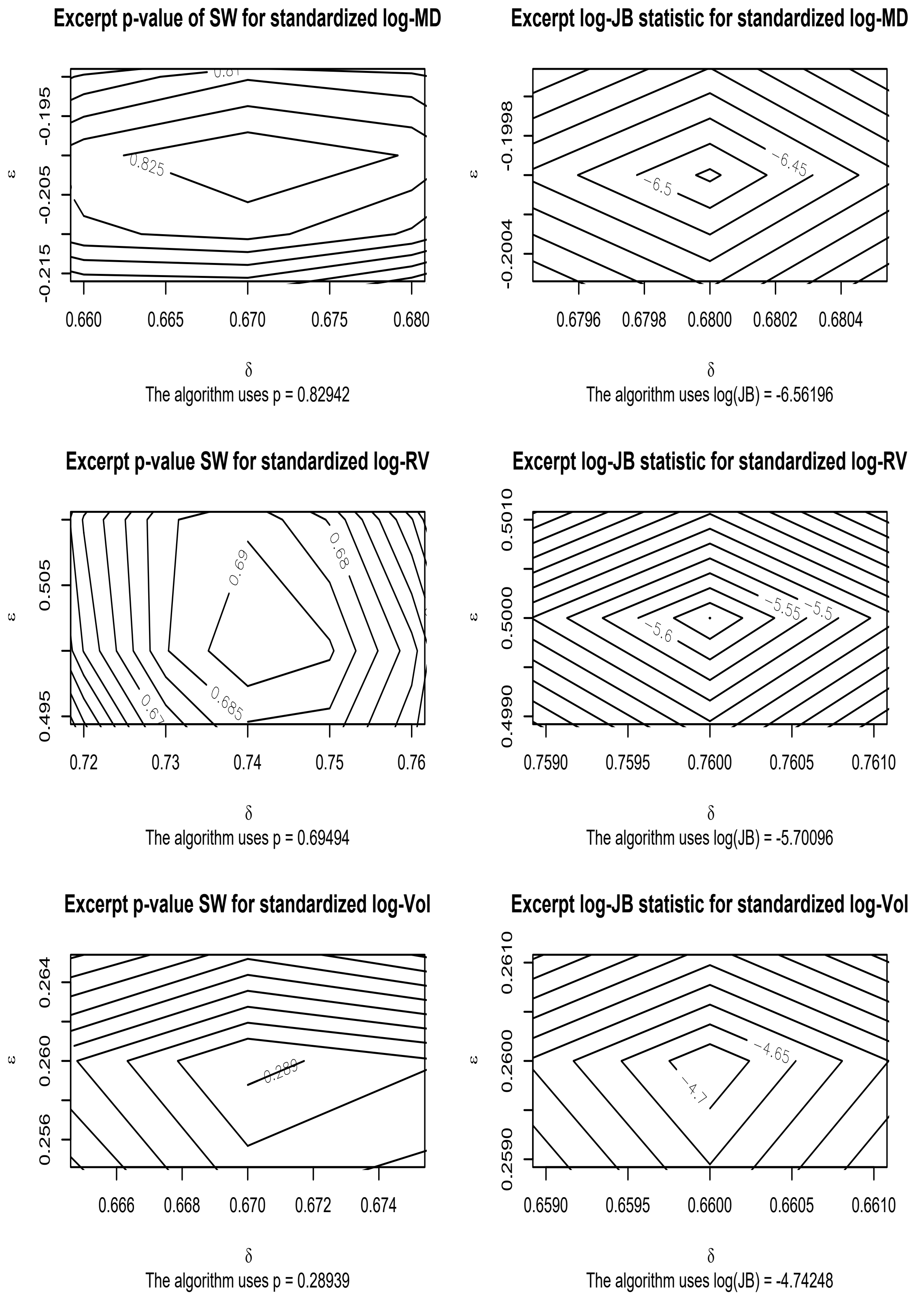

Figure A6.4 – Excerpt of contour plots for $\mathring{Y}_t^*$

Table A6.1 – Overview over denotation and composition of single and combination methods including NN

LLR	Local Linear Regression
RWC	Random Walk with constant drift
RWL	Random Walk with linear drift
RWLL	Random Walk with local linear drift
NNAR	Autoregressive Neural Network
NNARLL	Autoregressive Neural Network with local linear regression

Denotation	$C2_1$	$C2_2$	$C2_3$	$C2_4$	$C2_5$	$C2_6$	$C2_7$	$C2_8$	$C2_9$	$C2_{10}$
Combined Methods	LLR RWC	LLR RWL	LLR RWLL	LLR NNAR	LLR NNARLL	RWC RWL	RWC RWLL	RWC NNAR	RWC NNARLL	RWL RWLL
Denotation	$C2_{11}$	$C2_{12}$	$C2_{13}$	$C2_{14}$	$C2_{15}$	$C3_1$	$C3_2$	$C3_3$	$C3_4$	$C3_5$
Combined Methods	RWL NNAR	RWL NNARLL	RWLL NNAR	RWLL NNARLL	NNAR NNARLL	LLR RWC RWL	LLR RWC RWLL	LLR RWC NNAR	LLR RWC NNARLL	LLR RWL RWLL
Denotation	$C3_6$	$C3_7$	$C3_8$	$C3_9$	$C3_{10}$	$C3_{11}$	$C3_{12}$	$C3_{13}$	$C3_{14}$	$C3_{15}$
Combined Methods	LLR RWL NNAR	LLR RWL NNARLL	LLR RWLL NNAR	LLR RWLL NNARLL	LLR NNAR NNARLL	RWC RWL RWLL	RWC RWL NNAR	RWC RWL NNARLL	RWC RWLL NNAR	RWC RWLL NNARLL
Denotation	$C3_{16}$	$C3_{17}$	$C3_{18}$	$C3_{19}$	$C3_{20}$	$C4_1$	$C4_2$	$C4_3$	$C4_4$	$C4_5$
Combined Methods	RWC NNAR NNARLL	RWL RWLL NNAR	RWL RWLL NNARLL	RWL NNAR NNARLL	RWLL NNAR NNARLL	LLR RWC RWL RWLL	LLR RWC RWL NNAR	LLR RWC RWL NNARLL	LLR RWC RWLL NNAR	LLR RWC RWLL NNARLL
Denotation	$C4_6$	$C4_7$	$C4_8$	$C4_9$	$C4_{10}$	$C4_{11}$	$C4_{12}$	$C4_{13}$	$C4_{14}$	$C4_{15}$
Combined Methods	LLR RWC NNAR NNARLL	LLR RWL RWLL NNAR	LLR RWL RWLL NNARLL	LLR RWL NNAR NNARLL	LLR RWLL NNAR NNARLL	RWC RWL RWLL NNAR	RWC RWL RWLL NNARLL	RWC RWL NNAR NNARLL	RWC RWLL NNAR NNARLL	RWL RWLL NNAR NNARLL
Denotation	$C5_1$	$C5_2$	$C5_3$	$C5_4$	$C5_5$	$C5_6$	$C6$			
Combined Methods	LLR RWC RWL RWLL NNAR	LLR RWC RWL RWLL NNARLL	LLR RWC RWL NNAR NNARLL	LLR RWC RWLL NNAR NNARLL	LLR RWL RWLL NNAR NNARLL	RWC RWL RWLL NNAR NNARLL	LLR RWC RWL RWLL NNAR NNARLL			

Table A6.2 – Mean MASE, SMAE, MLCI and EoPoFi for log-GDP forecasts with NN methods and combinations

	LLR	RWC	RWL	RWLL	NNAR	NNARLL	$C2_1$	$C2_2$	$C2_3$	$C2_4$	$C2_5$	$C2_6$	$C2_7$
MASE	0.7804	0.7874	0.6402	1.0479	1.4814	0.9511	0.6573	0.5643	0.5604	0.5765	0.7991	0.4829	0.4558
SMAE	0.0025	0.0028	0.0024	0.0034	0.0053	0.0030	0.0021	0.0020	0.0020	0.0022	0.0025	0.0017	0.0015
mean CI	0.0083	0.0189	0.0180	0.0173	0.0180	0.0085	0.0086	0.0086	0.0084	0.0120	0.0064	0.0184	0.0178
PoFi	1.5625	0.2500	0.2500	0.3750	1.1250	1.9375	1.1875	0.6250	0.8750	0.6875	2.1250	0.2500	0.2500
	$C2_8$	$C2_9$	$C2_{10}$	$C2_{11}$	$C2_{12}$	$C2.13$	$C2_{14}$	$C2_{15}$	$C3_1$	$C3_2$	$C3_3$	$C3_4$	$C3_5$
MASE	0.7235	0.6600	0.6940	0.9624	0.6055	0.9985	0.6537	0.6678	0.4691	0.3463	0.5456	0.6303	0.4945
SMAE	0.0028	0.0020	0.0024	0.0034	0.0020	0.0033	0.0023	0.0025	0.0015	0.0012	0.0020	0.0019	0.0018
mean CI	0.0142	0.0112	0.0176	0.0139	0.0107	0.0134	0.0070	0.0108	0.0113	0.0113	0.0103	0.0074	0.0112
PoFi	0.6875	1.0000	0.2500	1.2500	0.6875	1.2500	1.0000	0.7500	0.3125	0.5000	0.7500	1.5625	0.4375
	$C3_6$	$C3_7$	$C3_8$	$C3_9$	$C3_{10}$	$C3_{11}$	$C3_{12}$	$C3_{13}$	$C3_{14}$	$C3_{15}$	$C3_{16}$	$C3_{17}$	$C3_{18}$
MASE	0.5465	0.6043	0.5851	0.5563	0.5375	0.3832	0.6531	0.4618	0.6379	0.4017	0.5548	0.8195	0.5492
SMAE	0.0020	0.0020	0.0020	0.0019	0.0018	0.0013	0.0025	0.0014	0.0022	0.0013	0.0020	0.0027	0.0020
mean CI	0.0101	0.0071	0.0099	0.0069	0.0090	0.0178	0.0145	0.0130	0.0142	0.0128	0.0106	0.0141	0.0126
PoFi	0.9375	1.5625	0.7500	1.0000	0.7500	0.2500	0.3125	0.3125	0.2500	0.3750	0.6250	0.7500	0.4375
	$C3_{19}$	$C3_{20}$	$C4_1$	$C4_2$	$C4_3$	$C4_4$	$C4_5$	$C4_6$	$C4_7$	$C4_8$	$C4_9$	$C4_{10}$	$C4_{11}$
MASE	0.5774	0.6331	0.3459	0.5101	0.4990	0.4432	0.4374	0.5191	0.5399	0.5103	0.4990	0.4898	0.5786
SMAE	0.0021	0.0022	0.0012	0.0019	0.0015	0.0016	0.0014	0.0017	0.0019	0.0018	0.0017	0.0017	0.0020
mean CI	0.0104	0.0101	0.0128	0.0110	0.0094	0.0108	0.0093	0.0086	0.0107	0.0091	0.0084	0.0083	0.0148
PoFi	1.0625	0.9375	0.2500	0.6250	0.6875	0.5000	0.8125	0.6250	0.6875	0.6875	0.8750	0.8125	0.2500
	$C4_{12}$	$C4_{13}$	$C4_{14}$	$C4_{15}$	$C5_1$	$C5_2$	$C5_3$	$C5_4$	$C5_5$	$C5_6$	$C6$		
MASE	0.3748	0.5072	0.4755	0.5695	0.4182	0.3978	0.4864	0.4152	0.4732	0.4415	0.4053		
SMAE	0.0013	0.0018	0.0017	0.0020	0.0015	0.0013	0.0017	0.0014	0.0017	0.0016	0.0014		
mean CI	0.0139	0.0116	0.0114	0.0113	0.0118	0.0108	0.0094	0.0093	0.0092	0.0124	0.0103		
PoFi	0.2500	0.5000	0.5000	0.7500	0.5000	0.4375	0.7500	0.6250	0.7500	0.4375	0.5000		

