CONTRIBUTIONS TO STATISTICAL MODELLING OF HIGH-FREQUENCY FINANCIAL DATA WITH APPLICATIONS TO FRANKFURT STOCK EXCHANGE

Dissertation
for the attainment of the degree of Doctor of Philosophy
at the Faculty of Economics and Business Administration
of the Georg-August-Universität Göttingen

presented by

Ta-Chao Kao

born in Taipei, Taiwan

Göttingen, 2011

First examiner:	Professor Dr. Walter Zucchini
Second examiner:	Professor Dr. Stefan Sperlich
Third examiner:	Professor Dr. Wolfgang Benner
Date of oral examination:	September 28, 2011

To my loving family ...

CONTENTS

iv

LIST OF FIGURES

LIST OF TABLES

NOTATION AND ABBREVIATIONS

Notation

$\bar{A}_d$	mean level of deseasonalised durations on day d
C_n or C_t	state occupied by Markov chain at $n-$th price event, or at time t
$\mathbf{C}^{(n)}$ or $\mathbf{C}^{(t)}$	$(C_1, C_2, \ldots, C_n)$, or $(C_1, C_2, \ldots, C_t)$
$e_{n,j}$	random variation of Z_n in state j
$f_{n\mid n-1,j}(\cdot)$	conditional distribution of Z_n given history in state j
$g_j(\cdot)$	distribution of e_n in state j
k	number of parameters of a model, or lagged term in model for μ_n
K	number of price events in dataset
l or l_N or l_T	log-likelihood
L or L_N or L_T	likelihood
$\log$	logarithm to the base e
M	minimum level for significant price change
M_t	conditional return given price change at time t
$\mathbf{M}^{(t)}$	$(M_1, M_2, \ldots, M_t)$
m	number of states in a Markov chain

$\mathbf{m}^{(t)}$ $(m_1, m_2, \ldots, m_t)$

$m_d(t)$ seasonal component of observed durations on day d

N number of trades contained in a dataset

N_t predicted value of V_{t+1} in exponential filter

$N(\mu, \sigma^2)$ normal distribution with mean μ and variance σ^2

n_k $k-$th price event in dataset

P_t share price at time t

$\mathbf{P}(x)$ diagonal matrix with $i-$th diagonal element $p_i(x)$

p order of lagged Z_n terms,

       or number of parameters requiring estimation

$p_i(\cdot)$ probability mass or density function in state i

p_n share price $n-$th transaction

$p_{n|n-1,j}(\cdot)$ conditional probability of C_n in state j

       given observation history

q order of lagged μ_n terms

R_t return at time t

r_t observed return at time t

T	length of an observed time series
t_0	opening time for trading on first day of dataset
$U(0,1)$	standard uniform distribution
u_n	forecast uniform pseudo-residual of $n-$th observation
V_t	log-volume traded at time t
v_n	volume traded of $n-$th transaction
w_t	sum of forward probabilities at time t
W_n	trade duration of $n-$th transaction
$W_{k,d}^p(M)$	$k-$th observed $M-$price duration on day d of dataset
X_t	indicator for price change at time t
$\mathbf{X}^{(t)}$	$(X_1, X_2, \ldots, X_t)$
$\mathbf{x}^{(t)}$	$(x_1, x_2, \ldots, x_t)$
$Z_{k,d}$	$k-$th standardised price duration on day d of dataset
$\mathbf{Z}^{(n)}$	$(Z_1, Z_2, \ldots, Z_n)$
$\mathbf{z}^{(n)}$	$(z_1, z_2, \ldots, z_n)$
α	coefficient of lagged Z_n term in model for μ_n,
	or weight for observation used in exponential filter

$\boldsymbol{\alpha}_t$	(row) vector of forward probabilities
α_{0i}	constant term of model for logit of probability in state i, or logarithm of variance in state i
α_{1i}	coefficient of predicted volume term of model for logit of probability, or logarithm of variance in state i
α_{jk}	coefficient of $k-$th lagged Z_n term of model for μ_n in state j
β_{jl}	coefficient of $l-$th lagged μ_n term of model for μ_n in state j
Γ or $\Gamma(\cdot)$	transition probability matrix of Markov chain, or gamma function
γ_{ij}	$(i,j)-$th element of Γ; probability of transition from state i to state j in a Markov chain
$\boldsymbol{\delta}$	stationary distribution of Markov chain
$\boldsymbol{\theta}$	vector containing $\boldsymbol{\theta}_\Gamma$ and $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_m)$; parameters of Markov-switching autoregressive conditional duration model
$\boldsymbol{\theta}_j$	vector containing parameters of autoregressive conditional duration model in state j
$\boldsymbol{\theta}_\Gamma$	vector containing parameters of Markov chain
$\left(\kappa_j, \sigma_j^2\right)$	parameters of 'scaled' Burr distribution in state j
$\mu_{n,j}$	model for conditional mean of Z_n in state j
$\nu(\cdot)$	mean function of log-duration
$\pi_{t,i}$	probability of price change at time in state j

$\sigma^2_{t,i}$ variance of conditional return given change at time t in state j

τ_n time of $n-$th transaction

τ^p_k time of $k-$th price event

$\Phi\left(\cdot\right)$ distribution function of standard normal

ϕ_n or ϕ_t "scaled" μ_n,

      or 'scaled' forward probabilities

ω_j constant term of model for μ_n in state j

$\mathbf{1}$ (row) vector of ones

Abbreviations

ACD	autoregressive conditional duration
ACF	autocorrelation function
AIC	Akaike's information criterion: $2 \cdot (-l + k)$
BIC	Bayesian information criterion: $2(-l) + (k \cdot \log(T))$
DNM	direct numerical maximisation
d.o.f.	degrees of freedom
HMM	hidden Markov model
MC	Markov chain
MLE	maximum likelihood estimator or estimation
MSACD	Markov-switching autoregressive conditional duration
MSFE	mean squared forecast error
MSFE_{exp}	mean squared forecast error of exponential filter
MSFE_{HMM}	mean squared forecast error of hidden Markov model
NPR	normal pseudo-residual
qq-plot	quantile-quantile plot
sm-M	'small-sized' category of minimum level for significant price change
t.p.m.	transition probability matrix
UPR	uniform pseudo-residual
XETRA	*exchange electronic trading* system of Frankfurt Stock Exchange

ACKNOWLEDGEMENTS

This doctoral dissertation was completed during my stay at the Institute for Statistics and Econometrics. In this duration I was also enrolled in the Ph.D. program "Applied Statistics and Empirical Methods." The former is of the Faculty of Economics and Business Administration and the latter is offered by the Centre for Statistics, both of which are at the Georg-August University Göttingen. I gratefully acknowledge the German Academic Exchange Service for its financial support.

It gives me great pleasure to acknowledge the assistance of the people and institutions that have made it possible to complete this manuscript.

First and foremost, I wish to thank my dissertation advisor, Prof. Dr. Walter Zucchini, for his academic guidance, constant support, encouragement and patience throughout the period of his supervision. I have benefitted greatly from his expertise on a wide range of topics in applied statistics, particularly his insight into the hidden Markov models - the central methodology used in this project. His enthusiasm for and meticulous approach to research shall influence me for life. I thank also Prof. Dr. Stefan Sperlich and Prof. Dr. Wolfgang Benner for acting as my second and third examiners respectively. My special thanks go to Prof. Dr. Manfred Denker for serving on my Ph.D. committee and for his encouragement. After his departure from Göttingen I thank Prof. Dr. Martin Schlather for taking over this duty.

I am indebted to Günther Kratz for contributing his computing expertise at the initial stage of this project. Dr. Mikhail Langovoy, Dr. Jan Bulla and Dr. Roland Langrock all scrutinized the mathematical parts in earlier drafts of this manuscript and provided constructive criticisms, to which I am most grateful.

I wish to thank the Institute for Statistics and Econometrics for providing the working environment and to its members over the years for providing help in various ways and a friendly atmosphere. In particular, I thank my "officemate" Ren Ohinata for helpful conversations, Daniel Adler for sharing his experience with software development and Dr. Oleg Nenadić for various help.

I am most grateful to Prof. Renkuan Guo of the University of Cape Town for encouraging me to pursue Ph.D. studies in Göttingen. To commence my studies here I wish to express my deepest gratitude to Mr. Chen Tsan-Mao, my second uncle, for providing the financial means and for his encouragement throughout; and to Mr. Chen Sheng-Jung, my fourth uncle, for providing the required computing equipment.

Finally, I reserve my heartfelt gratitude to my family - my parents Peng-Chou and Chung-Ming, and my sister Chien-Hui for their unconditional love, constant support and guidance throughout my life. I am especially indebted to my parents for the sacrifices they made and to my sister for financing the closing stages of this project. Without them I could not have completed this work. This dissertation is dedicated to them.

CHAPTER 1

INTRODUCTION

Shephard (1996) pointed out the importance of studying the volatility models for the financial markets. The availability of high quality transaction data at an affordable cost makes it possible to investigate the relationship between the intra-daily price movements and the corresponding trading volume (see Engle and Russell, 1998). The salient features of the transaction data are, firstly, that the transactions are recorded at irregular time intervals over the course of the trading day and, secondly, the price changes between consecutive transactions can have only a limited number of values.