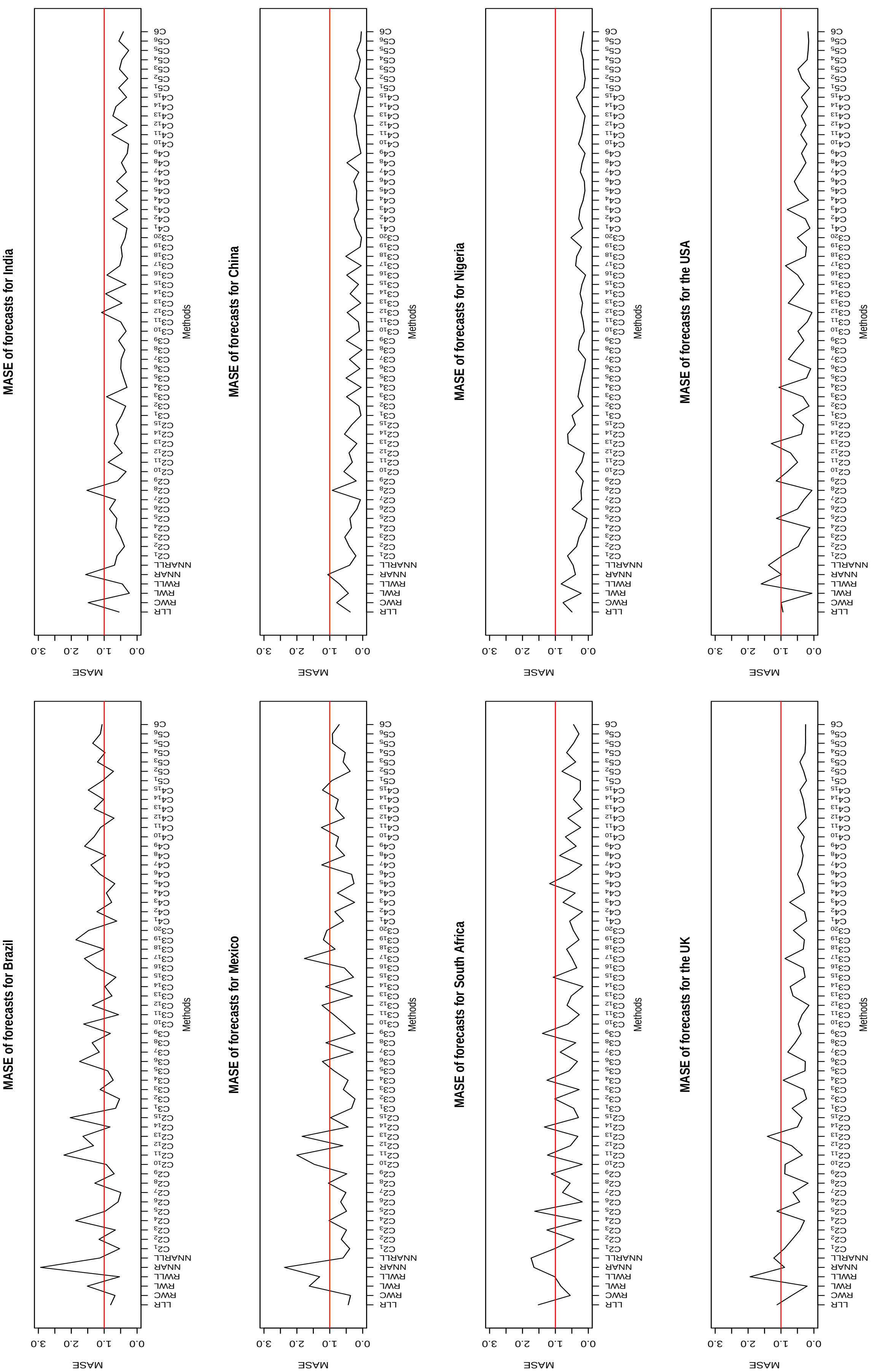

Figure A6.5 – MASE of all methods for all countries, k=5

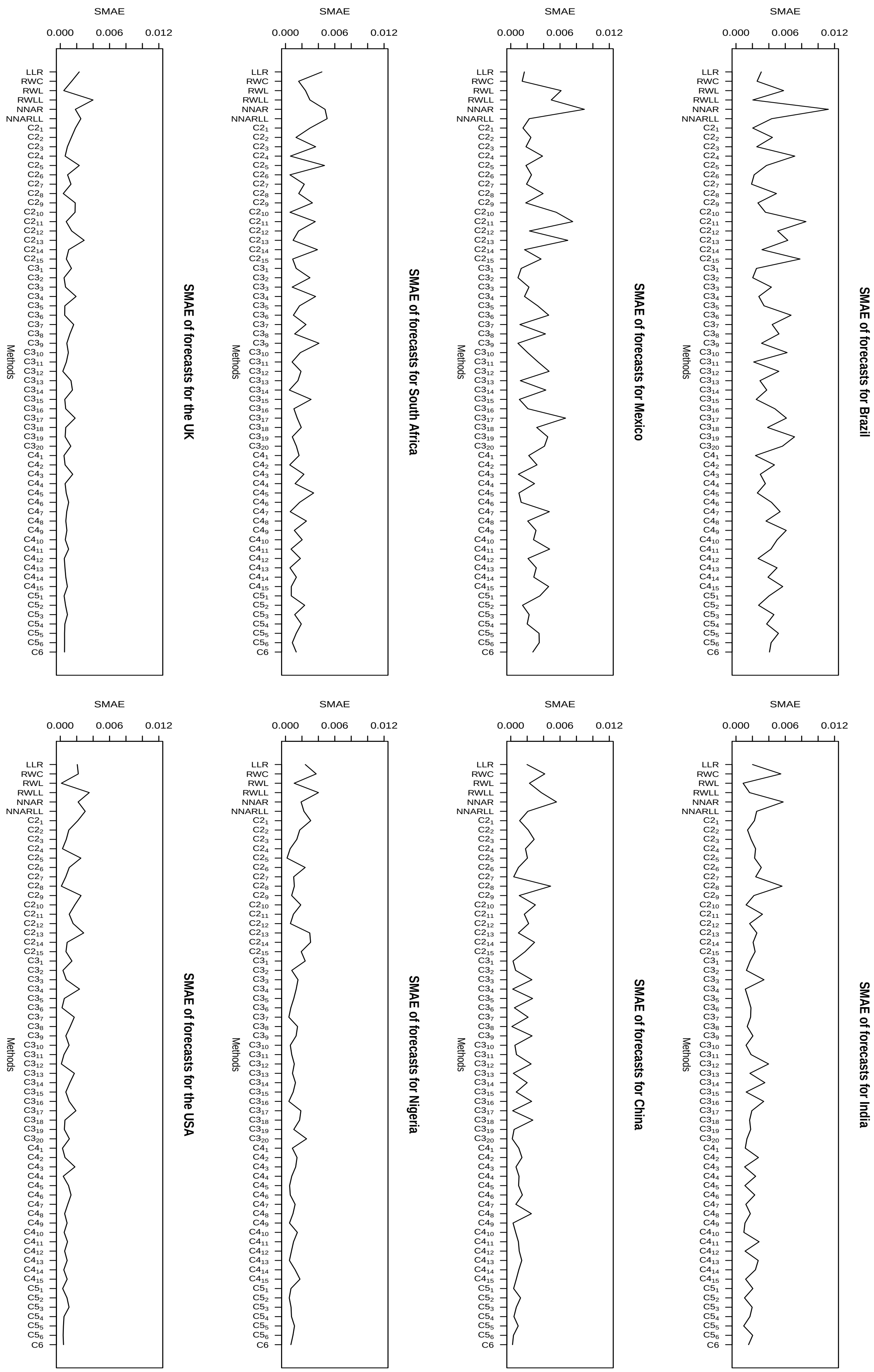

Figure A6.6 – Standardized MAE of all methods for all countries, k=5

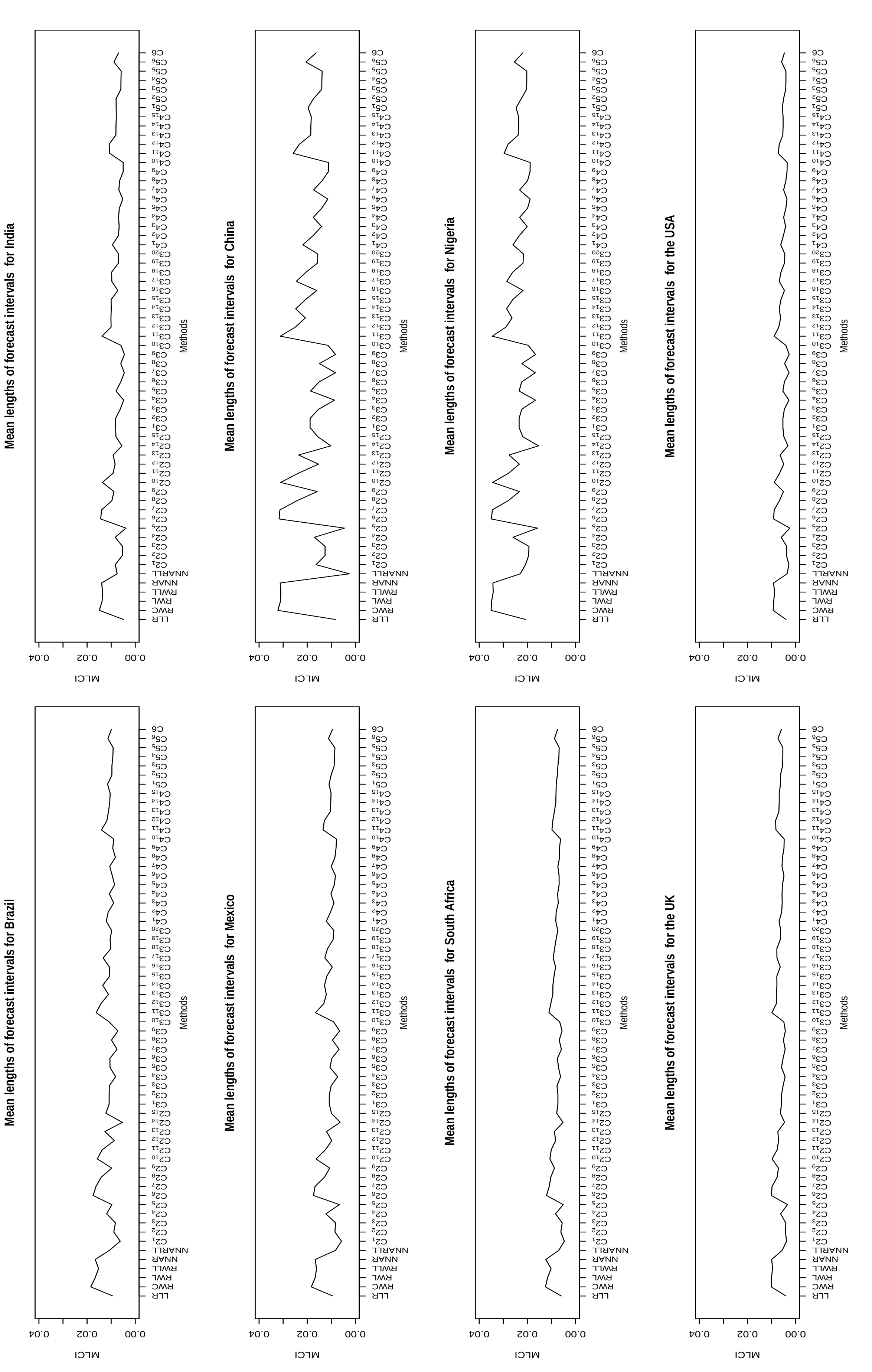

Figure A6.7 – Mean 95% FI lengths of all methods for all countries, k=5

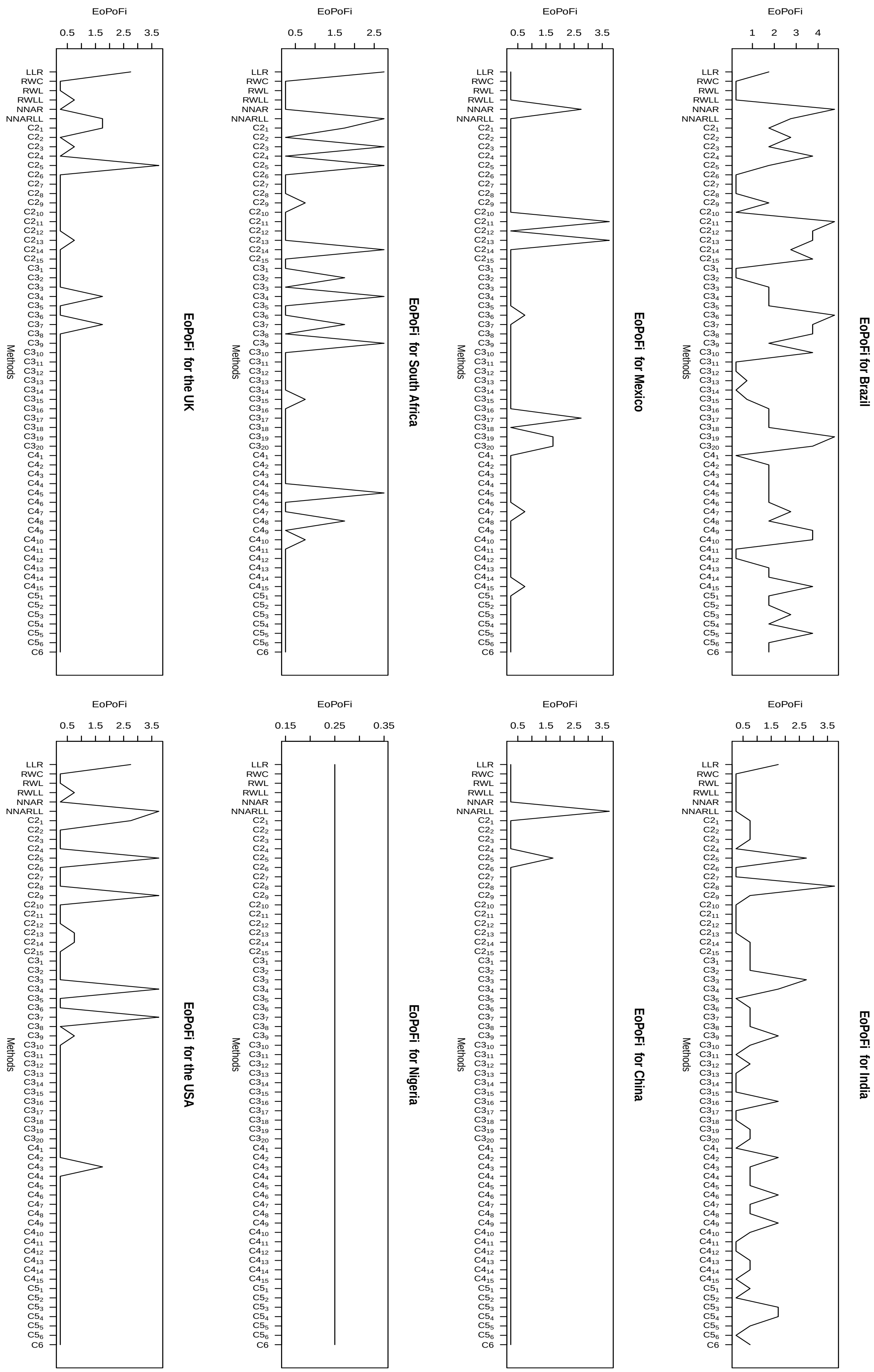

Figure A6.8 – EoPoFi of all methods for all countries, k=5

7 Summary of chapters and conclusion

The contribution of this work to current research can be derived individually from each of the contributions of the main chapters, but also as a development process as a whole. The main contribution of this work can be summarized as a semiparametric extension of the ACD model for the analysis of non-negative financial data. A two-stage model fitting process is presented, based on the decomposition of the data into a deterministic and a stochastic part, with the former being estimated using nonparametric methods and the latter using parametric methods. Motivated by the probable violation of the assumption of stationarity by a (typical) daily pattern of UHF financial data or long-term dynamics of HF data, the model extension presented here enables, among other things, the adaptation of already well-established models for the analysis of the stochastic part. The nonparametric local linear regression, the proposed IPI algorithm with an automatic bandwidth selection and also the extension to local polynomial regression for log-data prove to be very good tools for the estimation of the trend in empirical applications. The methods are improved and adapted and the arrangement of the chapters corresponds to the process of development. In detail, the chapters and their contributions to research in the corresponding fields can be summarized as follows:

Chapter 1 embeds the ACD model in a financial market microstructure theory framework and briefly describes its development since the original proposal in 1998. With regards to the contents of this work, a literature overview over semiparametric ACD model extensions is given and the chapters to follow are briefly summarized.

Chapter 2 describes the idea, first published by Feng (2014), of fitting the ACD model semiparametrically in a two-step procedure. The nonparametric smooth scale function, which is introduced into the ACD model to account for deterministic dynamics is estimated via local linear regression with a data-driven IPI for automatically choosing the optimal bandwidth required. The estimated scale function is removed from the data and any suitable parametric model can be fitted to the most likely stationary data. The trend estimation is a general step and does not require

prior assumptions on a parametric model to be fitted in the second step. A carried out simulation study evaluates the performance of the Semi-ACD model concerning the trend estimation and the model parameter estimation and also the automatic bandwidth selection of the IPI is examined. Different sample sizes are simulated to show the performance for an increase in sample size and also different daily patterns and different ACD models are simulated to discuss the performance of the proposed model for typical and difficult cases. The IPI is applied with different combinations of inflation method, inflation factor and variance factor for calculating the residual sum of autocovariances to derive a superior combination of these factors to propose for practical application. The simulation is repeated with a cubic spline estimation of the scale function with reference to the proposal of Engle and Russell (1998) and related works to compare the new methods to a method, that is already used in the prevailing literature. The evaluation of the simulation study provides the research contribution of this chapter: The here proposed Semi-ACD model with a local linear trend estimation using a bandwidth that is selected via a data-driven IPI, works well in practice and outperforms the cubic spline reference method. A superior combination of IPI factors is identified, which can be given as a recommendation for practical application. The application to two real financial trade duration examples supports the findings of the simulation study assessment.
Based on the results obtained for simulated data in chapter 2 the Semi-ACD model is further developed in chapter 3. In order to gain more flexibility of the model and ease its estimation, the semiparametric fitting idea is adapted to the (first-type) Log-ACD model of Bauwens and Giot (2000). It is further extended in the sense, that the model setup is generalized to a MEM, with the (Log-) ACD model as a special case and by not necessarily assuming a conditional distribution for the innovations. A general case is described with the log-data following a linear process and a special case with the log-data following a stationary and invertible ARMA process, i.e. under a conditional log-normal distribution assumption. Important properties of the model, i.e. the structure and moments for conditional distributions other than the log-normal one are discussed to not restrict the otherwise very flexible model at this point. Also conditions are derived under which the m-th moments of the original stationary data exist and under which heavy-tails of different levels can be modelled. The flexibility of the model is holistic, as the scale function estimation is a general procedure for which no assumption on the model to be fitted after removing the trend needs to be made. In addition to the model-theoretical research contribution, this chapter also provides rather application-relevant extensions of the IPI. Since there are no non-negativity constraints imposed on the parameters, the trend estimation is generalized to local polynomial regression and not limited to local linear regres-

sion anymore. Also the variance of the asymptotical optimal bandwidth, which had to be chosen manually before, is now estimated automatically as part of the IPI. Three methods are proposed, where two are parametric and AR(MA) model based and one is nonparametric using another IPI procedure. The research contributions of this chapters can be summarized as follows: The trend estimation is extended to local polynomial regression and includes an automatic variance factor estimation algorithm. The nonparametric method is proposed for flexibility reasons and the AR based method as a model based alternative. It is proven that the variance factor estimation does not affect the rate of convergence of the asymptotical optimal bandwidth and properties of the model are discussed for conditional distributions other than the log-normal one. The equivalent estimation of the scale function from the log-data is derived as recommendation.

Chapter 4 builds on the first two chapters by forecasting non-negative financial market data with the Semi-ACD and the Semi-Log-ACD model. The scale function is estimated following the automated IPI of chapter 3, i.e. it is estimated from the log-data with the nonparametric variance factor estimation method. The trend estimates are removed from the data and linearly extrapolated for the forecasts of the deterministic movement. Pursuing the aim of increased flexibility, model based bootstrap methods are applied to obtain point and 90% interval forecasts of the residuals. Again, the Semi-Log-ACD model under the conditional log-normal distribution is shown as a special case, for which Kalman filter methods are applied. Eventually the linearly extrapolated trend and the residual forecasts are combined to obtain the overall forecasts. The ACD, Log-ACD and Log-ACD model under the conditional normal distribution assumption are fitted semiparametrically and parametrically to three types of financial data of six firms. The assessment of the point and interval forecasting quality using two criteria each shows that, again, the semiparametric methods are clearly superior to the parametric ones. The model delivering the most precise and consistent results is the semiparametric Log-ACD model with the conditional normal distribution assumption. The contribution of this chapter to research in this field is the proposal of flexible forecasting methods for non-negative financial data. Also the prior made statements on the detrending procedure being universally applicable are confirmed, because it also works well in a forecasting framework. For the evaluation of the forecast quality a new criterion, EoPoFi, is developed and used. The comparison with the results obtained for the parametric models, once more show that the semiparametric methods are clearly better than the parametric ones. Overall, the results of the methods presented in chapters 2 to 4 should be of value for the ongoing research on analysing non-negative financial data. The nonparametric estimation method of the scale function presented here is not only shown to be clearly superior to the parametric methods, but it is also shown to be better than the originally proposed cubic spline method. Theoretical model properties, recommendations for application and further improvements of the methods for increasing flexibility are the main research contributions.