In order to account for the feature of irregular time intervals, Engle and Russell (1998) considered the time at which a transaction is recorded as a random variable. They then proposed the class of autoregressive conditional duration (ACD) models as a time series model for the sequence of times elapsed between consecutive trades. Engle and Russell (1998) also considered each variable (such as the transaction price and the trading volume) attached to a transaction time as a random variable. However, the changes in the successive transaction prices from their dataset have a negative sample autocorrelation at lag one[1]. As a result, Engle and Russell

---

[1]The cause of this phenomenon is explained, for example, in Bauwens and Giot (2001, Section 2.7).

(1998) used instead the midquotes[2] as the price variable to define the sequence of *quote times* elapsed between consecutive significant[3] price changes. They also showed that, conditional on the history, the intensity function for significant price change can be used to construct a measure of instantaneous return volatility (See Engle and Russell, 1998, p. 1154, Equation (32)).

A number of modifications to the original ACD model of Engle and Russell (1998) have been proposed in the literature. For a survey see, for example, Hautsch (2004, Section 5.3). The modifications may be grouped broadly into the specification of the conditional mean function and the specification of the error distribution. In particular, Hautsch (2004, Section 5.5.1) evaluated seven different specifications of the expected duration function of the ACD model and concluded that a nonlinear mean function provided a better description of the durations between consecutive significant midquote change than a linear mean function. Bauwens et al. (2004) applied a serially independent unit exponential model and thirteen separate specifications of the ACD model[4] to the waiting time between successive midquote changes of at least one-eighth of an US dollar. They compared the models with each other on the basis of their forecast performance, and

---

[2] The midquote is the average of the best bid quote (the price at which a buyer offers to buy an unit of the share) and the best ask quote (the price at which a seller offers to sell an unit of the share). The data used by Engle and Russell (1998) is obtained from trading on the New York Stock Exchange (NYSE). It contains, in addition to the trade data, the best bid quote, the best ask quote and the time at which the quotes are recorded. See paragraph one of Hautsch (2004, Section 4.2.1) for a discussion of data obtained from the NYSE.

[3] This means that the absolute value of the price change is at least a specified level - see later.

[4] The specifications used there selected from four types of model for the conditional mean, four families of distributions for the error process, and a class of models for duration expressed as a ratio of the quantiles of two gamma distributions.

concluded that the distribution for error in all the successful models have more parameters than the exponential and Weibull.

The transaction datasets analysed in this thesis contain only the following information for each transaction recorded: the time at which the transaction is recorded, the price at which the share is traded and the volume of the share traded. As a result, the price durations considered in this thesis are all defined as the time elapsed between consecutive significant *transaction* price changes. A financial duration model may, of course, be proposed for datasets that contain more information (for example, the bid and the ask quote) than are available in the dataset used in this thesis.

Hujer et al. (2002) proposed to model the durations between consecutive trades by allowing for some unobserved heterogeneity in the financial duration models. They reported that the models that allow for unobserved heterogeneity provided better forecasts than those provided by the ACD models that do not allow for unobserved heterogeneity. Hujer et al. (2002), however, did not apply their model to the more important variable of duration between significant price changes and they did not introduce covariates (such as the trading volume over the duration between significant price changes) to explain variations in the response variable. De Luca and Zuccolotto (2003) proposed a mixture of exponential distributions for the error of the ACD model. They reported that this model is adequate for the waiting time between successive price changes while the expo-

nential and the Weibull models considered by Engle and Russell (1998) disagree with the tail of the distribution of error. De Luca and Gallo (2004) reported that the agreement between the mixture of two exponentials and the error for the ACD model was not affected by introducing the volume of a day's first trade into the model for the conditional mean. De Luca and Gallo (2009) modelled the waiting time between successive price changes of at least one-sixteenth of an US dollar. They introduced covariate information into the model of De Luca and Zuccolotto (2003) in the following ways. Firstly, the conditional mean of the ACD model was allowed to depend on lagged trading intensity and lagged average volume. Secondly, the mixing proportion for the first exponential component was allowed to depend on lagged mixing proportion and lagged trading intensity. They reported that the model that incorporates covariates via the conditional mean fitted the waiting times better than that which does not. A comparison between the model that allows for covariates in the mixing proportion and that with a time-dependent mixing proportion is, however, not reported in De Luca and Gallo (2009).

Russell and Engle (2005) studied the relationship between the speed of the transaction price change and the size of the transaction price change. They considered the sequence of vectors containing the duration and the price change between consecutive transactions as a bivariate time series. The distribution of the bivariate vector, conditioned on the preceding observed vectors, is factorised[5] into a product of two component distributions: the distribution of price change, condi-

---

[5]This method was proposed by Engle (2000, p. 3, Equation (5)).

tioned on the duration, and the marginal distribution of the duration. Russell and Engle (2005) considered the price change between consecutive trades as a multinomial random variable in order to account for the feature of discrete changes in transaction price. However, they did not incorporate the volume information in their model.

The literature that followed from Russell and Engle's (2005) work has used either transaction data that contain variables in addition to transaction price and volume traded, or data that contain information about the requests made by the buyers and sellers of a share. Transaction data that contain additional variables are used by Zhang et al. (2009) to construct the sequences of midquote durations. These sequences are used to compare the out-of-sample forecasting performance of the model proposed by Russell and Engle (2005) with that of a simpler class of models. They reported that Russell and Engle's (2005) model produced more accurate forecasts than those produced by the simpler class of models, but that they are also computationally more demanding to apply than are the simpler models. Tay et al. (2004) chose to classify the transaction price changes into one of the three possible states: a price increase, a price decrease or no price change. They then studied the impact of the covariates (here they used trading volume, trading direction and their interaction as well as the lagged durations) on the durations between consecutive transactions when the price change over this duration is in a given state. Kwok et al. (2009) applied the model but used the

buy-sell initiated trade indicator[6], in addition to the trading volume, as covariates.

Ng (2010) analysed a dataset containing information on orders to buy or to sell a share. He considered the binary variable that indicates whether an order recorded on an electronic exchange is to buy or to sell a share. He viewed the sequence of these binary variables as a special case of the model proposed by Russell and Engle (2005) and used the resulting model to study the relationship between the buy-sell variable and the covariates, for example last transaction price, spread, bid-ask volume.

Another variation to model the discreteness of price change between consecutive trades is that proposed by Müller and Czado (2009) who categorised the transaction price changes into seven categories and regarded these as an ordinal-valued random variable. They then formulated their regression model to contain two components: a component to model the covariates (here, the two covariates used are the durations between consecutive transactions and the volume traded) and a component to model the correlation structure using a type of stochastic volatility model. Czado et al. (2010) applied that model to the case when other covariates, like the time-of-day and the number of quotes between trades, are used.

There is also literature that approached the modelling of irregular interval and discreteness of transaction price changes differently to that of Russell and Engle (2005). For example, Haug and Czado (2006) studied the relationship between the

---

[6]A trade is initiated by a buyer if the buyer sought for owners of share who are willing to sell and conversely for a trade initiated by a seller. For a discussion of buyer- or seller- initiated trades, see Hautsch (2004, Section 4.1.3).

absolute return volatility at the time of a transaction and the possible covariates (here they used only the bid-ask spread and the duration between consecutive trades). In contrast to Russell and Engle (2005) they considered the price change between consecutive trades as a continuous valued random variable. Instead of modelling the change in consecutive transaction prices directly, another approach is to model the aggregate of the square of each return within a trading day, called the *realized volatility* of intraday returns[7]. For example, Oomen (2006) considered the intraday return as the composition of a cumulative transaction price changes component and a cumulative error component. He used the Poisson process to describe the number of trades observed at the time instant t since the observation began, and the moving average process[8] to describe the error contained in the transaction price observed. Then the cumulative price changes component is modelled by the aggregate of the normally distributed random variables, and the cumulative error component is modelled by the aggregate of the moving average process. For both components, the number of variates summed is given by the difference in the outcome between the Poisson process at the end and at the beginning of a time interval. Oomen (2006) used the intraday return model considered to construct the realized volatility, which is an estimator for the integral of the instantaneous volatility over the duration of the trading day. He reported that the

---

[7]The price of the share is determined at a regular time interval within a trading day. Thus, the intraday return is the change in the share price at consecutive time intervals within a trading day. For a review of realized volatility, see McAleer et al. (2008).

[8]He used the moving average process with the constant term equals to zero and with the difference of standard normal variates used for each error term.

realized volatility has a smaller mean squared error when the transaction prices at the time of every fixed number of transactions are used than when the share prices are determined at a regular time interval within a trading day. Andersen et al. (2007) studied the realized volatility, where the intraday returns are computed over consecutive five-minute intervals. The realized volatility computed over the preceding day, the preceding week (they used the preceding five trading days) and the preceding month (they used the preceding twenty-two trading days) were used as the explanatory variables for the realized volatility computed over a proceeding number of days. They considered the realized volatility to consist of two components: a component for the continuous price evolution and a component for the aggregate "significant jumps" in the share prices. A nonparametric procedure was proposed to measure each component separately. Andersen et al. (2007) reported that the models that included the continuous evolution component and the aggregate jumps component as separate explanatory variables produced more accurate realized volatility forecasts than those produced by the models that include the realized volatility itself as the explanatory variable.