The origin of the methods presented lies in the analysis of (U)HF financial data. However, the proposed trend estimation methods also allow their use outside the financial data environment. Chapter 5 uses the proposed nonparametric trend estimation approach to improve the forecast accuracy of advanced and especially developing country GDP data. One of the methods applied is the adapted Semi-Log-ACD model forecasting method under the conditional normal distribution described in chapter 4. Furthermore, random walk models are used with a constant, a linear and a local linear drift estimated via the scale function estimation IPI as introduced in the previous chapters. These four individual methods are additionally combined in 11 ways, whereby the forecasts of the combined models are determined from the mean values of the individual forecast values. The application to examples of Log-GDP data of advanced and developing countries shows that the semiparametric regression model approach does not work consistently well due to structural breaks in the data. The conclusion to be drawn from the results of the point and interval forecasts obtained with the single methods and combined models is that the local linear trend estimation approach stabilizes the forecasts when it is combined with other methods. The flexibility of the LLR method improves the accuracy of trend identification by systematically adapting to the different phases of non-log-linear growth processes. The research contribution of this chapter is the adaptation of the Semi-Log-ACD model methods in a non-financial context. It is overall shown to work well, even though the results are inconclusive at some points. The general applicability and overall good practical performance of the local linear regression IPI is confirmed.

Chapter 6 gives possible future research questions, that are mostly derived from the findings of the individual chapters of this work. All of the proposals aim to further improve the accuracy of the fitted model in terms of describing the data's dynamics or in terms of forecasting. The approaches are applied at different points in the 2-step process. The contribution to research of this chapter is not that clearly identifiable (yet). Some of the first empirical results obtained are very promising, but all of these ideas need proper theoretical discussions and more profound empirical application. The contribution is considered to lie in the provision of future research questions. One promising proposal is to apply the nonparametric methods of the Semi-Log-ACD model in a GARCH model context to improve the estimation of the volatility of returns. The performance of the Semi-Log-GARCH model is examined in a simulation study and results on the bandwidth selection, the scale function estimation and the model parameter estimation are all in clear favour of this model proposal. A just comparison with the Semi-GARCH model performance is not yet possible, though. Motivated by the findings of chapter 4 that the method assuming a conditional normal distribution performs best, the Log-SAS transformation is proposed to further normalize the data. Two parameters of this transformation allow to adjust the skewness and kurtosis to the point where normality of the log-data is refined. A first algorithm is developed based on the JB and SW tests to find the optimal combination of these parameters. A first application to real financial data shows that this idea works well in the way intended and that it should be pursued in a timely manner. Based on the analysis carried out in chapter 5 the methods and combinations of methods used there are supplemented by NN methods with and without trend removal by local linear regression. The analysis is repeated with the increased number of single methods and combinations, but the results are inconclusive. The finding of chapter 5 that the local linear methods stabilize the forecasts is confirmed, but besides that no clear statement can be made. The combination of methods that was identified best in the prior analysis is not outperformed by any of the other methods or combinations of such. Further ideas are presented without first empirical results. The proposal of a local bandwidth for estimating the trend aims at improving the detrending procedure. The idea of using EF instead of (Q)ML methods would include another method that does not require prior knowledge on the distribution of the errors and, thus, allow for a semiparametric model parameter estimation. Applying a different bootstrap for forecasting is proposed, because the results of the applied bootstrap methods are shown to be not as good as the model assumption based methods. It could be investigated further, whether the bootstrap method is not as good, per se.

The arrangement of chapters shows on the one hand how the further developments of the Semi-ACD model and its extensions build on each other and on the other hand that a model, which already works very well, can still be further improved. Thus, research in this area is not yet near completion, but the contents of this work are expected to contribute positively to the ongoing research.

Bibliography

[1] I. S. Abramson. On bandwidth variation in kernel estimates-a square root law. *The annals of Statistics*, pages 1217–1223, 1982.

[2] A. R. Admati and P. Pfleiderer. A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies*, 1(1):3–40, 1988.

[3] N. K. Ahmed, A. F. Atiya, N. El Gayar, and H. El-Shishiny. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6):594–621, 2010.

[4] K. H. Aljuhani and L. I. Al Turk. Modification of the adaptive Nadaraya-Watson kernel regression estimator. *Scientific Research and Essays*, 9(22):966–971, 2014.

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

[6] D. Allen, K. Ng, and S. Peiris. Estimating and simulating Weibull models of risk or price durations: An application to ACD models. *The North American Journal of Economics and Finance*, 25:214–225, 2013a.

[7] D. Allen, K. Ng, and S. Peiris. The efficient modelling of high frequency transaction data: A new application of estimating functions in financial economics. *Economics Letters*, 120(1):117–122, 2013b.

[8] H. Allende, G. Ulloa, and H. Allende-Cid. Prediction Intervals in Linear and Nonlinear Time Series with Sieve Bootstrap Methodology. In J. Beran, Y. Feng, and H. Hebbel, editors, *Empirical Economic and Financial Research: Theory, Methods and Practice*, pages 255–273, Cham, 2015. Springer International Publishing.

[9] A. M. Alonso, D. Pena, and J. Romo. Forecasting time series with sieve bootstrap. *Journal of Statistical Planning and Inference*, 100(1):1–11, 2002.

[10] A. M. Alonso, D. Pena, and J. Romo. On sieve bootstrap prediction intervals. *Statistics and Probability Letters*, 65(1):13–20, 2003.

[11] A. M. Alonso, D. Pena, and J. Romo. Introducing model uncertainty in time series bootstrap. *Statistica Sinica*, pages 155–174, 2004.

[12] T. G. Andersen and T. Bollerslev. Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns. *The journal of Finance*, 52(3):975–1005, 1997.

[13] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Modeling and Forecasting Realized Volatility. *Econometrica*, 71(2):579–625, 2003.

[14] J. S. Armstrong and F. Collopy. Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8(1):69–80, 1992.

[15] J. M. Bates and C. W. Granger. The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969.

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

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

[18] L. Bauwens and P. Giot. Asymmetric ACD models: introducing price information in ACD models. *Empirical Economics*, 28(4):709–731, 2003.

[19] L. Bauwens, P. Giot, J. Grammig, and D. Veredas. A comparison of financial duration models via density forecasts. *International Journal of Forecasting*, 20(4):589–609, 2004.

[20] BCBS. *Basel III: Finalising post-crisis reforms*, 2017.

[21] M. Belfrage. *ACDm: Tools for Autoregressive Conditional Duration Models*, 2016. R package version 1.0.4.

[22] A. K. Bera and Y. Bilias. The MM, ME, ML, EL, EF and GMM approaches to estimation: a synthesis. *Journal of Econometrics*, 107(1):51–86, 2002.

[23] A. K. Bera, Y. Bilias, and P. Simlai. Estimating functions and equations: An essay on historical developments with applications to econometrics. *Palgrave Handbook of Econometrics*, 1:427–476, 2006.

[24] J. Beran and Y. Feng. SEMIFAR models - a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis*, 40(2):393–419, 2002.

[25] J. Beran and Y. Feng. Iterative plug-in algorithms for SEMIFAR models - definition, convergence, and asymptotic properties. *Journal of Computational and Graphical statistics*, 11(3):690–713, 2002a.

[26] J. Beran and Y. Feng. Local polynomial fitting with long-memory, short-memory and antipersistent errors. *Annals of the Institute of Statistical Mathematics*, 54(2):291–311, 2002b.

[27] J. Beran, Y. Feng, and S. Ghosh. Modelling long-range dependence and trends in duration series: an approach based on EFARIMA and ESEMIFAR models. *Statistical Papers*, 56(2):431–451, 2015.

[28] J. Beran, Y. Feng, and S. Heiler. Modifying the double smoothing bandwidth selector in nonparametric regression. *Statistical Methodology*, 6(5):447–465, 2009.

[29] J. Beran and D. Ocker. SEMIFAR forecasts, with applications to foreign exchange rates. *Journal of Statistical Planning and Inference*, 80(1-2):137–153, 1999.

[30] C. R. Bhatti. The Birnbaum–Saunders autoregressive conditional duration model. *Mathematics and Computers in Simulation*, 80(10):2062–2078, 2010.

[31] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[32] A. B. Bortoluzzo, P. A. Morettin, and C. M. Toloi. Time-varying autoregressive conditional duration model. *Journal of Applied Statistics*, 37(5):847–864, 2010.

[33] G. E. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252, 1964.

[34] M. Brockmann, T. Gasser, and E. Herrmann. Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association*, 88(424):1302–1309, 1993.

[35] P. Brockwell and R. Davis. *Time Series: Theory and Methods*. Springer-Verlag, New York, 1991 edition, 1991.

[36] P. J. Brockwell and R. A. Davis. *Introduction to time series and forecasting.* Springer-Verlag Inc, Berlin; New York, 2002.

[37] C. T. Brownlees, F. Cipollini, and G. M. Gallo. Intra-daily volume modeling and prediction for algorithmic trading. *Journal of Financial Econometrics*, 9(3):489–518, 2010.

[38] C. T. Brownlees, F. Cipollini, and G. M. Gallo. Multiplicative Error Models. In *Handbook of Volatility Models and Their Applications*, pages 223–247. Wiley-Blackwell, 2012.

[39] C. T. Brownlees and G. M. Gallo. Shrinkage estimation of semiparametric multiplicative error models. *International Journal of Forecasting*, 27(2):365–378, 2011.

[40] P. Bühlmann. Locally adaptive Lag-window spectral estimation. *Journal of Time Series Analysis*, 17(3):247–270, 1996.

[41] P. Bühlmann. Sieve bootstrap for time series. *Bernoulli*, 3(2):123–148, 1997.

[42] P. B¨uhlmann. Bootstraps for Time Series. *Statistical Science*, 17(1):52–72, 2002.

[43] P. B¨uhlmann and H. R. K¨unsch. Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis*, 31(3):295–310, 1999.

[44] P. Bühlmann and A. J. McNeil. Nonparametric GARCH models. In *Research Report/Seminar für Statistik and Department of Mathematics, Eidgenössische Technische Hochschule (ETH)*, volume 90. Seminar für Statistik, Eidgenössische Technische Hochschule, 1999.

[45] J. L. Castle, M. P. Clements, and D. F. Hendry. An overview of forecasting facing breaks. *Journal of Business Cycle Research*, 12(1):3–23, 2016.

[46] J. L. Castle, N. W. Fawcett, and D. F. Hendry. Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, 158(1):25–36, 2010.

[47] S. Chatterjee. Bootstrapping ARMA models: Some simulations. *IEEE transactions on systems, man, and cybernetics*, 16(2):294–299, 1986.

[48] B. Chen, Y. R. Gel, N. Balakrishna, and B. Abraham. Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes. *Journal of Forecasting*, 30(1):51–71, 2011.

[49] M.-Y. Cheng, J. Fan, and J. S. Marron. On automatic boundary corrections. *The Annals of Statistics*, pages 1691–1708, 1997.

[50] P. F. Christoffersen. Evaluating interval forecasts. *International economic review*, pages 841–862, 1998.

[51] S. Chung and S. Hwang. Stock return volatility based on intraday high frequency data: double-threshold ACD-GARCH model. *Korean Journal of Applied Statistics*, 29(1):221–230, 2016.

[52] M. P. Clements and D. F. Hendry. Forecasting economic processes. *International Journal of Forecasting*, 14(1):111–131, 1998.

[53] M. P. Clements and D. F. Hendry. Explaining Forecast Failure in Macroeconomics. In M. P. Clements and D. F. Hendry, editors, *A Companion to Economic Forecasting*, pages 539–571. Blackwell Publishing Ltd, 2002.

[54] K. O. Cogger. Proposals for research in time series forecasting. *International Journal of Forecasting*, 4(3):403–410, 1988. Special Issue on: The Future of Forecasting.

[55] A. Cosma and F. Galli. A Nonparametric ACD Model. Lsf research working paper series, Luxembourg School of Finance, University of Luxembourg, 2006.

[56] S. F. Crone, M. Hibon, and K. Nikolopoulos. Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, 27(3):635–660, 2011.

[57] C. Czado and S. Haug. An ACD-ECOGARCH (1, 1) Model. *Journal of Financial Econometrics*, 8(3):335–344, 2009.

[58] C. Czado and S. Haug. Finite sample properties of the QMLE in the ACD-ECOGARCH(1,1) model. Technical report, mediaTUM, Technical University Munich, 2010.

[59] D. B. Dahl. *xtable: Export Tables to LaTeX or HTML*, 2016. R package version 1.8-2.

[60] R. Dahlhaus. Fitting time series models to nonstationary processes. *The annals of Statistics*, 25(1):1–37, 1997.

[61] K. De Brabanter, J. De Brabanter, J. A. Suykens, and B. De Moor. Kernel regression in the presence of correlated errors. *Journal of Machine Learning Research*, 12(6):1955–1976, 2011.

[62] J. G. De Gooijer and R. J. Hyndman. 25 years of time series forecasting. *International Journal of Forecasting* , 22(3):443–473, 2006.

[63] G. De Luca and G. M. Gallo. Mixture processes for financial intradaily durations. *Studies in Nonlinear Dynamics & Econometrics*, 8(2): 1-20, 2004.

[64] G. De Luca and G. M. Gallo. Time-varying mixing weights in mixture autoregressive conditional duration models. *Econometric Reviews*, 28(1-3):102–120, 2008.

[65] G. De Luca and G. M. Gallo. A Time-varying Mixing Multiplicative Error Model for Realized Volatility. Econometrics Working Papers Archive wp2010_03, Universita' degli Studi di Firenze, Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti", 2010.

[66] G. De Luca and P. Zuccolotto. Finite and infinite mixtures for financial durations. *Metron*, 61(3):431–455, 2003.

[67] S. Demir and Ö. Toktamiş. On the adaptive nadaraya-watson kernel regression estimators. *Hacettepe Journal of Mathematics and Statistics*, 39(3), 2010.

[68] R. Deo, M. Hsieh, and C. M. Hurvich. Long memory in intertrade durations, counts and realized volatility of NYSE stocks. *Journal of Statistical Planning and Inference*, 140(12):3715–3733, 2010. Special Issue in Honor of Emanuel Parzen on the Occasion of his 80th Birthday and Retirement from the Department of Statistics, Texas A&M University.

[69] D. W. Diamond and R. E. Verrecchia. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics*, 18(2):277–311, 1987.

[70] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

[71] S. Dietz. A nonlinear model of economic data realted to the german automobile industry. *International Journal of Applied Operational Research*, 2(1):23–33, 2012.

[72] A. Dowla, E. Paparoditis, and D. N. Politis. Locally Stationary Processes and the Local Block Bootstrap. In M. G. Akritas and D. N. Politis, editors, *Recent Advances and Trends in Nonparametric Statistics*, pages 437 – 444. JAI, Amsterdam, 2003.

[73] A. Dowla, E. Paparoditis, and D. N. Politis. Local block bootstrap inference for trending time series. *Metrika*, 76(6):733–764, 2013.

[74] F. C. Drost and B. J. M. Werker. Semiparametric duration models. *Journal of Business & Economic Statistics*, 22(1):40–50, 2004.

[75] A. Dufour and R. F. Engle. The ACD model: predictability of the time between consecutive trades. *University of Reading and University of California at San Diego*, 35(3):463–497, 2000a.

[76] A. Dufour and R. F. Engle. Time and the price impact of a trade. *The Journal of Finance*, 55(6):2467–2498, 2000b.

[77] M. Dungey, X. Long, A. Ullah, and Y. Wang. A semiparametric conditional duration model. *Economics Letters*, 124(3):362–366, 2014.

[78] D. Easley, N. M. Kiefer, and M. O'Hara. The information content of the trading process. *Journal of Empirical Finance*, 4(2-3):159–186, 1997.

[79] D. Easley and M. O'Hara. Time and the process of security price adjustment. *The Journal of finance*, 47(2):577–605, 1992.

[80] B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.

[81] G. Elliott. Averaging and the optimal combination of forecasts. Technical report, University of California, San Diego, 2011.

[82] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.

[83] R. F. Engle. The econometrics of ultra-high-frequency data. *Econometrica*, 68(1):1–22, 2000.

[84] R. F. Engle. New frontiers for ARCH models. *Journal of Applied Econometrics*, 17(5):425–446, 2002.

[85] R. F. Engle and G. Lee. A long-run and short-run component model of stock return volatility. In R. F. Engle and H. White, editors, *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive WJ Granger*, pages 457–497. Oxford University Press, 1999.

[86] R. F. Engle and J. G. Rangel. The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies*, 21(3):1187–1222, 2008.