Problem statement and contributions made. The aim of this dissertation is to study the volatility of price movement in the transaction data of shares listed on the electronic trading system of the Frankfurt Stock Exchange. These data consist of observations recorded continuously during a trading day. The two research questions addressed in this dissertation are as follows. The first question

is whether the continuous-time models proposed by Hujer et al. (2002) can be used successfully to forecast the waiting times between successive price changes of at least a given level. The second question is whether the discrete-time models we propose can be used successfully to forecast the price changes observed at regular intervals during a trading day.

The contributions made in this dissertation are as follows. Firstly, we apply the class of models proposed by Hujer et al. (2002) to the waiting time between significant price changes. These waiting times contain both the variation in the local trend and the intra-daily seasonal variation. We model both variations then remove their effects from the observations. Although the presence of the intra-daily seasonality of the waiting time between successive transactions (and hence successive price changes of a given amount) are well documented (see Hautsch, 2004, Section 5.2.4), the variation in the local trend has been neglected in the literature on financial duration modelling. Our results show that the out-of-sample forecasts of the intra-daily seasonal variation play an important role for the accuracy of the forecasts made by the model. We fit the models of Hujer et al. (2002) to the standardised waiting time after both the effect of the intra-daily seasonal and of the local trend variations have been removed. The likelihood function of Hujer et al.'s (2002) model possesses several local maxima[9]. As a result, the local maximum identified by a numerical optimisation routine depends on the set of

---

[9]This is a problem common to all mixture models. The HMMs and the models of Hujer et al. (2002) are examples of dependent mixture models. For a description of the problem of multiple maxima in the likelihood of a HMM, see Zucchini and MacDonald (2009, Section 3.4.1).

starting values used. We develop a strategy to search for the set of starting values that improves the chances of locating the global maximum. We show that the forecasts given by the models that depend on an unobserved state process are more accurate than the forecasts made by the "standard" models for the waiting time between significant price changes. We show that the intra-daily seasonal variation is the chief issue for the success of the model. We also show that the parameters of the model vary substantially over successive rolling periods.

Secondly, we study the intra-daily price movement and its relationship with the trading volume. In contrast to the autoregressive multinomial model of Russell and Engle (2005), the models we propose here treat the intraday return as a composition of a binary variable that indicates the presence or absence of price change, and a continuous-valued variable for the conditional return given the presence of change. The variation in both the price change indicators and the conditional returns depends on an unobserved state process and the associated trading volume. We show that in most cases our models provide more accurate forecasts of the variance of price change than those computed by the exponential filter.

Overview of the dissertation. This dissertation contains two main chapters: Chapters 2 and 3. In Chapter 2 we give an account of our investigation for the first research question. In that chapter we study the waiting time between significant price changes. We show that these waiting times contain both a variation

in the intra-daily seasonality as well as a variation in the average waiting time in a trading day. The standardised waiting time is obtained by estimating and removing firstly the seasonal variation and then the variation in the daily mean waiting time. We fit the class of models proposed by Hujer et al. (2002) to these standardised waiting times. The parameters are estimated by numerical maximisation of the likelihood function. The likelihood function of Hujer et al.'s (2002) model has several local maxima. As a result, the local maximum identified by the algorithm depends on the set of starting values used to initialise the iterations. We develop a strategy to reduce the dangers of not locating the global maximum, and hence the maximum likelihood estimate.

In Chapter 3 we give an account of our investigation for the second research question. We propose another class of models for the variance of the price changes. The presence of price change is modelled separately from the size of the change given the occurrence of price change. We incorporate the volume into each of the component models to explain the intra-daily seasonality of the observations. For each component of price change we first select a model and then we assess the fit of this model in the out-of-sample period.

CHAPTER 2

THE MARKOV SWITCHING ACD MODELS FOR THE TIMES BETWEEN PRICE EVENTS

In this chapter we forecast the time until the level of the price moves by at least a given amount. The movement of the share price is modelled in continuous-time, since a trade (and hence a price movement) occurs at random throughout the trading hour.

Various measures of volatility can be constructed by using these waiting times. One example is the instantaneous volatility estimator that was applied in Engle and Russell (1997, Section 6; 1998, Section 8.1) and another example is the conditional volatility given preceding waiting times that was discussed in Gerhard and Hautsch (2002). For a discussion of volatility measures that use the waiting times between successive "significant" price changes, see Hautsch (2004, Section 3.3.1).

In this chapter we describe the data for three shares and then describe how the observations are generated from these data. We investigate the stationarity, the distribution and the serial correlation structure of these observations series. The models are considered in Sections 2.5 and 2.6. We apply these models to nine series of observations. The findings of our applications are discussed.

12

Period	Hours	Duration (hours)
01.1998 to 17.09.1999	8:30 to 17:00	8.5
20.09.1999 to 01.06.2000	9:00 to 17:30	8.5
02.06.2000 to 10.2003	9:00 to 20:00	11
11.2003 to 12.2004	9:00 to 17:30	8.5

Table 2.1: Trading hours of the XETRA. On the last trading day of each year XETRA closes at 14:00 .

The main references for this chapter are: Hautsch (2004) for the autoregressive conditional duration models, Zucchini and MacDonald (2009) - referred to as ZM hereafter - for the hidden Markov models, and Hujer et al. (2002) for the Markov switching models.

2.1 Description of the transaction data

We used a transaction dataset released by the Deutsche Finanzdatenbank (DFDB), Karlsruhe. This dataset contains all transactions recorded in each trading day. The period covered is 28 November 1997 to 30 December 2004, both inclusive. During this period the trading hour changed three times. Table 2.1 displays the trading hours and the periods to which they apply. For each transaction six variables were recorded: the securities identification number[1], the date, the time (precise to centiseconds), the price (in Euros), the volume (in units of shares

---

[1]It is a six-digit alphanumeric code used to identify a share listed on a stock market in Germany.

traded) and the code[2] that indicates the form[3] of trading. As of 2 January 2004, the International Security Identification Number[4](ISIN) is an additional variable contained in the dataset.

The transaction dataset we analysed was prepared from the records of the components of DAX30 traded on the exchange electronic trading (XETRA)[5] system of the FSE. Using the price of the components on the XETRA system, the DB calculates the value of DAX30 at every second during the trading hour.

2.1.1   Selection of shares and observation periods

We analysed the following three shares: Allianz (ALV, insurance sector), E.ON (EOA, utilities sector) and Badische Anilin- und Soda-Fabrik (BAS, chemicals sector). For all three shares the data cover the three-month period March to May in each of the years 2002-2004.

The three shares were selected for the following reasons. Firstly, each of the shares remained[6] as a component of the DAX30 throughout the period covered by the data. Secondly, in the year 2004 each of these shares was the most actively traded in its sector. Figure 2.1 depicts the annual trade for each selected share.

---

[2]See Table 4.1 on page 206.

[3]See Deutsche Börse AG (2004, pp. 18-31) for a detailed description.

[4]It is a code consisting of two letters (for example, DE to represent Germany) followed by a ten-digit alphanumeric code. This code is the international standard for quoting a share listed on a stock exchange.

[5]It was developed by the Deutsche Börse AG (DB) which operates the Frankfurt Stock Exchange (FSE). The system first came into operation on November 28, 1997. Hautsch (2004, Section 4.2.2) gives more details of its functioning.

[6]The DB reviews the composition of the DAX30 index in September of the trading year. Outside of this date, a company is excluded from the index composition if it is no longer one of the 45 largest companies listed on the FSE, and a company can be included if it ranks among the 25 largest companies listed.

In the top panel of this figure we note that there is an upward trend in the annual trades. The decrease in trading activity for the years 2003 and 2004 was due largely to the shortened length of trading hour. We note in the bottom panel of Figure 2.1 that for the years 2003 and 2004 Allianz is the most actively traded (i.e. top ranked) share among the components of DAX30. The amounts of annual trading for E.ON and BASF are very similar for the years 1998-2004.