[87] R. F. Engle and J. R. Russell. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, pages 1127–1162, 1998.

[88] J. Fan and I. Gijbels. Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, pages 2008–2036, 1992.

[89] J. Fan and I. Gijbels. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 371–394, 1995.

[90] J. Fan, I. Gijbels, T.-C. Hu, and L.-S. Huang. A study of variable bandwidth selection for local polynomial regression. *Statistica Sinica*, pages 113–127, 1996.

[91] R. C. Feenstra, R. Inklaar, and M. P. Timmer. The next generation of the Penn World Table. *American Economic Review*, 105(10):3150–82, 2015.

[92] Y. Feng. A class of semiparametric GARCH models applied to quantitative risk management. Unpublished.

[93] Y. Feng. Simultaneously modeling conditional heteroskedasticity and scale change. *Econometric Theory*, 20(3):563–596, 2004a.

[94] Y. Feng. *Non- and Semiparametric Regression with Fractional Time Series Errors*. Theory and Applications to Financial Data. Habilitation Work, University of Konstanz, 2004b.

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

[96] Y. Feng. Long-memory non-negative subordinated Gaussian processes and the log-sinh-arcsinh transformation. 2018.

[97] Y. Feng and J. Beran. Robust estimation of the volatility tend in varying scale multiplicative error models. Unpublished.

[98] Y. Feng and J. Beran. Optimal convergence rates in non-parametric regression with fractional time series errors. *Journal of Time Series Analysis*, 34(1):30–39, 2013.

[99] Y. Feng, S. Forstinger, and C. Peitz. On the iterative plug-in algorithm for estimating diurnal patterns of financial trade durations. *Journal of Statistical Computation and Simulation*, 86(12):2291–2307, 2016.

[100] Y. Feng, T. Gries, and M. Fritz. Data-driven local polynomial for the trend and its derivatives in economic time series. Technical report, Paderborn University, CIE Center for International Economics, 2018. Under Revision for Statistical Modelling.

[101] Y. Feng and S. Heiler. A simple bootstrap bandwidth selector for local polynomial fitting. *Journal of Statistical Computation and Simulation*, 79(12):1425–1439, 2009.

[102] Y. Feng and C. Zhou. Forecasting financial market activity using a semiparametric fractionally integrated Log-ACD. *International Journal of Forecasting*, 31(2):349–363, 2015.

[103] M. Fernandes and J. Grammig. A family of autoregressive conditional duration models. *Journal of Econometrics*, 130(1):1–23, 2006.

[104] S. Figlewski. Conference Presentation at UCSD. 1995.

[105] R. Fildes and H. Stekler. The state of macroeconomic forecasting. *Journal of macroeconomics*, 24(4):435–468, 2002.

[106] S. Forstinger and Y. Feng. A semi-MEM for modelling long-term and conditional dynamics in non-negative financial data. Technical report, pre-print, Paderborn University, 2018.

[107] C. Francq, O. Wintenberger, and J.-M. Zakoïan. GARCH models without positivity constraints: Exponential or Log GARCH? *Journal of Econometrics*, 177(1):34–46, 2013.

[108] P. H. Franses. A note on the Mean Absolute Scaled Error. *International Journal of Forecasting*, 32(1):20–22, 2016.

[109] M. Fritz, S. Forstinger, Y. Feng, and T. Gries. Forecasting Economic Growth Processes: Improved Forecast Accuracy by Combining Local Linear and Standard Measures. Technical report, pre-print, Paderborn University, 2018.

[110] T. Gasser, A. Kneip, and W. Köhler. A flexible and fast method for automatic smoothing. *Journal of the american statistical association*, 86(415):643–652, 1991.

[111] A. A. Gerunov. Automating Analytics: Forecasting Time Series in Economics and Business. *Journal of Economics and Political Economy*, 3(2):340-349, 2016.

[112] J. Geweke. Modeling the persistence of conditional variances: a comment. *Econometric Reviews*, 5(1):57–61, 1986.

[113] E. Ghysels and J. Jasiak. GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model. *Studies in Nonlinear Dynamics & Econometrics*, 2(4):1–19, 1998.

[114] P. Giot. Time Transformations, intraday data and volatility models. *Journal of Computational Finance*, 4:31–62, 2000.

[115] L. R. Glosten and P. R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1):71–100, 1985.

[116] V. P. Godambe. An Optimum Property of Regular Maximum Likelihood Estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211, 1960.

[117] E. Gómez-Déniz and J. V. Pérez-Rodríguez. Conditional duration model and the unobserved market heterogeneity of traders: an infinite mixture of non-exponentials. *Revista Colombiana de Estadística*, 39(2):307–325, 2016.

[118] S. Gonçalves and D. Politis. Discussion: Bootstrap methods for dependent data: A review. *Journal of the Korean Statistical Society*, 40(4):383–386, 2011.

[119] C. Gourieroux, J. Jasiak, and G. Le Fol. Intra-day market activity. *Journal of Financial Markets*, 2(3):193–226, 1999.

[120] J. Grammig and K.-O. Maurer. Non-monotonic hazard functions and the autoregressive conditional duration model. *The Econometrics Journal*, 3(1):16–38, 2000.

[121] J. Grammig and M. Wellner. Modeling the interdependence of volatility and inter-transaction duration processes. *Journal of Econometrics*, 106(2):369–400, 2002.

[122] F. A. Graybill. *Theory and application of the linear model.* London, Duxberry Press, 1976.

[123] M. Hallin and D. La Vecchia. A Simple R-Estimation Method for Semiparametric Duration Models. Ecares working papers, 2017.

[124] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust statistics: the approach based on influence functions*. New York: Wiley, 1986.

[125] L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054, 1982.

[126] J. D. Hart. Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 173–187, 1991.

[127] A. C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990.

[128] A. C. Harvey and C. R. McKenzie. Algorithm AS 182: Finite Sample Prediction from ARIMA Processes. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 31(2):180–187, 1982.

[129] N. Hautsch. *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models*. Springer: Berlin Heidelberg, 2004.

[130] N. Hautsch. *Econometrics of financial high-frequency data*. Springer Science & Business Media, 2011.

[131] N. Hautsch and W. Pohlmeier. Econometric Analysis of Financial Transaction Data: Pitfalls and Opportunities. Technical report, Center of Finance and Econometrics (CoFE), University of Konstanz, 2001.

[132] D. F. Hendry and M. P. Clements. Pooling of forecasts. *The Econometrics Journal*, 7(1):1–31, 2004.

[133] E. Herrmann. Local bandwidth choice in kernel regression estimation. *Journal of Computational and Graphical Statistics*, 6(1):35–54, 1997.

[134] E. Herrmann and T. Gasser. Iterative plug-in algorithm for bandwidth selection in kernel regression estimation. *Preprint, Darmstadt Institute of Technology and University of Zurich*, 1994.

[135] E. Herrmann, T. Gasser, and A. Kneip. Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika*, 79(4):783–795, 1992.

[136] M. Hibon and T. Evgeniou. To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting*, 21(1):15–24, 2005.

[137] E. Hwang and D. W. Shin. Stationary bootstrap prediction intervals for garch (p, q). *Communications for Statistical Applications and Methods*, 20(1):41–52, 2013.

[138] R. J. Hyndman. Prediction intervals for NNETAR models. Hyndsight post, Dec. 2017. retrieved from https://robjhyndman.com/hyndsight/nnetar-prediction-intervals/.

[139] R. J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice.* OTexts, 2014.

[140] R. J. Hyndman and Y. Khandakar. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22, 2008.

[141] R. J. Hyndman and A. B. Köhler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, pages 679–688, 2006.

[142] C. M. Jarque and A. K. Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6(3):255–259, 1980.

[143] J. Jasiak. Persistence in Intertrade Durations. Technical report, York University, Department of Economics, 1999.

[144] N. Jeyasreedharan, D. E. Allen, and J. W. Yang. Yet another ACD model: The autoregressive conditional directional duration (ACDD) model. *Annals of Financial Economics*, 09(01):1450004, 2014.

[145] J. John and N. Draper. An alternative family of transformations. *Applied Statistics*, pages 190–197, 1980.

[146] M. Jones and A. Pewsey. Sinh-arcsinh distributions. *Biometrika*, 96(4):761–780, 2009.

[147] R. Karaa, S. Slim, and D. M. Hmaied. Trading Intensity and Informed Trading in the Tunis Stock Exchange. In M. e. H. Arouri, S. Boubaker, and D. K. Nguyen, editors, *Emerging Markets and the Global Economy: A Handbook*, 2013.

[148] R. Karaa, S. Slim, and D. M. Hmaied. Trading intensity and the volume-volatility relationship on the Tunis Stock Exchange. *Research in International Business and Finance*, 2017.

[149] M. Karanasos. The statistical properties of exponential ACD models. *Quantitative and Qualitative Analysis in Social Sciences*, 2(1):29–49, 2008.

[150] A. Kassambara. *ggpubr: 'ggplot2' Based Publication Ready Plots*, 2017. R package version 0.1.6.

[151] S. Kaur Bhogal and R. Thekke Variyam. Conditional Duration models for high-frequency data: A review on recent developments: Conditional and Duration models for high frequency data. *Journal of Economic Surveys*, 2018.

[152] M. Y. Khan and S. Mittnik. Nonlinear time series modeling and forecasting the seismic data of the Hindu Kush region. *Journal of Seismology*, 22(1):353–376, 2018.

[153] T. Krivobokova and G. Kauermann. A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association*, 102(480):1328–1337, 2007.

[154] S. S. M. Kwok, W. K. Li, and P. L. H. Yu. The autoregressive conditional marked duration model: Statistical inference to market microstructure. *Journal of Data Science*, 7, 2009.

[155] S. N. Lahiri. Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):386–404, 1999.

[156] M. Leadbetter. On a basis for "Peaks over Threshold" modeling. *Statistics & Probability Letters*, 12(4):357–362, 1991.

[157] H. Li and G. S. Maddala. Bootstrapping time series models. *Econometric Reviews*, 15(2):115–158, 1996.

[158] A. Lindner and K. Meyer. Extremal behavior of finite EGARCH Processes, 2003.

[159] X. Long, L. Su, and A. Ullah. Estimation and forecasting of dynamic conditional covariance: A semiparametric multivariate model. *Journal of Business & Economic Statistics*, 29(1):109–125, 2011.

[160] S. Makridakis and M. Hibon. The M3-Competition: results, conclusions and implications. *International journal of forecasting*, 16(4):451–476, 2000.

[161] S. Manganelli. Duration, volume and volatility impact of trades. *Journal of Financial Markets*, 8(4):377–399, 2005.

[162] E. Marçal and E. Junior. Is It Possible to Beat the Random Walk Model in Exchange Rate Forecasting? More Evidence for Brazilian Case. *Brazilian Review of Finance*, 14(1):65–88, 2016.

[163] T. H. McInish and R. A. Wood. An analysis of intraday patterns in bid/ask spreads for NYSE stocks. *the Journal of Finance*, 47(2):753–764, 1992.

[164] N. Meddahi, E. Renault, and B. Werker. GARCH and irregularly spaced data. *Economics Letters*, 90(2):200–204, 2006.

[165] R. A. Meese and K. Rogoff. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of international economics*, 14(1-2):3–24, 1983.

[166] A. Milhøj. A Multiplicative Parametrization of ARCH Models. Research report 101, University of Copenhagen: Institute of Statistics, 1987.

[167] C. Min, W. Guomin, W. Guofu, and J. Xuelei. ACD-GARCH model and its application to Chinese stock market. *Statistical Research*, 11:012, 2003.

[168] S. Mishra, L. Su, and A. Ullah. Semiparametric estimator of time series conditional variance. *Journal of Business & Economic Statistics*, 28(2):256–274, 2010.

[169] D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370, 1991.

[170] K. Ng and M. Peiris. Modelling high frequency transaction data in financial economics: A comparative study based on simulations. *Economic Computation and Economic Cybernetics Studies and Research*, 47(2):189–202, 2013.

[171] K. Ng, S. Peiris, and R. Gerlach. Estimation and forecasting with logarithmic autoregressive conditional duration models: A comparative study with an application. *Expert Systems with Applications*, 41(7):3323–3332, 2014.

[172] D. J. Nordman. A note on the stationary bootstrap's variance. *The Annals of Statistics*, 37(1):359–370, 2009.

[173] J. Opsomer, Y. Wang, and Y. Yang. Nonparametric regression with correlated errors. *Statistical Science*, pages 134–153, 2001.

[174] M. Pacurar. Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. *Journal of Economic Surveys*, 22(4):711–751, 2008.

[175] S. G. Pantula. Modeling the persistence of conditional variances: a comment. *Econometric Reviews*, 5(1):71–74, 1986.

[176] E. Paparoditis and D. Politis. The tapered block bootstrap for general statistics from stationary sequences. *The Econometrics Journal*, 5(1):131–148, 2002.

[177] L. Pascual, J. Romo, and E. Ruiz. Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics & Data Analysis*, 50(9):2293 -2312, 2006.

[178] D. Pathmanathan, K. Ng, and S. Peiris. On estimation of autoregressive conditional duration (ACD) models based on different error distributions. *Sri Lankan Journal of Applied Statistics*, 10:251–269, 2010.

[179] A. Patton, D. N. Politis, and H. White. Correction to 'Automatic block-length selection for the dependent bootstrap' by D. Politis and H. White. *Econometric Reviews*, 28(4):372–375, 2009.

[180] M. Peiris, K. Ng, and I. Ibrahim. A Review of Recent Developments of Financial Time Series: ACD Modelling using the Estimation Function Approach. *Sri Lankan Journal of Applied Statistics*, 8:1–17, 2007.

[181] C. Peitz. *Die parametrische und semiparametrische Analyse von Finanzzeitreihen: neue Methoden, Modelle und Anwendungsmöglichkeiten*. Springer-Verlag, 2016.

[182] M. Perlin. *fACD: Autoregressive Conditional Duration Models in R*, 2014. R package version 1.0/r5974.

[183] R. R. Picard and K. N. Berk. Data splitting. *The American Statistician*, 44(2):140–147, 1990.

[184] W. Pohlmeier and F. Gerhard. Identifying intraday volatility. In *Beiträge zur Mikro-und zur Makroökonomik*, pages 347–362. Springer, 2001.

[185] D. N. Politis and J. P. Romano. The stationary bootstrap. *Journal of the American Statistical association*, 89(428):1303–1313, 1994.

[186] D. N. Politis and H. White. Automatic Block-Length Selection for the Dependent Bootstrap. *Econometric Reviews*, 23(1):53–70, 2004.

[187] R. Prado and M. West. *Time series: modeling, computation, and inference*. CRC Press, 2010.

[188] M. B. Priestley. *Spectral analysis and time series / M.B. Priestley*. Academic Press, London; New York, 1981.

[189] W. Qian, C. A. Rolling, G. Cheng, and Y. Yang. On the Forecast Combination Puzzle. *arXiv preprint arXiv:1505.00475*, 2015.

[190] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[191] F.-É. Racicot, R. Théoret, and A. Coën. Forecasting irregularly spaced UHF financial data: Realized volatility vs UHF-GARCH models. *International Advances in Economic Research*, 14(1):112–124, 2008.

[192] K. Ranasinghe and M. J. Silvapulle. Estimation Under Inequality Constraints: Semiparametric Estimation of Conditional Duration Models. *Econometric Reviews*, 30(4):359–378, 2011.

[193] J. J. Reeves. Bootstrap prediction intervals for ARCH models. *International Journal of Forecasting*, 21(2):237–248, 2005.

[194] J. M. Rodríguez-Poo, D. Veredas, and A. Espasa. Semiparametric estimation for financial durations. In *High frequency financial econometrics*, pages 225–251. Springer, 2008.

[195] F. J. Rubio, E. O. Ogundimu, and J. L. Hutton. On modelling asymmetric data using two-piece sinh–arcsinh distributions. *Brazilian Journal of Probability and Statistics*, 30(3):485–501, 2016.

[196] J. R. Russell and R. F. Engle. Chapter 7 - Analysis of High-Frequency Data. In Y. AïT-SAHALIA and L. P. HANSEN, editors, *Handbook of Financial Econometrics: Tools and Techniques*, volume 1 of *Handbooks in Finance*, pages 383 -426. North-Holland, San Diego, 2010.

[197] P. W. Saart, J. Gao, and D. E. Allen. Semiparametric autoregressive conditional duration model: Theory and practice. *Econometric Reviews*, 34(6-10):849–881, 2015.

[198] R. Sakia. The Box-Cox transformation technique: a review. *The statistician*, pages 169–178, 1992.

[199] C. Savu and W. L. Ng. The SCoD Model: Analyzing Durations with a Semiparametric Copula Approach*. *International Review of Finance*, 5(1–2):55 – 74, 2006.

[200] D. Sena and N. K. Nagwani. A neural network autoregressive model to forecast per Capita disposable income. *Asian Research Publishing Network (ARPN)*, 11:13123–13128, 2016.

[201] T. A. Severini and W. H. Wong. Profile likelihood and conditionally parametric models. *The Annals of statistics*, pages 1768–1802, 1992.

[202] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[203] G. Shmueli and K. C. Lichtendahl. *Practical Time Series Forecasting with R: A Hands-On Guide.* Axelrod Schnall Publishers, 2016.

[204] R. Shumway and D. Stoffer. *Time Series Analysis and Its Applications: With R Examples.* Springer Texts in Statistics. Springer: New York, 2011.

[205] O. Simonsen. An empirical model for durations in stocks. *Annals of Finance*, 3(2):241–255, 2007.

[206] J. Šindelář. GDP Forecasting by Czech Institutions: An Empirical Evaluation. *Prague Economic Papers*, 2017(2):155–169, 2017.

[207] K. Singh. On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics*, pages 1187–1195, 1981.

[208] J. Smith and K. F. Wallis. A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, 71(3):331–355, 2009.

[209] R. Solgi and A. Mira. A Bayesian Semiparametric Multiplicative Error Model with an Application to Realized Volatility. *Journal of Computational and Graphical Statistics*, 22(3):558–583, 2013.

[210] L. Spierdijk. An empirical analysis of the role of the trading intensity in information dissemination on the NYSE. *Journal of empirical finance*, 11(2):163–184, 2004.

[211] J. G. Staniswalis. Local bandwidth selection for kernel estimates. *Journal of the American Statistical Association*, 84(405):284–288, 1989.

[212] H. O. Stekler. The future of macroeconomic forecasting: Understanding the forecasting process. *International Journal of Forecasting*, 23(2):237–248, 2007.

[213] R. A. Stine. Estimating Properties of Autoregressive Forecasts. *Journal of the American Statistical Association*, 82(400):1072–1078, 1987.

[214] J. H. Stock and M. W. Watson. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430, 2004.

[215] J. H. Stock and M. W. Watson. Twenty Years of Time Series Econometrics in Ten Pictures. *Journal of Economic Perspectives*, 31(2):59–86, 2017.

[216] A. Tay, C. Ting, Y. K. Tse, and M. Warachka. Using high-frequency transaction data to estimate the probability of informed trading. *Journal of financial Econometrics*, 7(3):288–311, 2009.

[217] A. S. Tay, C. Ting, Y. K. Tse, and M. Warachka. Transaction-data analysis of marked durations and their implications for market microstructure. Technical report, Singapore Management University School of Economics, Paper No 09-2004, 2004.

[218] A. Timmermann. Forecast combinations. *Handbook of economic forecasting*, 1:135–196, 2006.

[219] D. Veredas, J. Rodriguez-Poo, and A. Espasa. On the (intradaily) seasonality and dynamics of a financial point process: A semiparametric approach. CORE Discussion Paper 23, Université Catholique de Louvain, 2001.

[220] E. I. Vlahogianni, M. G. Karlaftis, and K. Kepaptsoglou. Nonlinear autoregressive conditional duration models for traffic congestion estimation. *Journal of Probability and Statistics*, 2011.

[221] H. White. Learning in artificial neural networks: A statistical perspective. *Neural computation*, 1(4):425–464, 1989.

[222] P. Wongsaart, J. Gao, and D. E. Allen. An alternative semiparametric regression approach to nonlinear duration modeling: theory and application. In *Econometric Society World Congress (10th: 2010: Shanghai, China) ESWC*, 2010.

[223] D. Wuertz, T. Setz, Y. Chalabi, C. Boudt, P. Chausse, and M. Miklovac. *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*, 2017. R package version 3042.83.

[224] H. Yang and X. Wu. Semiparametric EGARCH model with the case study of China stock market. *Economic Modelling*, 28(3):761–766, 2011.

[225] G. P. Zhang and D. M. Kline. Quarterly time-series forecasting with neural networks. *IEEE transactions on neural networks*, 18(6):1800–1814, 2007.

[226] M. Y. Zhang, J. R. Russell, and R. S. Tsay. A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, 104(1):179–207, 2001.

[227] Y. Zheng, Y. Li, and G. Li. On Fréchet autoregressive conditional duration models. *Journal of Statistical Planning and Inference*, 175:51–66, 2016.

[228] P. Zuccolotto. Modelling the impact of open volume on inter-trade autoregressive durations. *Metron*, 60(3-4):51–65, 2002.

[229] P. Zuccolotto. Quantile estimation in ultra-high frequency financial data: a comparison between parametric and semiparametric approach. *Statistical Methods and Applications*, 12(2):243–257, 2003.

Supplementary Appendix of Chapter 4

The tests for significance of the estimated scale functions test whether the difference between the time-varying scale function is statistically different to the constant trend, i.e. the mean. The asymptotic variances for correctly obtaining the confidence interval bounds are determined via the automatically determined variance factor and lag-window estimator with Bartlett-window weights. The underlying hypothesis is, that there is no difference between the time-varying and the constant scale function. The Null for the Shapiro Wilk tests is, that the data is normally distributed. These tests are carried out via the *shapiro.test* function implemented in R for a signficance level of $\alpha = 5\%$, as well. The results show, that the data and their transformations each are with a trend that is statistically significant different to the mean of the data and that for all data and their transformations the hypothesis of the Shapiro Wilk test that the data is normally distributed is rejected at the 5% significance level. A thorough discussion on the Shapiro Wilk test is not done here.

Table B4.1 – Scale function significance tests and Shapiro Wilk tests

		Scale function significance log-data	Scale function significance original data	Shapiro Wilk log-data	Shapiro Wilk original
		% outside 95% CI	% outside 95% CI	p-value	value
ALV	MD	68.9	68.5	0.00028	$< 2.2 * 10^{-16}$
	RV	53.3	16.2	$< 2.2 * 10^{-16}$	$< 2.2 * 10^{-16}$
	Vol	73.0	66.7	0.00028	$< 2.2 * 10^{-16}$
BMW	MD	59.2	67.6	0.01153	$< 2.2 * 10^{-16}$
	RV	60.9	28.9	$< 2.2 * 10^{-16}$	$< 2.2 * 10^{-16}$
	Vol	57.9	51.8	0.0011	$< 2.2 * 10^{-16}$
DBK	MD	71.0	76.5	0.00073	$< 2.2 * 10^{-16}$
	RV	61.9	15.8	$< 2.2 * 10^{-16}$	$< 2.2 * 10^{-16}$
	Vol	42.5	38.7	$8.7 * 10^{-8}$	$< 2.2 * 10^{-16}$
SIE	MD	60.4	63.4	$2.1 * 10^{-8}$	$< 2.2 * 10^{-16}$
	RV	54.1	16.8	$< 2.2 * 10^{-16}$	$< 2.2 * 10^{-16}$
	Vol	82.8	88.2	$1.7 * 10^{-5}$	$< 2.2 * 10^{-16}$
TKA	MD	70.7	79.5	0.00025	$< 2.2 * 10^{-16}$
	RV	49.2	20.5	$< 2.2 * 10^{-16}$	$< 2.2 * 10^{-16}$
	Vol	60.4	55.3	0.0051	$< 2.2 * 10^{-16}$
VW	MD	95.9	93.1	$< 2.2 * 10^{-16}$	$< 2.2 * 10^{-16}$
	RV	63.0	17.8	$< 2.2 * 10^{-16}$	$< 2.2 * 10^{-16}$
	Vol	56.5	23.5	$4.0 * 10^{-9}$	$< 2.2 * 10^{-16}$

Table B4.2 – Point forecast evaluation for k=10

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE
ALV	MD	2.3154	0.5574	2.5512	0.6141	2.5454	0.6127	2.8620	0.6889	2.8044	0.6751	2.8399	0.6836
	RV	0.2033	0.0846	0.1499	0.0623	0.1470	0.0611	0.1739	0.0723	0.3200	0.1331	0.1983	0.0825
	Vol	0.5112	0.1648	0.4548	0.1467	0.4519	0.1457	0.5813	0.1875	0.6934	0.2236	0.5748	0.1854
BMW	MD	0.9810	0.2425	1.0277	0.2541	1.0406	0.2572	1.0237	0.2531	0.8270	0.2045	1.0285	0.2543
	RV	0.2477	0.0946	0.2207	0.0842	0.2150	0.0821	0.2972	0.1134	0.4967	0.1896	0.3127	0.1194
	Vol	0.4946	0.1633	0.4528	0.1495	0.4525	0.1494	0.5089	0.1680	0.5684	0.1876	0.5031	0.1661
DBK	MD	1.3167	0.3337	1.3177	0.3340	1.3129	0.3328	1.3316	0.3375	1.3284	0.3367	1.3402	0.3397
	RV	0.2063	0.0887	0.1987	0.0854	0.2003	0.0861	0.2033	0.0874	0.3244	0.1395	0.2178	0.0936
	Vol	0.9041	0.2896	0.9002	0.2883	0.9007	0.2885	0.9026	0.2891	0.9441	0.3023	0.9037	0.2894
SIE	MD	0.9172	0.2171	0.9968	0.2360	0.9992	0.2365	1.0699	0.2533	0.9333	0.2209	1.0619	0.2514
	RV	0.2068	0.0869	0.1580	0.0664	0.1540	0.0647	0.2104	0.0884	0.3924	0.1649	0.2313	0.0972
	Vol	0.2863	0.0971	0.2771	0.0940	0.2720	0.0923	0.3001	0.1018	0.3273	0.1110	0.2993	0.1015
TKA	MD	1.1215	0.2995	1.3408	0.3580	1.2987	0.3468	1.3695	0.3657	1.7562	0.4689	1.3838	0.3695
	RV	1.1476	0.4627	0.4777	0.1926	0.4499	0.1814	0.6070	0.2448	2.0547	0.8285	0.6233	0.2513
	Vol	0.7440	0.2598	0.7163	0.2502	0.7061	0.2466	0.7713	0.2694	1.6357	0.5713	0.7800	0.2724
VW	MD	0.2974	0.1135	0.3739	0.1427	0.3450	0.1317	0.3343	0.1276	0.2874	0.1097	0.3317	0.1266
	RV	0.2104	0.1007	0.1505	0.0721	0.1403	0.0672	0.1887	0.0904	0.3249	0.1556	0.2009	0.0962
	Vol	0.8139	0.3151	0.7200	0.2787	0.5672	0.2196	0.5311	0.2056	0.8277	0.3204	0.5854	0.2266
mean criterion		0.7978	0.2243	0.8082	0.2243	0.8074	0.2240	0.8761	0.2441	0.9230	0.2658	0.8799	0.2460
mean rank		2.8248	2.8095	2.0221	2.0084	1.7215	1.7082	3.7297	3.7113	4.5497	4.5190	4.2313	4.2119

Table B4.3 – Point forecast evaluation for k=20

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE
ALV	MD	3.1158	0.7501	3.3990	0.8182	3.3942	0.8171	3.8264	0.9211	3.8502	0.9268	3.7917	0.9127
	RV	0.2973	0.1236	0.2049	0.0852	0.1991	0.0828	0.2432	0.1011	0.5362	0.2229	0.3008	0.1251
	Vol	0.5953	0.1920	0.5141	0.1658	0.5088	0.1641	0.6980	0.2251	0.8528	0.2750	0.6946	0.2240
BMW	MD	1.3266	0.3279	1.3955	0.3450	1.4155	0.3499	1.3750	0.3399	1.0179	0.2516	1.3912	0.3439
	RV	0.2996	0.1143	0.2642	0.1008	0.2573	0.0982	0.3627	0.1384	0.7243	0.2765	0.3924	0.1498
	Vol	0.5292	0.1747	0.4628	0.1528	0.4618	0.1525	0.5536	0.1827	0.6479	0.2139	0.5584	0.1843
DBK	MD	1.5256	0.3867	1.5956	0.4044	1.5833	0.4013	1.6341	0.4142	1.5437	0.3913	1.6433	0.4165
	RV	0.2800	0.1204	0.2504	0.1076	0.2427	0.1044	0.2800	0.1204	0.4754	0.2044	0.3071	0.1320
	Vol	1.3630	0.4365	1.3304	0.4261	1.3340	0.4272	1.3290	0.4256	1.3938	0.4464	1.3368	0.4281
SIE	MD	1.4609	0.3458	1.5931	0.3771	1.6013	0.3791	1.7151	0.4060	1.4711	0.3482	1.6772	0.3970
	RV	0.2865	0.1204	0.2140	0.0899	0.2115	0.0889	0.2904	0.1220	0.6269	0.2634	0.3283	0.1379
	Vol	0.3310	0.1123	0.2970	0.1007	0.2845	0.0965	0.3707	0.1257	0.4287	0.1454	0.3646	0.1236
TKA	MD	1.2398	0.3311	1.5446	0.4125	1.4749	0.3938	1.6543	0.4417	2.0419	0.5452	1.6531	0.4414
	RV	0.8431	0.3400	0.4337	0.1749	0.3996	0.1611	0.6166	0.2486	2.0787	0.8382	0.6533	0.2634
	Vol	0.7143	0.2495	0.6631	0.2316	0.6433	0.2247	0.8012	0.2798	1.7311	0.6046	0.8109	0.2832
VW	MD	0.4819	0.1839	0.5694	0.2173	0.5316	0.2029	0.5188	0.1980	0.4675	0.1785	0.5143	0.1963
	RV	0.2345	0.1123	0.1841	0.0881	0.1689	0.0809	0.2312	0.1107	0.3993	0.1911	0.2545	0.1218
	Vol	1.1630	0.4502	1.1317	0.4381	0.9544	0.3694	0.9213	0.3566	1.1931	0.4618	0.9771	0.3782
mean criterion		1.0369	0.2918	1.0463	0.2895	1.0441	0.2886	1.1446	0.3187	1.2269	0.3565	1.1574	0.3241
mean rank		2.5300	2.5114	2.4264	2.4101	1.9257	1.9098	3.4363	3.4138	4.5724	4.5276	4.2392	4.2150

Table B4.4 – Point forecast evaluation for k=30

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE
ALV	MD	3.4012	0.8187	3.7265	0.8971	3.7214	0.8958	4.2791	1.0301	4.4253	1.0653	4.2214	1.0162
	RV	0.3440	0.1431	0.2256	0.0938	0.2174	0.0904	0.2869	0.1193	0.6972	0.2899	0.3747	0.1558
	Vol	0.6620	0.2135	0.5757	0.1856	0.5696	0.1837	0.7754	0.2500	0.9422	0.3038	0.7704	0.2485
BMW	MD	1.2920	0.3194	1.3754	0.3400	1.4005	0.3462	1.3233	0.3271	0.8907	0.2202	1.3446	0.3324
	RV	0.3117	0.1190	0.2592	0.0989	0.2482	0.0948	0.4180	0.1596	0.9410	0.3592	0.4727	0.1804
	Vol	0.5142	0.1697	0.4541	0.1499	0.4514	0.1490	0.5414	0.1787	0.6536	0.2157	0.5455	0.1801
DBK	MD	1.3965	0.3540	1.3610	0.3450	1.3668	0.3465	1.3735	0.3481	1.3476	0.3416	1.3764	0.3489
	RV	0.2410	0.1036	0.2210	0.0950	0.2162	0.0930	0.2438	0.1048	0.6005	0.2582	0.2820	0.1212
	Vol	1.7456	0.5590	1.7962	0.5752	1.7997	0.5764	1.7884	0.5727	1.6764	0.5369	1.7427	0.5581
SIE	MD	1.5991	0.3785	1.7657	0.4180	1.7750	0.4202	1.8985	0.4494	1.5644	0.3703	1.8493	0.4378
	RV	0.3057	0.1285	0.2187	0.0919	0.2162	0.0908	0.3168	0.1331	0.7910	0.3323	0.3843	0.1614
	Vol	0.3611	0.1225	0.3279	0.1112	0.3168	0.1074	0.4100	0.1391	0.4803	0.1629	0.4021	0.1364
TKA	MD	1.3697	0.3658	1.6596	0.4432	1.5909	0.4248	1.8146	0.4846	2.1656	0.5783	1.8054	0.4821
	RV	0.6863	0.2767	0.3938	0.1588	0.3562	0.1436	0.6312	0.2545	2.0693	0.8344	0.6834	0.2755
	Vol	0.6724	0.2349	0.6173	0.2156	0.6019	0.2102	0.7918	0.2766	1.7018	0.5944	0.8060	0.2815
VW	MD	0.3932	0.1501	0.5228	0.1996	0.4640	0.1771	0.4415	0.1685	0.3600	0.1374	0.4319	0.1649
	RV	0.2413	0.1155	0.1883	0.0901	0.1674	0.0802	0.2549	0.1220	0.4575	0.2190	0.2909	0.1393
	Vol	1.0746	0.4160	1.0978	0.4249	0.8175	0.3165	0.7746	0.2998	1.1341	0.4390	0.8808	0.3410
mean criterion		1.1009	0.3111	1.1105	0.3090	1.1101	0.3084	1.2255	0.3434	1.3527	0.3990	1.2367	0.3491
mean rank		2.8312	2.8119	2.5259	2.5099	2.2248	2.2095	3.6418	3.6160	3.9941	3.9359	3.9473	3.9180