Figure 2.2 displays the plot of consecutive daily returns on Allianz for each of the years 2002-2004. We computed the daily returns as $100 \cdot \ln(p_d/p_{d-1})$, where p_d is the opening price on day d. In Figure 2.2 we note that values of return in excess of $3-$standard deviations occurred more frequently in the year 2004 than in 2003. This indicates that the daily returns in the year 2004 were more volatile than in 2003. We also note in Figure 2.2 that in the years 2002 and 2003 the daily returns were not volatile during the three-month period March to May. In the year 2004, however, the daily returns were more volatile during March to May than during the second-half of the year. In Appendix B. we plot the daily returns on EOA (in Figure 4.1) and BAS (in Figure 4.2) for the years 2002-2004. In each of these figures we again note the increased volatility in March to May of the year 2004.
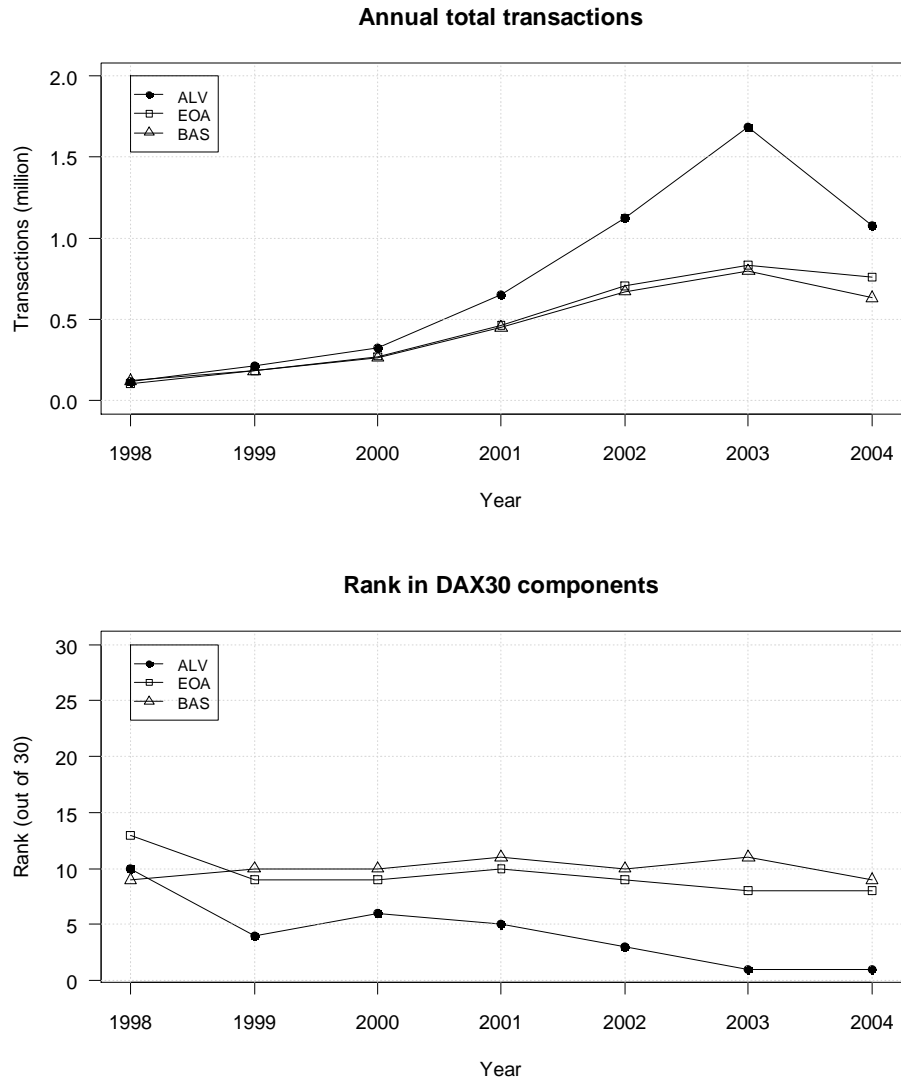
Figure 2.1: Number of annual transactions for three shares, and their ranks (rank 1 indicates the most frequently traded).
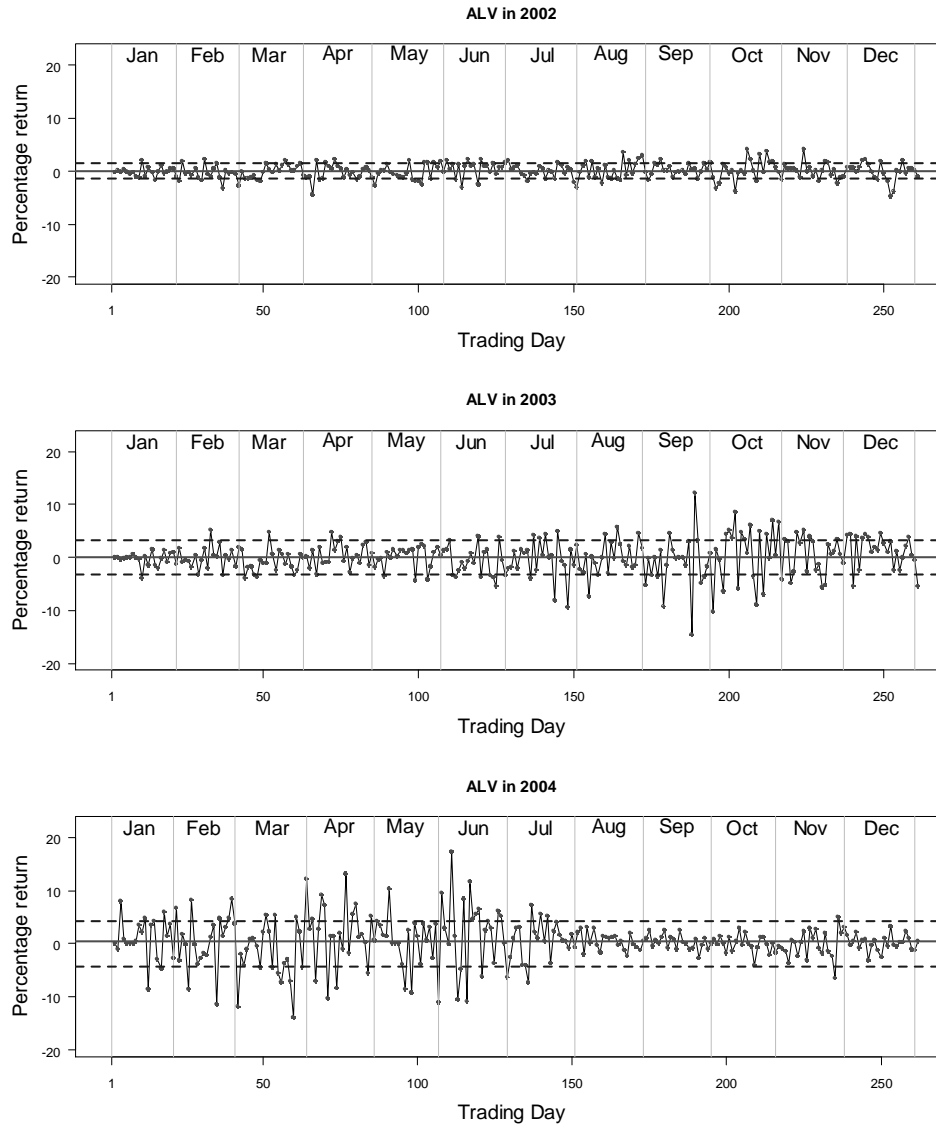
Figure 2.2: Daily returns on the opening price of the Allianz shares. The mean

return (solid line) and 3-standard deviation (dashed lines) over the year are shown.

2.1.2   Adjustments to transaction data

Two features of our transaction dataset are: firstly, instances of several trans-
actions that occur simultaneously, and secondly, the conspicuous delay to execute
the transaction after an auction[7]. For the first feature we treated the volumes of
the elements as part of the volume of one order. We therefore replaced each se-
quence of trades by a single transaction executed at the common transaction time
of the sequence, at the volume-weighted average price of the elements, and with
the sum of the elements' volume traded. For a discussion of other treatments[8]
of the first feature, see Hautsch (2004, Section 4.1.2). No trading is permitted
by the XETRA system during an auction. In the case that this pause occurred
between two consecutive trades we deduct the length of the period of pause from
the corresponding waiting time (See Section 2.2.1).

2.2   Generation of price durations

From the dataset adjusted for trades sharing identical transaction time, we
extract the nine sub-datasets specified in Section 2.1.1, i.e. the trades of ALV,
EOA and BAS, each covering the period March to May in the years 2002-2004.
For a given sub-dataset we denote the number of trades by N, and the opening

---

[7]Three auctions, each of a different type, are held during each trading day of the year 2002-
2004: the opening auction at the opening time, the intraday auction at 13:00 (and at 17:30 for
trading days in the period January 2002 to November 2003, both inclusive), and the closing
auction at the closing time. See Deutsche Börse AG (2004, pp. 19-26) for an account. The
volatility interrupted auction, in addition, can also be held during a trading day. See Deutsche
Börse AG (2004, pp. 27-31) for an account.

[8]Note that, however, after applying the method he selected, Hautsch (2004, p. 49, footnote 4)
deleted the trades sharing a common transaction time that remain before applying the models.

time for trading in the first trading day by t_0. Then, for $n = 1, 2, \ldots, N$, the $n-$th transaction since t_0 is recorded at the transaction time τ_n, and executed at the transaction price p_n (Euros per unit of share traded) for the transaction volume of v_n (units of shares traded).