Table B4.5 – Point forecast evaluation for k=40

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE
ALV	MD	3.2340	0.7785	3.5581	0.8565	3.5489	0.8543	4.1473	0.9984	4.4093	1.0614	4.0924	0.9851
	RV	0.3640	0.1514	0.2284	0.0950	0.2184	0.0908	0.3127	0.1300	0.8187	0.3404	0.4363	0.1814
	Vol	0.6466	0.2085	0.5563	0.1794	0.5483	0.1768	0.7692	0.2481	0.9661	0.3116	0.7678	0.2476
BMW	MD	1.3026	0.3220	1.3916	0.3440	1.4206	0.3512	1.3085	0.3235	0.7659	0.1893	1.3325	0.3294
	RV	0.3115	0.1189	0.2493	0.0952	0.2359	0.0900	0.4568	0.1744	1.0988	0.4194	0.5383	0.2055
	Vol	0.5233	0.1728	0.4482	0.1479	0.4459	0.1472	0.5686	0.1877	0.7064	0.2332	0.5731	0.1892
DBK	MD	1.2973	0.3288	1.2695	0.3218	1.2672	0.3212	1.2726	0.3226	1.2640	0.3204	1.2782	0.3240
	RV	0.2127	0.0914	0.1899	0.0817	0.1899	0.0816	0.2295	0.0987	0.7509	0.3228	0.2988	0.1284
	Vol	1.5807	0.5062	1.6523	0.5292	1.6518	0.5290	1.6226	0.5196	1.5266	0.4889	1.5701	0.5028
SIE	MD	1.6456	0.3895	1.8273	0.4326	1.8367	0.4348	1.9461	0.4607	1.5554	0.3682	1.8845	0.4461
	RV	0.3219	0.1352	0.2282	0.0959	0.2257	0.0948	0.3490	0.1466	0.9236	0.3881	0.4411	0.1853
	Vol	0.3277	0.1111	0.2930	0.0994	0.2814	0.0954	0.3942	0.1337	0.4780	0.1621	0.3821	0.1296
TKA	MD	1.3399	0.3578	1.6024	0.4279	1.5303	0.4086	1.8075	0.4826	2.1145	0.5646	1.8047	0.4819
	RV	0.5982	0.2412	0.3633	0.1465	0.3217	0.1297	0.6608	0.2664	2.0739	0.8362	0.7366	0.2970
	Vol	0.6331	0.2211	0.5741	0.2005	0.5570	0.1946	0.8003	0.2795	1.6809	0.5871	0.8278	0.2891
VW	MD	0.3782	0.1444	0.5119	0.1954	0.4511	0.1722	0.4255	0.1624	0.3250	0.1241	0.4124	0.1574
	RV	0.2294	0.1098	0.1774	0.0849	0.1552	0.0743	0.2624	0.1256	0.4897	0.2344	0.3095	0.1482
	Vol	1.0547	0.4083	1.1212	0.4340	0.7949	0.3077	0.7446	0.2882	1.1246	0.4353	0.8685	0.3362
mean criterion		1.0525	0.2976	1.0604	0.2945	1.0585	0.2936	1.1876	0.3337	1.3674	0.4097	1.2097	0.3437
mean rank		2.8311	2.8119	2.7249	2.7095	2.0236	2.0090	3.6457	3.6174	4.0099	3.9419	3.9538	3.9205

Table B4.6 – Point forecast evaluation for k=50

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE
ALV	MD	2.7910	0.6719	3.0808	0.7416	3.0781	0.7410	3.6880	0.8878	3.9883	0.9601	3.6189	0.8712
	RV	0.3574	0.1486	0.2190	0.0910	0.2096	0.0871	0.3217	0.1338	0.8985	0.3736	0.4761	0.1980
	Vol	0.5956	0.1921	0.5132	0.1655	0.5086	0.1640	0.7221	0.2329	0.9173	0.2958	0.7234	0.2333
BMW	MD	1.1983	0.2962	1.2815	0.3168	1.3102	0.3239	1.1842	0.2927	0.7626	0.1885	1.2100	0.2991
	RV	0.2856	0.1090	0.2234	0.0853	0.2108	0.0805	0.4657	0.1777	1.1979	0.4572	0.5774	0.2204
	Vol	0.5018	0.1657	0.4280	0.1413	0.4274	0.1411	0.5592	0.1846	0.7083	0.2338	0.5678	0.1874
DBK	MD	1.1911	0.3019	1.1594	0.2939	1.1588	0.2937	1.1657	0.2955	1.1576	0.2934	1.1668	0.2957
	RV	0.1926	0.0828	0.1777	0.0764	0.1863	0.0801	0.2191	0.0942	0.8866	0.3811	0.3122	0.1342
	Vol	1.4812	0.4744	1.5664	0.5016	1.5665	0.5017	1.5243	0.4882	1.3988	0.4480	1.4466	0.4633
SIE	MD	1.6146	0.3822	1.7929	0.4244	1.8045	0.4271	1.8894	0.4473	1.4862	0.3518	1.8177	0.4303
	RV	0.3211	0.1349	0.2173	0.0913	0.2149	0.0903	0.3703	0.1556	1.0359	0.4352	0.4946	0.2078
	Vol	0.3360	0.1140	0.2980	0.1011	0.2854	0.0968	0.4074	0.1382	0.4908	0.1665	0.3962	0.1344
TKA	MD	1.4347	0.3831	1.7121	0.4572	1.6327	0.4360	1.9573	0.5227	2.2318	0.5960	1.9444	0.5192
	RV	0.5404	0.2179	0.3395	0.1369	0.2928	0.1181	0.6859	0.2766	2.0773	0.8376	0.7862	0.3170
	Vol	0.6413	0.2240	0.5750	0.2008	0.5576	0.1948	0.8381	0.2927	1.6907	0.5905	0.8710	0.3042
VW	MD	0.3453	0.1318	0.4731	0.1806	0.4108	0.1568	0.3856	0.1472	0.3321	0.1268	0.3752	0.1432
	RV	0.2152	0.1030	0.1657	0.0793	0.1462	0.0700	0.2593	0.1241	0.5103	0.2443	0.3193	0.1529
	Vol	1.0157	0.3932	1.1014	0.4264	0.7703	0.2982	0.7271	0.2815	1.1066	0.4284	0.8505	0.3292
mean criterion		0.9550	0.2714	0.9611	0.2682	0.9618	0.2681	1.0945	0.3097	1.3240	0.4035	1.1221	0.3225
mean rank		2.9286	2.9109	2.5223	2.5085	2.2211	2.2080	3.4466	3.4178	4.0198	3.9457	4.0577	4.0220

Table B4.7 – Point forecast evaluation for k=60

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE
ALV	MD	2.6714	0.6431	2.9399	0.7077	2.9398	0.7077	3.5267	0.8489	3.8458	0.9258	3.4641	0.8339
	RV	0.3680	0.1530	0.2335	0.0971	0.2260	0.0940	0.3479	0.1446	0.9552	0.3972	0.5239	0.2179
	Vol	0.5887	0.1899	0.5076	0.1637	0.5037	0.1624	0.7245	0.2336	0.9314	0.3004	0.7279	0.2347
BMW	MD	1.0943	0.2705	1.1609	0.2870	1.1844	0.2928	1.0966	0.2711	0.9014	0.2228	1.1223	0.2774
	RV	0.2796	0.1067	0.2281	0.0871	0.2203	0.0841	0.4568	0.1743	1.2268	0.4683	0.5791	0.2211
	Vol	0.4928	0.1627	0.4357	0.1438	0.4341	0.1433	0.5373	0.1774	0.6815	0.2250	0.5494	0.1814
DBK	MD	1.2399	0.3143	1.1435	0.2898	1.1564	0.2931	1.1159	0.2828	1.1224	0.2845	1.1113	0.2817
	RV	0.1888	0.0812	0.1932	0.0831	0.2077	0.0893	0.2102	0.0903	0.9781	0.4205	0.3233	0.1390
	Vol	1.4872	0.4763	1.6030	0.5134	1.6028	0.5133	1.5362	0.4920	1.3477	0.4316	1.4103	0.4517
SIE	MD	1.5665	0.3708	1.7288	0.4092	1.7427	0.4125	1.8097	0.4284	1.4532	0.3440	1.7290	0.4093
	RV	0.3227	0.1356	0.2291	0.0963	0.2279	0.0957	0.3820	0.1605	1.0949	0.4600	0.5081	0.2135
	Vol	0.3424	0.1161	0.3074	0.1043	0.2968	0.1007	0.4134	0.1402	0.4948	0.1678	0.4040	0.1370
TKA	MD	1.3837	0.3695	1.6144	0.4311	1.5496	0.4138	1.8599	0.4966	2.1034	0.5617	1.8500	0.4940
	RV	0.5034	0.2030	0.3295	0.1329	0.2877	0.1160	0.6997	0.2821	2.0116	0.8111	0.8125	0.3276
	Vol	0.6130	0.2141	0.5481	0.1914	0.5319	0.1858	0.8340	0.2913	1.6551	0.5781	0.8726	0.3048
VW	MD	0.3403	0.1299	0.4329	0.1653	0.3836	0.1464	0.3707	0.1415	0.3791	0.1447	0.3667	0.1400
	RV	0.2101	0.1006	0.1693	0.0810	0.1543	0.0739	0.2550	0.1221	0.5189	0.2484	0.3184	0.1524
	Vol	0.9589	0.3712	1.0434	0.4039	0.7592	0.2939	0.7292	0.2823	1.0468	0.4052	0.8206	0.3176
mean criterion		0.9345	0.2664	0.9384	0.2636	0.9417	0.2645	1.0613	0.3017	1.3323	0.4084	1.0902	0.3154
mean rank		2.6280	2.6107	2.8228	2.8087	2.5220	2.5084	3.3457	3.3174	4.2227	4.1468	3.6579	3.6221

Table B4.8 – Point forecast evaluation for k=70

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE
ALV	MD	2.5655	0.6176	2.7897	0.6715	2.7971	0.6733	3.3547	0.8076	3.6807	0.8860	3.2847	0.7907
	RV	0.3611	0.1501	0.2386	0.0992	0.2316	0.0963	0.3550	0.1476	0.9872	0.4105	0.5599	0.2328
	Vol	0.5681	0.1832	0.4883	0.1575	0.4862	0.1568	0.7122	0.2297	0.9385	0.3026	0.7191	0.2319
BMW	MD	1.0697	0.2644	1.1286	0.2790	1.1500	0.2843	1.0617	0.2625	0.9620	0.2378	1.0793	0.2668
	RV	0.2905	0.1109	0.2519	0.0961	0.2469	0.0942	0.4569	0.1744	1.2534	0.4784	0.5869	0.2240
	Vol	0.4909	0.1620	0.4418	0.1458	0.4412	0.1456	0.5337	0.1762	0.6730	0.2221	0.5416	0.1788
DBK	MD	1.2119	0.3072	1.1415	0.2893	1.1495	0.2914	1.1243	0.2850	1.1297	0.2863	1.1271	0.2857
	RV	0.1830	0.0787	0.1913	0.0822	0.2089	0.0898	0.2106	0.0905	1.0957	0.4710	0.3522	0.1514
	Vol	1.3861	0.4439	1.4914	0.4776	1.4916	0.4777	1.4293	0.4577	1.2748	0.4083	1.3298	0.4259
SIE	MD	1.5088	0.3571	1.6542	0.3916	1.6656	0.3943	1.7173	0.4065	1.4117	0.3342	1.6340	0.3868
	RV	0.3257	0.1368	0.2403	0.1009	0.2384	0.1002	0.3892	0.1635	1.1686	0.4910	0.5318	0.2234
	Vol	0.3332	0.1130	0.3047	0.1034	0.2966	0.1006	0.3966	0.1345	0.4799	0.1628	0.3877	0.1315
TKA	MD	1.3609	0.3634	1.5107	0.4034	1.4739	0.3936	1.6956	0.4528	1.8965	0.5064	1.6968	0.4531
	RV	0.4970	0.2004	0.3477	0.1402	0.3134	0.1264	0.7000	0.2823	1.9532	0.7876	0.8288	0.3342
	Vol	0.6010	0.2099	0.5489	0.1917	0.5389	0.1882	0.7827	0.2734	1.5408	0.5382	0.8320	0.2906
VW	MD	0.3188	0.1217	0.4019	0.1534	0.3557	0.1358	0.3471	0.1325	0.4198	0.1603	0.3445	0.1315
	RV	0.2030	0.0972	0.1685	0.0807	0.1588	0.0760	0.2476	0.1185	0.5315	0.2545	0.3212	0.1538
	Vol	0.9372	0.3628	1.0388	0.4021	0.7535	0.2917	0.7224	0.2796	1.0364	0.4012	0.8113	0.3141
mean criterion		0.9030	0.2576	0.9070	0.2554	0.9114	0.2566	1.0265	0.2923	1.3328	0.4115	1.0645	0.3098
mean rank		2.7290	2.7111	2.6252	2.6096	2.7247	2.7094	3.1457	3.1174	4.2253	4.1478	3.7587	3.7224

Table B4.9 – Point forecast evaluation for k=80

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE
ALV	MD	2.7479	0.6615	2.9933	0.7206	2.9888	0.7195	3.6289	0.8736	4.0166	0.9669	3.5362	0.8512
	RV	0.3718	0.1546	0.2412	0.1003	0.2329	0.0968	0.3846	0.1599	1.0798	0.4490	0.6221	0.2587
	Vol	0.5867	0.1892	0.5004	0.1614	0.4958	0.1599	0.7472	0.2410	1.0005	0.3226	0.7533	0.2429
BMW	MD	1.1267	0.2785	1.1877	0.2936	1.2111	0.2994	1.0754	0.2659	0.9127	0.2256	1.0942	0.2705
	RV	0.2815	0.1075	0.2388	0.0911	0.2333	0.0890	0.4885	0.1865	1.3252	0.5058	0.6456	0.2464
	Vol	0.4874	0.1609	0.4299	0.1419	0.4285	0.1415	0.5506	0.1817	0.7129	0.2353	0.5619	0.1855
DBK	MD	1.2161	0.3082	1.1974	0.3035	1.1946	0.3028	1.2151	0.3080	1.2072	0.3060	1.2118	0.3072
	RV	0.1805	0.0776	0.1764	0.0758	0.1886	0.0811	0.2362	0.1016	1.2087	0.5196	0.4130	0.1775
	Vol	1.2948	0.4147	1.3798	0.4419	1.3802	0.4420	1.3313	0.4264	1.2084	0.3870	1.2476	0.3995
SIE	MD	1.5101	0.3575	1.6617	0.3934	1.6724	0.3959	1.6910	0.4003	1.3848	0.3278	1.5981	0.3783
	RV	0.3280	0.1378	0.2355	0.0990	0.2337	0.0982	0.4193	0.1762	1.2286	0.5162	0.5970	0.2508
	Vol	0.3169	0.1075	0.2880	0.0977	0.2798	0.0949	0.3960	0.1343	0.4848	0.1644	0.3841	0.1303
TKA	MD	1.4264	0.3809	1.5864	0.4236	1.5354	0.4100	1.8167	0.4851	2.0089	0.5364	1.8075	0.4827
	RV	0.4721	0.1904	0.3329	0.1342	0.2946	0.1188	0.7298	0.2943	1.9773	0.7973	0.8568	0.3455
	Vol	0.6015	0.2101	0.5424	0.1894	0.5278	0.1844	0.8309	0.2902	1.5762	0.5505	0.8756	0.3058
VW	MD	0.3071	0.1172	0.4067	0.1552	0.3509	0.1339	0.3296	0.1258	0.4394	0.1677	0.3268	0.1248
	RV	0.2016	0.0965	0.1638	0.0784	0.1497	0.0717	0.2658	0.1272	0.5801	0.2777	0.3603	0.1725
	Vol	0.9499	0.3677	1.0847	0.4199	0.7288	0.2821	0.6823	0.2641	1.0987	0.4253	0.8323	0.3222
mean criterion		0.9215	0.2614	0.9272	0.2589	0.9282	0.2591	1.0731	0.3049	1.4080	0.4353	1.1206	0.3266
mean rank		2.8281	2.8107	2.4239	2.4091	2.2233	2.2089	3.6489	3.6186	4.2325	4.1506	3.8646	3.8246