2.2.1 Basic types of duration

Three basic types of durations can be computed by using our transaction data[9]. The first type of duration, which Hautsch (2004, p. 32) called the trade duration, refers to the waiting time between successive transactions. We compute the trade duration of the $n-$transaction, denoted by W_n, by using $W_1 = \tau_1 - t_0$ and

$$W_n = \tau_n - \tau_{n-1} \quad (\text{for } n = 2, 3, \ldots, N). \tag{2.1}$$

For each sub-dataset we make two adjustments to the trade durations obtained by using Equation (2.1). Firstly, we exclude the waiting time between the last transaction in a trading day and the opening time of trading in the subsequent day. Secondly, we subtract the minimum[10] duration of an auction[11] from the trade duration of the transaction after this auction. We use the trade durations obtained after we make these two adjustments.

---

[9] See Hautsch (2004, Section 3.1.1) for an account of the types of duration for other transaction datasets.

[10] The opening auction ends at the opening time. The volatility interrupted and the closing auction last at least two minutes long. The intraday auction lasts at least two minutes, but at least five minutes on a trading day when the financial security is traded under the derivative contract transacted on the European Exchange (Eurex).

[11] For a description of the code used to indicate the auction corresponding to the transaction, see Appendix A., Table 4.1.

The second type of duration, which Hautsch (2004, p. 33) called the price duration, refers to the waiting time between successive "significant" price changes. Define the price event as the transaction in which a significant price change occurs. For $k = 1, 2, \ldots$, we denote the $k-$th price event in a dataset by n_k. We compute the price durations in two steps. In the first step we identify the events according to the rule in Engle and Russell (1997, p. 196). This is done by setting the initial transaction as n_1, and then defining n_k as the first transaction after n_{k-1} in which p_n satisfies

$$\left| p_n - p_{n_{k-1}} \right| \geq M \quad \text{(for } k = 2, 3, \ldots\text{)},$$

where M denotes the minimum magnitude (in Euros) for a price change to be significant. The total number of events in a dataset is denoted by K, and this is a function of the M we use. In the second step, denote by $\tau_k^p(M)$ the time at which the n_k-th event occurs. We then compute the price duration until n_k, which we denote by $W_k^p(M)$, according to the following equation

$$W_k^p(M) = \tau_k^p(M) - \tau_{k-1}^p(M) \quad \text{(for } k = 2, 3, \ldots, K\text{).} \tag{2.2}$$

We refer to $\{W_k^p(M)\}$ thus generated as the "$M-$" price durations whenever it is necessary to identify the M we used. Our datasets contain transactions in successive trading days. We restart the computation in each day. In doing so the serial order applies only to the day in which the durations series are computed.

The third type of duration, which Hautsch (2004, pp. 33-34) called the volume duration, refers to the waiting time between successive "significant" cumulative

volume traded. This is not analysed in this chapter.

2.2.2 Issues

In order to generate price durations for a given sub-dataset we require a price change level M to be given. One approach is by using the sub-dataset to determine M[12]. But, three issues need to be addressed first. These are the division of sub-dataset, a criterion for M and Ms for the out-of-sample periods.

Division of sub-dataset. We partition a given sub-dataset into subsamples in three consecutive monthly periods. Then we generate price durations for each subsample separately.

Criterion. Given a subsample we generate series of price durations at a variety of Ms. For each series generated we compute the series of counts of trades in consecutive price durations. We use three categories of price change levels: small-sized M (sm-M), medium-sized M (md-M) and big-sized M (bg-M). In order to select an M for each category we adopt the criterion that the average of observations in the counts series is roughly equal to a value that we assign. Following Hautsch (2004, p. 62) we assign the value of 10 (trades per unit time) for sm-M and 500 for bg-M. For md-M we assign the value of 150. Given a one-month period the price durations series at the same category of M are comparable, since these series have similar lengths.

---

[12]Alternatively, M can be supplied by the "decision-maker" (for example, an investor or a risk manager): see Gerhard and Hautsch (2002, p. 62).

Table 2.2 presents, for each category of price change, the averages of trade count and the lengths of the price durations series for ALV in March of the years 2002-2004. In this table we note that the lengths of both the md-M and the bg-M series are less than 1000 in each of the three years. These series are not sufficiently long for the application of the models we consider in this chapter. For each category of price change we also present the averages of trade counts and the lengths of price durations series for EOA (in Table 2.3) and BAS (in Table 2.4) in March of the years 2002-2004. In both of these tables we again note that the lengths of both the md-M and the bg-M series are less than 1000[13]. We focus on the sm-M price durations in what follows.

Out-of-sample periods. In order to generate the price durations series for the out-of-sample periods we need to define M for these periods. One way of doing so is as follows. In each sub-dataset we determine M by using the first sub-sample. Then we assign this value of M for each of the two successive monthly periods that follows the first month. By using the subsample for each out-of-sample period we generate price durations at the assigned M.

We now examine the stability of averages of trade count in the three consecutive monthly periods. Table 2.5 presents, in each of the years 2002-2004, the averages of trade count and the lengths of sm-M price durations series for ALV in each of the months March to May. In this table we note that in the year 2002 the values of the average for March, April and May remain roughly 10. In this

---

[13]This problem could be solved by considering periods longer than one month.

month	sm-M	av. trades	#(obs.)
3.2002	0.25	9.4	4505
3.2003	0.10	9.7	14248
3.2004	0.08	10.7	10614
md-M			
3.2002	1.10	151.8	268
3.2003	0.40	137.0	998
3.2004	0.30	145.0	769
bg-M			
3.2002	2.10	502.8	70
3.2003	0.85	513.7	248
3.2004	0.60	507.0	197

Table 2.2: Price durations of Allianz at three categories of M (sm-M, md-M, bg-M). Averages of trade count (av. trades), and the lengths of price durations at each category of M.

month	sm-M	av. trades	#(obs.)
3.2002	0.06	11.0	3260
3.2003	0.06	10.0	7730
3.2004	0.04	8.5	8569
md-M			
3.2002	0.25	152.9	219
3.2003	0.30	174.9	434
3.2004	0.20	169.8	408
bg-M			
3.2002	0.55	501.2	44
3.2003	0.50	497.7	148
3.2004	0.35	516.0	123

Table 2.3: Price durations of E.ON at three categories of M (sm-M, md-M, bg-M).

month	**sm-M**	av. trades	#(obs.)
3.2002	0.05	10.1	3437
3.2003	0.05	9.4	8616
3.2004	0.04	10.7	5965
	md-M		
3.2002	0.20	137.9	245
3.2003	0.25	171.4	454
3.2004	0.15	128.3	473
	bg-M		
3.2002	0.40	504.8	56
3.2003	0.45	510.0	139
3.2004	0.30	496.7	104

Table 2.4: Price durations of BASF at three categories of M (sm-M, md-M, bg-M).

2002	Mar	Apr	May
sm-M	0.25	0.25	0.25
av. trades	9.4	10.8	10.1
obs.	4505	3760	4289
2003			
sm-M	0.10	0.10	0.10
av. trades	9.7	17.6	15.9
obs.	14248	9611	7943
2004			
sm-M	0.08	0.08	0.08
av. trades	10.7	12.7	14.8
obs.	10614	5349	5524

Table 2.5: Comparison of price durations of Allianz corresponding to the same small-M in different months of the year 2002, 2003, 2004: average number of trades (av. trades) between significant price changes, and the number of price changes (obs.).

case the lengths of 0.25−price durations series across the three monthly periods also remain roughly stable. In the year 2003, however, the average for April is 1.8 times that for March. The corresponding length of series for April is two thirds of that for March. In the year 2004 the averages remain relatively stable across the three monthly periods. The corresponding length of series for April is however only half of that for March. We also present, in each of the years 2002-2004, the averages of trade count and the lengths of sm-M price durations for EOA (in Table 2.6) and BAS (in Table 2.7) in each of the months March to May. We note in both tables that in each of the years the averages remain relatively stable across the three monthly periods.

2.3 Analysis of the observed durations

In this section we outline the results of an exploratory analysis of sm-M price durations for three shares. We investigate the seasonal variation within the day, the distribution of the durations and the serial correlation structure of the series. Changes in the daily means are considered in Section 2.3.4. The findings of this preliminary analysis are summarised.

2.3.1 Features of variation

Figure 2.3 displays, for each of the years 2002-2004, the index plots of sm-M price durations for ALV in the left column and a smooth of the points in each

2002	Mar	Apr	May
sm-M	0.06	0.06	0.06
av. trades	11.0	11.0	12.5
obs.	3260	2971	3031
2003			
sm-M	0.06	0.06	0.06
av. trades	10.0	11.3	12.9
obs.	7730	5848	4725
2004			
sm-M	0.04	0.04	0.04
av. trades	8.5	9.9	9.0
obs.	8569	5973	6111

Table 2.6: Comparison of price durations of E.ON corresponding to the same small-M in different months of the year 2002, 2003, 2004.

2002	*Mar*	*Apr*	*May*
sm-M	0.05	0.05	0.05
av. trades	10.1	10.4	11.5
obs.	3437	3430	3563
2003			
sm-M	0.05	0.05	0.05
av. trades	9.4	10.2	11.9
obs.	8616	6823	5463
2004			
sm-M	0.04	0.04	0.04
av. trades	10.7	12.3	13.1
obs.	5965	4030	3587

Table 2.7: Comparison of price durations of BASF corresponding to the same small-M in different months of the year 2002, 2003, 2004.

of the index plot in the right column[14]. In this figure we note the following two features. Firstly, each of the three index plots displays "spikes," which are most pronounced for the period March to May 2002 and the least for the period March to May 2003. Secondly, a smooth of each observations series displays an upward trend, which suggests that the series are not stationary.

In Appendix B. we present the index plot and its smooth of the sm-M price durations for EOA (Figure 4.3) and for BAS (Figure 4.4). In both these figures we note features that are similar to those noted in Figure 2.3.

2.3.2 Distributional features

The left column of Figure 2.4 displays the plot of kernel density estimates[15] of the observed sm-M price durations for Allianz for each of the years 2002-2004. We have also plotted, to the right of each figure, the observed against the exponential distribution quantiles of the price durations. In Figure 2.4 we note two properties that are common to all three sequences of sm-M price durations. Firstly, the unconditional distributions (as represented by the plots in left column) have the general shape of an exponential distribution, i.e. their modes are close to zero. Their tails, however, are all considerably longer than that of the exponential. The numbers of observations whose values exceed 800 seconds indicate this; there are 545 such instances in March to May 2002, 184 in March to May 2003 and 201

---

[14]We use the locally-weighted polynomial regression method to smooth the points. These smooths are computed by the **R** function `lowess`.

[15]Kernel density with Epanechnikov kernel and bandwidth calculated by Silverman's rule-of-thumb. The estimates are computed by `density` function of **R**.
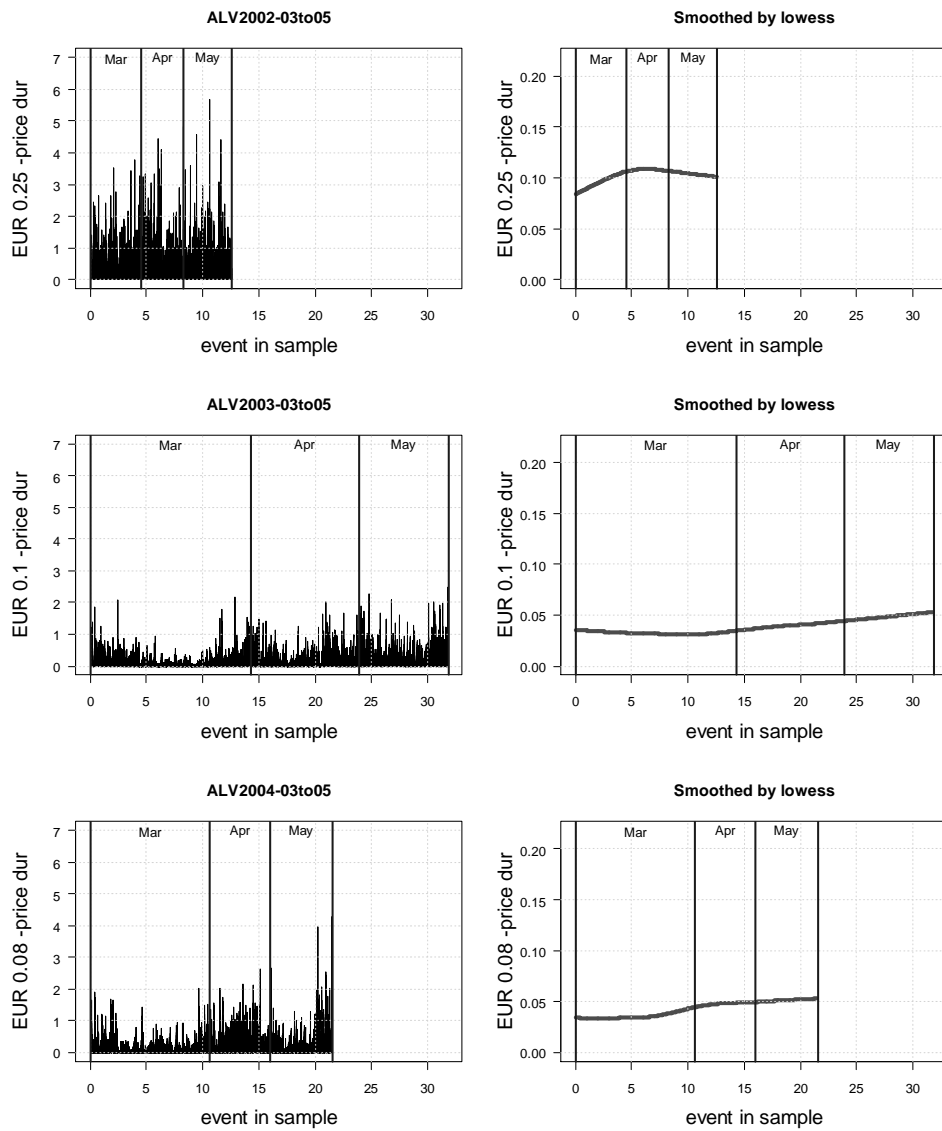
Figure 2.3: Plots of sm-M price durations (in units of 1000 seconds) for ALV against index of price event (in units of 1000) for the months March to May in the years 2002-2004 (left column). Nonparametric regression lines computed by the **R** function `lowess` that is a smooth of the points in each of the years (right column).
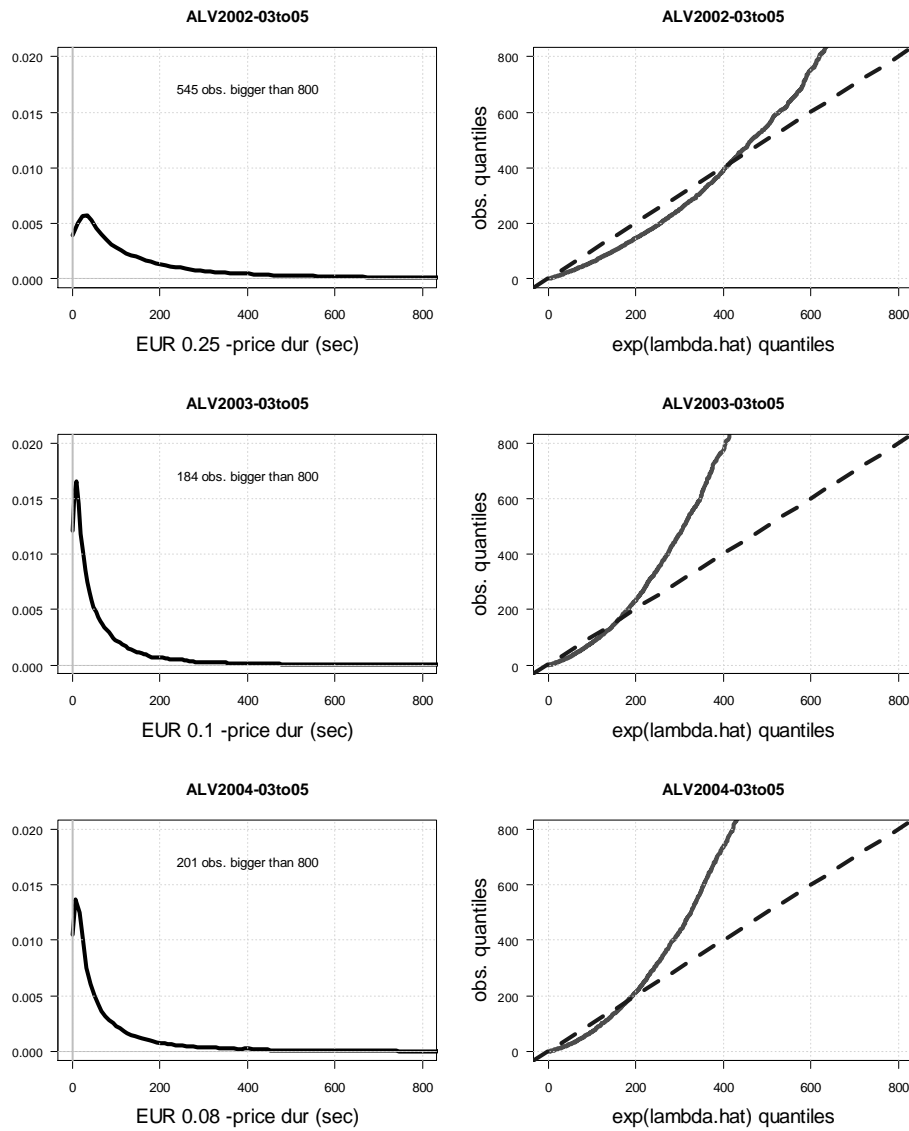
Figure 2.4: Kernel density estimates of the observed sm-M price durations of Allianz (left column). qq-plot of the observed price durations based on an exponential distribution (right column).

in March to May 2004. Secondly, the qq-plots (right column) show that for the shorter durations, i.e. less than 400 seconds (top) or less than 200 seconds (middle and bottom), the observed durations occur less frequently than the corresponding exponential distribution. But, for the longer durations, the observed durations occur much more frequently than the corresponding exponential distribution.