Table B4.10 – Point forecast evaluation for k=90

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE
ALV	MD	2.7949	0.6728	3.0398	0.7318	3.0363	0.7309	3.7081	0.8926	4.1119	0.9898	3.6262	0.8729
	RV	0.3766	0.1566	0.2441	0.1015	0.2350	0.0977	0.4082	0.1697	1.1403	0.4742	0.6842	0.2845
	Vol	0.5955	0.1920	0.5099	0.1644	0.5050	0.1629	0.7625	0.2459	1.0338	0.3334	0.7730	0.2493
BMW	MD	1.1865	0.2933	1.2459	0.3080	1.2704	0.3141	1.1059	0.2734	0.9051	0.2237	1.1209	0.2771
	RV	0.2807	0.1072	0.2374	0.0906	0.2315	0.0884	0.5148	0.1965	1.3858	0.5290	0.7004	0.2674
	Vol	0.4910	0.1621	0.4319	0.1426	0.4311	0.1423	0.5670	0.1872	0.7514	0.2480	0.5766	0.1903
DBK	MD	1.2909	0.3272	1.2973	0.3288	1.2911	0.3273	1.3346	0.3383	1.3279	0.3366	1.3272	0.3364
	RV	0.1808	0.0777	0.1721	0.0740	0.1820	0.0782	0.2589	0.1113	1.3575	0.5836	0.4724	0.2031
	Vol	1.2622	0.4042	1.3320	0.4266	1.3361	0.4279	1.2982	0.4158	1.2005	0.3845	1.2403	0.3972
SIE	MD	1.5369	0.3638	1.6846	0.3988	1.6961	0.4015	1.6922	0.4006	1.3899	0.3290	1.6106	0.3812
	RV	0.3411	0.1433	0.2459	0.1033	0.2438	0.1024	0.4507	0.1894	1.2966	0.5447	0.6598	0.2772
	Vol	0.3172	0.1076	0.2876	0.0975	0.2792	0.0947	0.4035	0.1369	0.4941	0.1676	0.3943	0.1338
TKA	MD	1.4792	0.3950	1.6451	0.4393	1.5877	0.4240	1.9055	0.5088	2.0796	0.5553	1.8996	0.5073
	RV	0.4451	0.1795	0.3160	0.1274	0.2779	0.1120	0.7533	0.3037	1.9826	0.7994	0.9122	0.3678
	Vol	0.6063	0.2118	0.5431	0.1897	0.5271	0.1841	0.8714	0.3044	1.6023	0.5597	0.9245	0.3229
VW	MD	0.2991	0.1142	0.4001	0.1527	0.3446	0.1315	0.3196	0.1220	0.4572	0.1745	0.3190	0.1218
	RV	0.1997	0.0956	0.1621	0.0776	0.1462	0.0700	0.2784	0.1333	0.6171	0.2954	0.3881	0.1858
	Vol	0.9458	0.3661	1.1072	0.4286	0.7158	0.2771	0.6678	0.2585	1.1373	0.4402	0.8112	0.3140
mean criterion		0.9399	0.2659	0.9456	0.2631	0.9465	0.2633	1.1065	0.3145	1.4682	0.4559	1.1690	0.3420
mean rank		2.3281	2.3107	2.5237	2.5091	2.3232	2.3088	3.7515	3.7196	4.4386	4.3529	3.8700	3.8267

Table B4.11 – Point forecast evaluation for k=100

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE	MASE	SMAE
ALV	MD	2.7299	0.6572	2.9244	0.7040	2.9237	0.7038	3.5174	0.8467	3.8941	0.9374	3.4353	0.8270
	RV	0.3930	0.1634	0.2701	0.1123	0.2633	0.1095	0.4336	0.1803	1.1501	0.4782	0.7241	0.3011
	Vol	0.6732	0.2171	0.6014	0.1940	0.5979	0.1928	0.8201	0.2645	1.0800	0.3483	0.8337	0.2689
BMW	MD	1.1389	0.2816	1.1945	0.2953	1.2195	0.3015	1.0622	0.2626	0.9507	0.2350	1.0811	0.2673
	RV	0.2705	0.1033	0.2327	0.0888	0.2294	0.0876	0.5229	0.1996	1.4280	0.5451	0.7245	0.2766
	Vol	0.4954	0.1635	0.4420	0.1459	0.4423	0.1460	0.5668	0.1871	0.7454	0.2461	0.5794	0.1913
DBK	MD	1.2595	0.3193	1.2749	0.3232	1.2635	0.3203	1.3334	0.3380	1.3194	0.3344	1.3177	0.3340
	RV	0.1770	0.0761	0.1654	0.0711	0.1754	0.0754	0.2776	0.1194	1.9952	0.8577	0.5167	0.2221
	Vol	1.2091	0.3872	1.2829	0.4108	1.2848	0.4115	1.2418	0.3977	1.1620	0.3721	1.1950	0.3827
SIE	MD	1.4778	0.3498	1.6211	0.3837	1.6342	0.3868	1.6133	0.3819	1.3322	0.3154	1.5263	0.3613
	RV	0.3322	0.1396	0.2335	0.0981	0.2327	0.0978	0.4674	0.1964	1.3401	0.5630	0.6986	0.2935
	Vol	0.3251	0.1103	0.3012	0.1022	0.2961	0.1004	0.4087	0.1386	0.4972	0.1686	0.4007	0.1359
TKA	MD	1.3799	0.3685	1.5369	0.4104	1.4837	0.3962	1.8273	0.4880	1.9864	0.5304	1.7999	0.4806
	RV	0.4092	0.1650	0.2920	0.1178	0.2585	0.1042	0.7660	0.3089	1.9826	0.7994	0.9408	0.3794
	Vol	0.5698	0.1990	0.5074	0.1772	0.4936	0.1724	0.8773	0.3064	1.5946	0.5570	0.9297	0.3247
VW	MD	0.2984	0.1139	0.3809	0.1454	0.3337	0.1274	0.3200	0.1221	0.5148	0.1965	0.3239	0.1237
	RV	0.1875	0.0898	0.1546	0.0740	0.1421	0.0680	0.2833	0.1356	0.6431	0.3079	0.4069	0.1948
	Vol	0.9130	0.3534	1.0777	0.4172	0.7186	0.2782	0.6843	0.2649	1.0943	0.4236	0.8018	0.3104
mean criterion		0.9274	0.2632	0.9320	0.2606	0.9333	0.2609	1.0862	0.3106	1.5250	0.4838	1.1564	0.3412
mean rank		2.4271	2.4103	2.4233	2.4089	2.3229	2.3088	3.7523	3.7200	4.4428	4.3545	3.8725	3.8277

Table B4.12 – Forecast interval evaluation for k=10

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi
ALV	MD	1.7053	1	1.6357	1	1.6915	0	1.6402	2	1.7018	2	1.6020	3
	RV	0.4963	1	0.4660	1	0.4417	1	0.4903	1	0.6208	0	0.5270	0
	Vol	0.6525	0	0.6578	2	0.6459	1	0.7476	3	0.7674	3	0.7483	3
BMW	MD	0.7056	0	0.6866	0	0.7085	0	0.7451	1	0.8812	1	0.7122	0
	RV	0.4619	1	0.4500	1	0.4274	1	0.5192	1	0.7044	0	0.5391	0
	Vol	0.6275	1	0.6258	1	0.6156	1	0.6870	1	0.6909	1	0.6892	1
DBK	MD	1.3359	1	1.2618	1	1.3321	1	1.3386	1	1.5314	0	1.2870	1
	RV	0.4090	0	0.3697	0	0.3408	0	0.4097	0	0.5903	0	0.4565	0
	Vol	1.0479	1	1.0013	1	1.0047	0	1.0307	0	1.1204	0	1.0722	0
SIE	MD	0.9429	0	0.8979	0	0.9374	0	0.9258	0	1.0035	0	0.8942	0
	RV	0.4466	1	0.4174	0	0.4017	1	0.4738	0	0.6395	1	0.5017	0
	Vol	0.5774	1	0.5693	1	0.5444	1	0.6168	1	0.6513	1	0.6261	1
TKA	MD	1.1201	1	1.0296	0	1.1132	1	1.0987	0	0.9675	3	1.0309	0
	RV	1.2830	7	0.7104	7	0.6596	0	0.7963	6	2.0868	9	0.8426	7
	Vol	1.1178	1	1.0158	1	1.0042	0	1.1105	2	1.6671	6	1.1127	2
VW	MD	0.4606	1	0.4180	0	0.4571	1	0.4881	1	0.4976	1	0.4671	1
	RV	0.4563	1	0.4506	0	0.4003	1	0.4662	0	0.5801	1	0.5215	1
	Vol	1.8228	1	1.9289	1	1.7368	1	1.7495	1	1.9435	1	1.8152	1
mean criterion		0.8269	0.6667	0.7950	0.8889	0.8009	0.5556	0.8454	1.1111	0.9565	0.7778	0.8482	0.8889
mean rank		3.1462	3.2500	1.6450	3.4500	1.7427	2.9000	3.5519	3.8500	5.3704	2.8500	3.7539	3

Table B4.13 – Forecast interval evaluation for k=20

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi
ALV	MD	1.7839	2	1.7053	6	1.7566	4	1.6919	7	1.8902	7	1.6625	8
	RV	0.5835	0	0.5196	0	0.4954	0	0.5992	0	0.8474	2	0.6741	2
	Vol	0.6696	4	0.6706	6	0.6583	1	0.7895	8	0.8176	8	0.7896	8
BMW	MD	0.7201	1	0.7002	2	0.7187	2	0.7772	1	1.0542	2	0.7375	2
	RV	0.4943	0	0.4734	0	0.4483	0	0.6099	1	0.9487	6	0.6625	2
	Vol	0.6398	1	0.6344	1	0.6243	1	0.7163	1	0.7295	1	0.7289	1
DBK	MD	1.3608	3	1.2724	3	1.3485	3	1.3663	3	1.6775	2	1.3109	3
	RV	0.4655	0	0.4019	2	0.3698	1	0.5029	0	0.8488	1	0.5763	1
	Vol	1.1180	3	1.0722	3	1.0761	2	1.1411	1	1.2815	2	1.2112	2
SIE	MD	1.0033	3	0.9519	3	0.9887	3	1.0034	3	1.1510	1	0.9764	3
	RV	0.5093	0	0.4570	2	0.4428	0	0.5761	2	0.8955	4	0.6281	4
	Vol	0.5719	2	0.5659	1	0.5390	2	0.6342	1	0.6634	1	0.6407	1
TKA	MD	1.2440	1	1.1493	0	1.2432	1	1.2089	0	1.0975	5	1.1375	0
	RV	1.0715	9	0.6936	10	0.6311	2	0.8545	12	2.3014	17	0.9142	14
	Vol	0.9998	1	0.9637	2	0.9493	1	1.0978	4	1.6410	13	1.0913	5
VW	MD	0.5139	0	0.4337	4	0.4888	0	0.5514	0	0.6026	1	0.5360	0
	RV	0.4598	0	0.4616	4	0.4105	0	0.5219	4	0.6546	7	0.5831	7
	Vol	1.8680	0	2.1473	2	1.8331	1	1.9036	1	2.1186	1	2.0412	1
mean criterion		0.8706	1.5556	0.8278	2.5556	0.8329	1.5556	0.9105	2.4444	1.1217	3.4444	0.9282	3.2222
mean rank		2.9494	2.3500	1.7473	3.4000	1.6448	2.6500	3.6610	2.8500	5.4949	4.3000	3.7662	4.2500

Table B4.14 – Forecast interval evaluation for k=30

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi
ALV	MD	1.8075	4	1.7318	7	1.7818	6	1.7233	10	1.9748	10	1.7027	12
	RV	0.6295	1	0.5493	1	0.5253	1	0.6974	1	1.0370	5	0.8106	3
	Vol	0.6729	7	0.6721	8	0.6623	3	0.8156	13	0.8528	14	0.8217	13
BMW	MD	0.7192	3	0.7009	4	0.7177	4	0.8002	1	1.1785	2	0.7678	3
	RV	0.5079	0	0.4849	0	0.4567	0	0.6931	2	1.1551	15	0.7729	5
	Vol	0.6439	1	0.6391	1	0.6271	1	0.7371	3	0.7640	3	0.7528	3
DBK	MD	1.3770	3	1.2813	3	1.3516	3	1.3856	3	1.7660	1	1.3523	3
	RV	0.4974	1	0.4199	1	0.3833	0	0.5896	1	1.0968	1	0.7104	0
	Vol	1.1528	5	1.1067	4	1.1096	4	1.2177	2	1.3854	3	1.3130	3
SIE	MD	1.0248	5	0.9710	6	1.0076	5	1.0604	4	1.2352	1	1.0412	3
	RV	0.5431	0	0.4820	2	0.4654	0	0.6695	2	1.1280	9	0.7618	7
	Vol	0.5670	2	0.5602	2	0.5342	2	0.6472	0	0.6716	0	0.6545	0
TKA	MD	1.2978	1	1.2045	1	1.3036	1	1.2541	2	1.1820	7	1.1935	2
	RV	0.9399	8	0.6721	12	0.6064	1	0.9229	17	2.3962	24	1.0093	20
	Vol	0.9517	4	0.9339	4	0.9198	2	1.1097	7	1.6162	18	1.1207	8
VW	MD	0.5421	1	0.4413	4	0.5049	1	0.6072	1	0.7014	2	0.5954	1
	RV	0.4681	1	0.4726	3	0.4155	0	0.5771	4	0.7348	12	0.6634	10
	Vol	1.9021	1	2.2906	1	1.8869	2	2.0255	2	2.2847	2	2.2343	2
mean criterion		0.8898	2.7778	0.8429	3.2222	0.8461	2.4444	0.9622	4	1.2456	6	1.0004	5
mean rank		3.0508	2.7500	1.6485	3.1000	1.5457	2.5000	3.6693	3.3000	5.5155	5.2500	3.8773	4.2000

Table B4.15 – Forecast interval evaluation for k=40

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi
ALV	MD	1.8156	4	1.7394	8	1.7938	6	1.7472	13	1.9891	12	1.7261	15
	RV	0.6620	2	0.5690	2	0.5433	2	0.7879	2	1.2036	5	0.9493	2
	Vol	0.6765	8	0.6745	9	0.6642	5	0.8371	17	0.8877	19	0.8445	17
BMW	MD	0.7172	6	0.6986	6	0.7138	6	0.8205	0	1.2715	4	0.7874	3
	RV	0.5139	1	0.4878	1	0.4600	1	0.7709	2	1.3241	22	0.8729	7
	Vol	0.6455	1	0.6390	1	0.6284	1	0.7555	3	0.7898	5	0.7729	5
DBK	MD	1.3836	3	1.2854	3	1.3533	3	1.4027	2	1.7847	0	1.3639	2
	RV	0.5266	2	0.4356	1	0.3895	1	0.6718	2	1.3630	0	0.8406	1
	Vol	1.1693	5	1.1290	5	1.1282	4	1.2736	2	1.4384	4	1.4020	3
SIE	MD	1.0354	5	0.9801	6	1.0149	6	1.1053	4	1.2919	1	1.0883	3
	RV	0.5597	1	0.4932	2	0.4793	0	0.7570	2	1.3094	17	0.8965	10
	Vol	0.5652	2	0.5574	1	0.5305	2	0.6589	0	0.6832	0	0.6638	0
TKA	MD	1.3255	1	1.2349	4	1.3376	1	1.2809	4	1.2328	10	1.2203	4
	RV	0.8578	7	0.6539	10	0.5857	0	0.9898	20	2.4373	33	1.1134	27
	Vol	0.9240	6	0.9125	6	0.9011	4	1.1268	9	1.6009	21	1.1451	11
VW	MD	0.5607	2	0.4443	3	0.5143	1	0.6597	2	0.8010	3	0.6546	2
	RV	0.4682	0	0.4748	2	0.4173	1	0.6314	3	0.7954	12	0.7383	11
	Vol	1.9199	1	2.3932	1	1.9184	3	2.1217	3	2.4266	3	2.3679	3
mean criterion		0.9011	3.5556	0.8509	4	0.8527	3.2222	1.0075	4.7778	1.3391	7.8889	1.0622	6.1111
mean rank		3.0514	3.2000	1.6488	3.2500	1.4460	2.7500	3.7771	3.2500	5.5324	5.8000	3.8873	4.0500