In Appendix B. we present the plots of the kernel density estimates and the qq-plot of the sm-M price durations for E.ON (Figure 4.5) and BASF (Figure 4.6). In these figures we also note the property of less frequent small durations and more frequent large durations for these two sets of sm-M price durations.

Table 2.8 shows the descriptive statistics for sm-M price durations generated from the transactions for Allianz, E.ON and BASF from March to May. The last column of Table 2.8 shows that the sm-M price durations generated from transactions on all three shares are over-dispersed relative to an exponential distribution, i.e. the sample standard deviation exceeds its sample mean.

2.3.3 Autocorrelation

The ACF of the observed price durations for Allianz, in the years 2002-2004, are displayed in Figure 2.5. The price durations in each of the year correspond to sm-M price changes in March, April and May. Note in Figure 2.5 two features that are common to all three sample ACFs. Firstly, the sample autocorrelation at lag 1 is substantially larger than zero[16]. Secondly, at longer lags, the sample

---

[16]In all three sample ACFs it is in excess of 0.3.

ALV

year	*sm-M*	*#(obs.)*	*min.*	*med.*	*max.*	*s.d.*	*mean*	$\frac{s.d.}{mean}$
2002	0.25	12554	0.01	89.95	5668.10	319.87	195.58	1.64
2003	0.10	31802	0.01	32.27	2499.44	134.03	76.35	1.76
2004	0.08	21487	0.01	36.06	4262.34	167.13	90.58	1.85

EOA

year	*sm-M*	*#(obs.)*	*min.*	*med.*	*max.*	*s.d.*	*mean*	$\frac{s.d.}{mean}$
2002	0.06	9262	0.01	128.25	6137.39	424.79	266.26	1.60
2003	0.06	18303	0.01	65.99	3645.99	212.28	132.92	1.60
2004	0.04	20653	0.01	42.81	2896.16	157.74	94.33	1.67

BAS

year	*sm-M*	*#(obs.)*	*min.*	*med.*	*max.*	*s.d.*	*mean*	$\frac{s.d.}{mean}$
2002	0.05	10430	0.01	110.50	5823.18	379.55	236.06	1.61
2003	0.05	20902	0.01	56.75	4634.91	191.41	116.45	1.64
2004	0.04	13582	0.01	64.32	5537.23	245.56	143.39	1.71

Table 2.8: Descriptive statistics of observed small-M price durations (in seconds) for Allianz, E.ON and BASF for the period March to May, 2002-2004. The eight columns are: size of small-M (Euro), length, shortest, median, longest, standard deviation, average and relative dispersion.
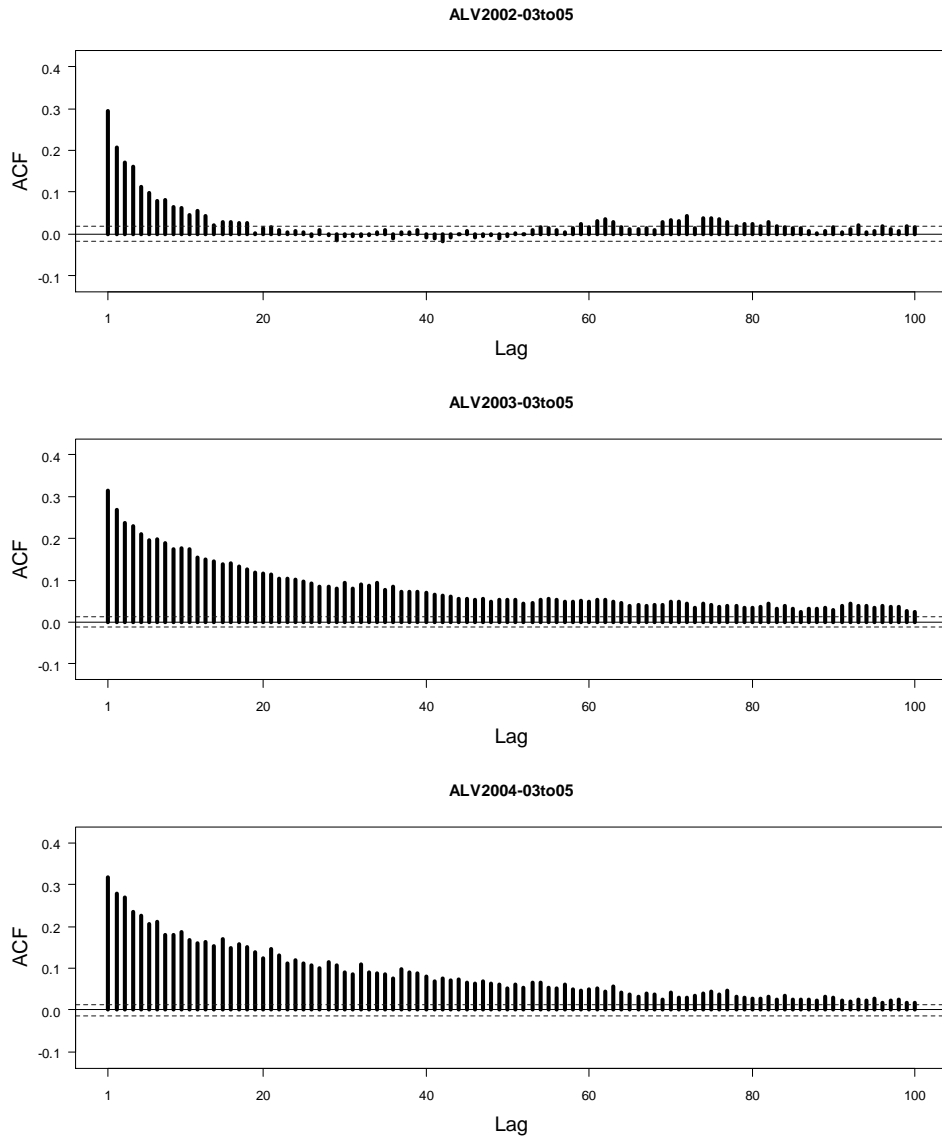
Figure 2.5: ACF of observed price durations of Allianz corresponding to small-M price changes for the period March to May in years 2002, 2003 and 2004.

autocorrelations decrease slowly towards zero. This suggests that the observation sequence may contain trend and / or seasonal variation over the period covered.

We have also presented the ACF of the observed price durations, for the years 2002-2004, of the E.ON share (Figure 4.7) and of the BASF share (Figure 4.8) in Appendix B.. The price durations in each year correspond, again, to sm-M price changes in March, April and May. These ACFs exhibit similar features to those noted in Figure 2.5.

2.3.4   Components of variation

Figure 2.6 displays each of the two components of variation exhibited by the price durations series for Allianz in the years 2002-2004. The observations for each of the years correspond to sm-M price changes for the months March to May. In order to reveal the intraday seasonality in these series we estimate the regression of price duration on the time at which the last significant price change occurred by a smooth function[17].

We note in Figure 2.6 two properties that are common to all three series of observations. Firstly, the expected duration functions estimated (left column) show the systematic variation of price duration sequences over the period of a

---

[17]This is computed by the **R** function `smooth.spline`. We define here twenty-one knots for series in the years 2002 and 2003, and sixteen knots for those in 2004. By following Hautsch (2004, p. 65), who places a knot at 30-minute intervals during the trading day, we deduce these numbers from the length of duration of the trading day for the periods covered by the series (see Table 2.1).