Table B4.16 – Forecast interval evaluation for k=50

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi
ALV	MD	1.8279	3	1.7548	8	1.8002	5	1.7666	12	2.0103	10	1.7660	13
	RV	0.6821	3	0.5813	3	0.5547	3	0.8716	3	1.3387	4	1.0942	2
	Vol	0.6791	8	0.6764	9	0.6653	4	0.8564	16	0.9192	18	0.8659	16
BMW	MD	0.7118	5	0.6934	7	0.7088	6	0.8393	1	1.3599	4	0.8037	2
	RV	0.5143	2	0.4884	2	0.4610	2	0.8442	1	1.4703	26	0.9910	6
	Vol	0.6454	0	0.6394	0	0.6292	0	0.7727	3	0.8161	6	0.7953	6
DBK	MD	1.3866	2	1.2868	2	1.3551	2	1.4184	1	1.8036	1	1.3855	1
	RV	0.5374	3	0.4410	2	0.3920	2	0.7502	3	1.5671	1	0.9810	2
	Vol	1.1811	6	1.1419	6	1.1399	5	1.3158	1	1.4790	2	1.4591	2
SIE	MD	1.0403	5	0.9849	7	1.0170	6	1.1416	3	1.3386	0	1.1420	2
	RV	0.5745	1	0.5052	1	0.4884	1	0.8390	1	1.4892	18	1.0424	12
	Vol	0.5628	2	0.5539	1	0.5275	2	0.6700	0	0.6932	0	0.6801	0
TKA	MD	1.3431	1	1.2544	5	1.3595	1	1.2999	6	1.2707	12	1.2421	6
	RV	0.8068	6	0.6393	10	0.5684	1	1.0533	24	2.4551	42	1.2042	34
	Vol	0.9062	8	0.9003	8	0.8876	6	1.1448	13	1.5941	26	1.1693	15
VW	MD	0.5716	3	0.4442	4	0.5202	2	0.7102	3	0.8900	4	0.7036	3
	RV	0.4703	1	0.4779	1	0.4174	2	0.6847	2	0.8626	13	0.8200	11
	Vol	1.9336	2	2.4726	1	1.9378	4	2.1983	4	2.5182	4	2.4975	4
mean criterion		0.9073	3.5556	0.8559	4.3333	0.8562	3.2222	1.0483	4.5556	1.4183	8	1.1268	5.5556
mean rank		3.0514	3.3500	1.5488	3.6000	1.4461	3.0500	3.7844	2.8500	5.5470	6.2000	3.9991	3.7500

Table B4.17 – Forecast interval evaluation for k=60

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi
ALV	MD	1.8315	4	1.7574	7	1.8037	5	1.7827	12	2.0100	10	1.7830	13
	RV	0.6942	3	0.5879	3	0.5622	3	0.9489	4	1.4503	5	1.2263	1
	Vol	0.6795	7	0.6757	9	0.6660	4	0.8745	17	0.9462	20	0.8854	17
BMW	MD	0.7075	4	0.6901	6	0.7033	6	0.8572	1	1.4296	4	0.8163	3
	RV	0.5142	2	0.4878	2	0.4608	2	0.9132	0	1.5718	25	1.0988	5
	Vol	0.6452	0	0.6388	0	0.6297	0	0.7892	2	0.8410	5	0.8113	5
DBK	MD	1.3952	1	1.2913	1	1.3571	1	1.4331	0	1.8163	3	1.3863	0
	RV	0.5407	4	0.4441	3	0.3924	2	0.8246	4	1.7643	2	1.1024	3
	Vol	1.1923	7	1.1515	6	1.1479	6	1.3483	0	1.5086	1	1.4889	1
SIE	MD	1.0409	7	0.9841	8	1.0165	7	1.1715	2	1.3725	0	1.1864	3
	RV	0.5787	2	0.5093	1	0.4949	1	0.9160	0	1.5991	20	1.1470	12
	Vol	0.5584	2	0.5501	2	0.5249	2	0.6809	1	0.7015	1	0.6907	1
TKA	MD	1.3561	0	1.2678	4	1.3753	0	1.3148	6	1.2965	13	1.2499	6
	RV	0.7675	6	0.6204	11	0.5539	1	1.1129	25	2.4396	48	1.2839	37
	Vol	0.8931	8	0.8895	8	0.8769	5	1.1626	14	1.5890	30	1.1834	17
VW	MD	0.5767	4	0.4451	3	0.5241	3	0.7589	4	0.9696	5	0.7544	4
	RV	0.4721	1	0.4801	1	0.4165	2	0.7368	1	0.9321	12	0.8832	10
	Vol	1.9448	2	2.5369	0	1.9505	5	2.2597	5	2.6008	5	2.6100	5
mean criterion		0.9111	3.5556	0.8583	4.1111	0.8581	3.2222	1.0857	4.4444	1.4820	8.3333	1.1776	5.3333
mean rank		3.0514	3.2000	1.5488	3.3500	1.4461	2.8500	3.6913	2.8500	5.5572	6.6000	4.1099	3.6500

Table B4.18 – Forecast interval evaluation for k=70

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi
ALV	MD	1.8341	5	1.7596	8	1.8056	6	1.7961	13	2.0168	11	1.7967	14
	RV	0.7022	4	0.5931	4	0.5671	4	1.0204	5	1.5508	4	1.3588	0
	Vol	0.6804	7	0.6759	10	0.6666	5	0.8916	18	0.9715	22	0.9067	18
BMW	MD	0.6999	7	0.6839	8	0.6975	7	0.8744	2	1.4806	6	0.8341	1
	RV	0.5149	2	0.4885	2	0.4601	2	0.9780	1	1.6782	26	1.1798	7
	Vol	0.6462	1	0.6405	1	0.6301	1	0.8050	2	0.8670	6	0.8265	6
DBK	MD	1.3948	0	1.2920	0	1.3593	0	1.4469	1	1.8073	4	1.4233	1
	RV	0.5434	5	0.4444	4	0.3915	3	0.8951	5	1.9269	3	1.2226	4
	Vol	1.1954	5	1.1573	5	1.1536	5	1.3739	1	1.5247	0	1.5292	0
SIE	MD	1.0376	8	0.9806	10	1.0145	8	1.1964	1	1.3889	1	1.2192	2
	RV	0.5874	1	0.5151	1	0.4996	2	0.9880	0	1.7388	22	1.2786	14
	Vol	0.5574	3	0.5483	3	0.5225	3	0.6915	2	0.7138	1	0.7022	2
TKA	MD	1.3701	1	1.2818	5	1.3874	1	1.3272	5	1.3250	10	1.2657	6
	RV	0.7390	6	0.6104	10	0.5414	0	1.1687	24	2.4805	52	1.3968	40
	Vol	0.8846	7	0.8821	7	0.8678	4	1.1797	13	1.5903	29	1.2283	16
VW	MD	0.5815	5	0.4451	2	0.5267	4	0.8062	5	1.0546	6	0.8179	5
	RV	0.4752	3	0.4819	0	0.4151	3	0.7877	0	0.9969	12	0.9803	8
	Vol	1.9496	4	2.5763	1	1.9590	6	2.3093	6	2.6787	6	2.6864	6
mean criterion		0.9123	4	0.8595	4.6667	0.8590	3.6667	1.1202	5.3333	1.5360	9.1111	1.2309	5.6667
mean rank		2.9515	3.0500	1.5489	3.3000	1.4460	2.6500	3.6978	3.6500	5.4678	6.3000	4.3180	3.9500

Table B4.19 – Forecast interval evaluation for k=80

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi
ALV	MD	1.8327	7	1.7594	11	1.8064	9	1.8074	18	1.9876	17	1.8111	17
	RV	0.7100	5	0.5967	5	0.5704	5	1.0864	6	1.6463	8	1.4745	6
	Vol	0.6813	11	0.6773	12	0.6670	6	0.9081	22	1.0029	25	0.9167	23
BMW	MD	0.6936	11	0.6784	11	0.6915	11	0.8909	3	1.5389	7	0.8512	2
	RV	0.5123	3	0.4864	3	0.4591	3	1.0390	2	1.7644	30	1.2790	8
	Vol	0.6458	1	0.6400	0	0.6304	2	0.8204	2	0.8883	6	0.8399	6
DBK	MD	1.3980	1	1.2947	1	1.3616	1	1.4599	2	1.7992	4	1.4328	1
	RV	0.5454	6	0.4450	5	0.3899	4	0.9618	6	2.0898	4	1.3687	5
	Vol	1.2005	5	1.1630	4	1.1581	4	1.3945	2	1.5380	1	1.5624	1
SIE	MD	1.0331	10	0.9778	12	1.0114	10	1.2174	1	1.4084	2	1.2322	2
	RV	0.5916	0	0.5187	0	0.5033	2	1.0553	1	1.8288	27	1.3967	18
	Vol	0.5542	4	0.5450	3	0.5202	4	0.7019	2	0.7181	1	0.7059	2
TKA	MD	1.3776	2	1.2901	7	1.3974	2	1.3378	6	1.3347	11	1.2871	7
	RV	0.7146	5	0.6015	10	0.5305	1	1.2206	30	2.4903	60	1.4532	47
	Vol	0.8740	8	0.8751	8	0.8598	5	1.1959	17	1.5844	35	1.2317	20
VW	MD	0.5880	6	0.4462	1	0.5285	5	0.8522	6	1.1681	7	0.8707	6
	RV	0.4744	4	0.4813	1	0.4133	4	0.8373	1	1.0606	14	1.0607	12
18	1.9620	5	2.6152	1	1.9650	7	2.3496	7	2.7777	7	2.8266	7	
mean criterion		0.9133	5.5556	0.8601	5.7778	0.8594	5	1.1520	7	1.5839	11.3333	1.2818	7.6667
mean rank		2.9512	3.2000	1.5486	3.0500	1.2459	2.8000	3.8039	3.6500	5.4764	7	4.4279	4.1000

Table B4.20 – Forecast interval evaluation for k=90

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi
ALV	MD	1.8347	8	1.7629	12	1.8065	10	1.8169	21	1.9782	20	1.8321	19
	RV	0.7149	6	0.6000	6	0.5726	6	1.1472	7	1.7173	13	1.6134	12
	Vol	0.6821	14	0.6773	14	0.6673	9	0.9239	26	1.0257	30	0.9432	26
BMW	MD	0.6867	15	0.6725	16	0.6855	15	0.9069	4	1.5656	7	0.8701	0
	RV	0.5126	4	0.4863	4	0.4578	3	1.0962	2	1.8342	38	1.3682	16
	Vol	0.6468	2	0.6404	0	0.6307	1	0.8354	5	0.9101	10	0.8502	9
DBK	MD	1.3985	0	1.2949	2	1.3640	2	1.4723	3	1.7866	5	1.4556	3
	RV	0.5440	7	0.4445	6	0.3878	5	1.0248	7	2.2454	2	1.4954	2
	Vol	1.2047	5	1.1672	4	1.1616	4	1.4113	3	1.5517	2	1.5872	2
SIE	MD	1.0304	11	0.9757	13	1.0078	11	1.2352	0	1.4217	3	1.2448	1
	RV	0.5974	0	0.5242	1	0.5062	2	1.1181	3	1.9407	34	1.5142	25
	Vol	0.5527	5	0.5432	5	0.5181	5	0.7122	2	0.7295	1	0.7177	2
TKA	MD	1.3845	3	1.2980	9	1.4058	3	1.3472	9	1.3500	17	1.2917	10
	RV	0.6958	4	0.5907	8	0.5207	2	1.2690	35	2.5006	66	1.5348	55
	Vol	0.8650	10	0.8684	11	0.8525	6	1.2112	23	1.5815	41	1.2502	27
VW	MD	0.5889	7	0.4451	1	0.5298	6	0.8970	7	1.2400	8	0.9235	7
	RV	0.4738	5	0.4808	2	0.4114	5	0.8856	2	1.1211	19	1.1394	14
	Vol	1.9698	6	2.6500	1	1.9692	8	2.3826	8	2.8416	8	2.8492	8
mean criterion		0.9139	6.7778	0.8607	7.1111	0.8593	6.1111	1.1816	8.6667	1.6239	14.1111	1.3350	9.8889
mean rank		2.9513	3.3000	1.5486	3.3000	1.2458	2.6500	3.8096	3.6500	5.4834	7.9000	4.4368	4.8000

Table B4.21 – Forecast interval evaluation for k=100

		SemiACD BS		Semi LogACD BS		SemiARMA		ARMA		ACD BS		LogACD BS	
		length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi	length CI	PoFi
ALV	MD	1.8366	9	1.7622	13	1.8061	11	1.8249	21	1.9718	19	1.8549	20
	RV	0.7170	6	0.6014	6	0.5739	6	1.2032	8	1.7688	13	1.7305	12
	Vol	0.6826	18	0.6777	17	0.6675	11	0.9391	26	1.0497	31	0.9675	26
BMW	MD	0.6795	15	0.6663	16	0.6795	16	0.9224	5	1.6083	9	0.8815	0
	RV	0.5115	5	0.4844	5	0.4565	4	1.1498	2	1.8936	37	1.4279	15
	Vol	0.6473	1	0.6410	1	0.6309	1	0.8501	4	0.9309	9	0.8684	9
DBK	MD	1.4019	0	1.2984	2	1.3664	2	1.4840	3	1.7904	5	1.4762	2
	RV	0.5424	8	0.4420	7	0.3854	6	1.0841	8	2.3545	2	1.5969	1
	Vol	1.2075	4	1.1701	4	1.1644	4	1.4252	4	1.5588	3	1.6103	3
SIE	MD	1.0256	11	0.9717	14	1.0038	11	1.2505	1	1.4329	3	1.2782	0
	RV	0.5990	1	0.5238	1	0.5086	2	1.1766	3	1.9830	36	1.6161	26
	Vol	0.5499	5	0.5400	4	0.5160	5	0.7223	2	0.7378	2	0.7310	2
TKA	MD	1.3909	2	1.3044	7	1.4131	2	1.3555	8	1.3600	15	1.3021	10
	RV	0.6762	3	0.5814	8	0.5118	3	1.3141	37	2.4975	74	1.5891	60
	Vol	0.8571	8	0.8616	11	0.8457	5	1.2255	23	1.5830	44	1.2572	26
VW	MD	0.5897	8	0.4442	0	0.5306	7	0.9406	8	1.3339	9	0.9690	8
	RV	0.4717	5	0.4780	3	0.4093	6	0.9325	3	1.1717	19	1.2029	15
	Vol	1.9676	7	2.6652	2	1.9723	9	2.4099	9	2.8629	9	2.9133	9
mean criterion		0.9140	7.3333	0.8604	7.8889	0.8589	6.7778	1.2092	9	1.6585	14.2222	1.3794	9.7778
mean rank		2.7512	3.1500	1.5484	3.4500	1.3456	2.9000	3.8150	3.8500	5.4894	7.7000	4.5428	4.6500

Supplementary Appendix of Chapter 5

Table B5.1 – RMSSE of all methods for all countries with mean and standard deviations

	Brazil	India	Mexico	China	S. Africa	Nigeria	USA	UK	mean	s.e.
LLR	0.771	0.542	0.427	0.366	1.643	0.384	0.929	1.123	0.773	0.446
RCS	0.708	1.371	0.471	0.713	0.623	0.612	0.993	0.643	0.767	0.285
RSL	1.309	0.291	1.505	0.434	0.780	0.188	0.067	0.217	0.599	0.545
RLL	0.629	0.442	0.315	0.054	0.518	0.385	0.645	0.650	0.455	0.205
$C2_1$	1.005	0.361	0.580	0.400	0.473	0.278	0.476	0.658	0.529	0.227
$C2_2$	0.559	0.754	0.604	0.151	0.176	0.386	0.507	0.415	0.444	0.207
$C2_3$	0.817	0.316	1.351	0.563	0.232	0.346	0.785	0.963	0.672	0.382
$C2_4$	1.309	0.291	1.505	0.434	0.780	0.188	0.067	0.217	0.599	0.545
$C2_5$	0.615	0.485	0.436	0.530	1.366	0.241	0.357	0.591	0.578	0.341
$C2_6$	0.488	0.429	1.199	0.693	1.096	0.764	1.579	2.037	1.036	0.559
$C3_1$	0.520	0.553	0.436	0.178	1.129	0.497	0.958	0.880	0.644	0.315
$C3_2$	0.526	0.461	0.797	0.148	0.348	0.146	0.201	0.463	0.386	0.224
$C3_3$	0.504	0.363	0.248	0.124	1.117	0.158	0.148	0.299	0.370	0.327
$C3_4$	0.506	0.588	0.468	0.060	0.859	0.166	0.300	0.730	0.460	0.273
$C4$	0.586	0.341	0.518	0.200	0.659	0.164	0.123	0.253	0.355	0.206

Ehrenwörtliche Erklärung

Hiermit versichere ich, durch eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbstständig und ohne unerlaubte Hilfe Dritter angefertigt habe. Alle Stellen, die inhaltlich oder wörtlich aus Veröffentlichungen stammen, sind kenntlich gemacht. Diese Arbeit lag nach meinem Informationsstand in gleicher oder ähnlicher Weise noch keiner Prüfungsbehörde vor und wurde bisher nicht veröffentlicht.

Paderborn, den 02.Juli 2018

(Sarah Forstinger)