The generalised cross-validation (GCV) method has also been used, but the numbers of knots we obtained by using this method are unreasonably large. For example, we obtain roughly two hundred knots for each of the series for March in the years 2002-2004. From these numbers we can deduce that, for March in the years 2002 and 2003, the knots are placed roughly at intervals of 3 minutes and 20 seconds, and, for March 2004, at intervals of 2 minutes and 30 seconds.
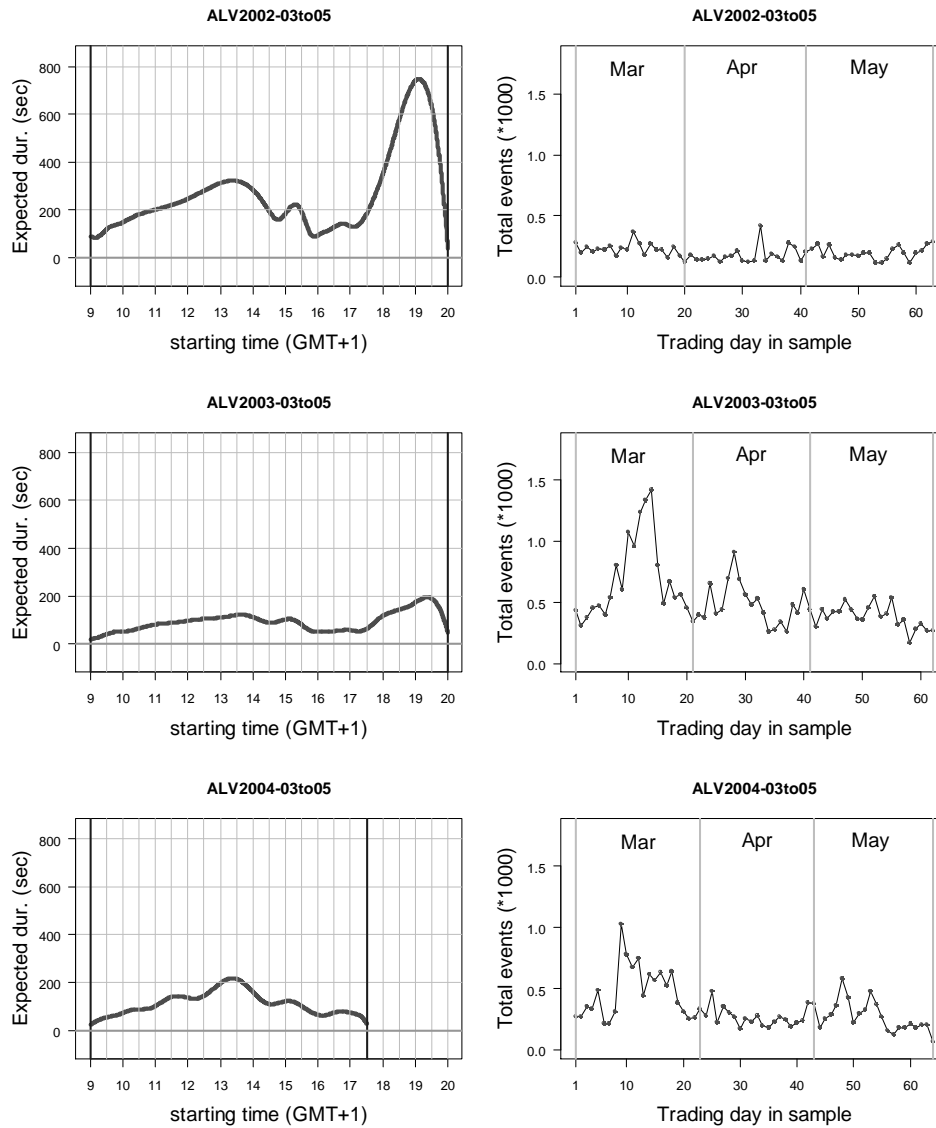
Figure 2.6: Observed price durations of Allianz: nonparametric estimate of intraday cyclical pattern (left column) and daily average (right column). The durations correspond to small-M price changes for the period March to May in years 2002, 2003 and 2004..

trading day. The general pattern of this can be described as follows: durations between price events are, on average, the shortest at the beginning and at the end of the day, but longer between one and two o'clock in the afternoon. The longer trading hours (see Table 2.1) for trading days from the period March to May 2002 and March to May 2003 introduces an additional pattern to the daily variation (top and middle): price durations are, on average, the longest around seven o'clock in the evening. Secondly, the index plots show noticeable fluctuation of the daily total number of price events (right column) around some downward sloping trend over the sample period. This is evidence for fluctuation in daily mean duration and the mean is increasing over the sample period (since the number of price events in a trading day is inversely related to the average price duration over the day). We also present the two components of variation of the observed price durations of the E.ON share (Figure 4.9) and the BASF share (Figure 4.10) in Appendix B.. The price durations correspond to sm-M price changes in the same three periods as those used in Figure 2.6. The intraday cycle of these price durations exhibits similar general pattern as those estimated from the price durations of the Allianz share. Furthermore, the daily average of these price durations also exhibit considerable variation.

2.3.5   Properties of observed durations

Based on our analysis, we list four features that are common to all nine sequences of sm-M price durations.

1. The mean is not stationary. The sequences all display both an upward overall trend and seasonal variation.

2. The unconditional distribution is overdispersed (i.e. the sample standard deviation exceeds the sample mean) relative to the exponential.

3. The positive serial correlation is substantial. Furthermore, the sample ACF decays slowly. In addition to the nonstationarity in the mean, this suggests that price durations "cluster," i.e. a long duration tends to follow a long duration and a short duration tends to follow a short duration.

4. The overall trend in the series of total number of price events in each day is downward. In the case of the series of average price duration in each day, this is upward.

2.4   Stationary price durations

In this section we outline the results of an analysis of the components of sm-M price durations for three shares. We consider a multiplicative model for the series, which consists of three factors: the model for the intraday seasonal variation, the level of the daily mean and the model for the intraday random variation. We

preprocess all series in two stages: firstly, the observations are deseasonalised, and secondly, the deseasonalised series are standardised by the levels of their daily averages. We investigate the procedure to deseasonalise the series in Section 2.4.1. The seasonally-adjusted series and the mean-standardised series are then examined. The findings of this analysis are summarised.

2.4.1 Adjustment for intraday seasonal pattern

On trading day d, the deseasonalised duration, denoted by $\mathbf{a}_d = (a_{n,d})$, is constructed by removing the seasonal variation from the observed durations. For a sample consisting of D trading days, the seasonal component is assumed to be the regression of observed durations over the entire sample $\mathbf{w} = (\mathbf{w}_d)$ on its starting times $\mathbf{t} = (\mathbf{t}_d)$, i.e.

$$w_{n,d} = m\left(t_{n-1,d}\right) + e_{n,d} \quad (\text{for } n = 1, 2, \ldots, n_d), \tag{2.3}$$

where $t_{n-1,d}$ denotes the observed starting time of duration $w_{n,d}$ and n_d denotes the number of price durations observed on day d. The errors in Equation (2.3) are serially correlated. The conditional expectation of the duration when its starting time is t centiseconds from midnight on day d, denoted by $m(t)$ in Equation (2.3), is estimated by a smoothing spline function[18].

Note that the regression model (2.3) implies that the seasonal variation of

---

[18]Estimation uses the `smooth.spline` function contained in the base package of R(version 2.9.0). For transactions made in the year 2002 and until the end of October 2003, twenty one knots were specified for the spline function. For transactions made from the beginning of November 2003, sixteen knots were specified. The specified number of knots correspond roughly to knots placed at thirty minute intervals from the start of a trading day.

duration is exactly the same over each trading day in the sample. Of course, more refined models that are capable of better capturing the shape of the seasonal variation can be assumed[19]. However, from our experience, the dominant seasonal variation of duration is that durations tend to be short at the beginning and at the end of the trading day, but long at around one o'clock in the afternoon.

Using the estimated smoothing spline function $\hat{m}(t)$, the seasonal variation is removed from the observed durations according to

$$a_{n,d} = \frac{w_{n,d}}{\hat{m}(t_{n-1,d})}. \tag{2.4}$$

The descriptive statistics of the deseasonalised price durations for Allianz, E.ON and BASF are presented in Table 2.9. In Table 2.9 we note the following. Firstly, the second column from the right of the panels for all three shares show that the means of deseasonalised durations are almost always slightly less than 1. Secondly, the first column from the right of the panels for all three shares show that the unconditional distributions of the deseasonalised price durations remained overdispersed relative to the exponential distribution. Finally, in the fourth column from the right of the panel for BAS we see that the smallest deseasonalised price duration over the period March to May 2002 is a negative value (-30.139).

The smoothing spline function estimated from the durations observed over this period is shown in the top left hand plot of Figure 2.7. The estimated mean duration function displayed by the top left plot of Figure 2.7 takes on negative

---

[19]One possibility would be to assume a different seasonal component for each weekday (see for example, Hujer and Vuletić (2007)).

ALV

year	*sm-M*	*#(obs.)*	*min.*	*med.*	*max.*	*s.d.*	*mean*	$\frac{s.d.}{mean}$
2002	0.25	12554	0.00	0.55	19.97	1.33	1.001	1.33
2003	0.10	31802	0.00	0.48	25.94	1.51	0.999	1.51
2004	0.08	21487	0.00	0.46	63.02	1.58	0.998	1.59

EOA

year	*sm-M*	*#(obs.)*	*min.*	*med.*	*max.*	*s.d.*	*mean*	$\frac{s.d.}{mean}$
2002	0.06	9262	0.00	0.56	19.97	1.31	0.996	1.32
2003	0.06	18303	0.00	0.58	28.53	1.29	0.999	1.29
2004	0.04	20653	0.00	0.51	36.80	1.46	0.999	1.46

BAS

year	*sm-M*	*#(obs.)*	*min.*	*med.*	*max.*	*s.d.*	*mean*	$\frac{s.d.}{mean}$
2002	0.05	10430	*-30.14*	0.55	29.14	1.44	0.998	1.45
2003	0.05	20902	0.00	0.57	29.04	1.30	0.999	1.30
2004	0.04	13582	0.00	0.51	35.24	1.49	0.999	1.49

Table 2.9: Descriptive statistics of (seasonally-) adjusted small-M price durations of Allianz, E.ON and BASF for the period March to May, 2002-2004. The eight columns are: size of small-M (Euro), length, smallest, median, largest, standard deviation, average and relative dispersion.
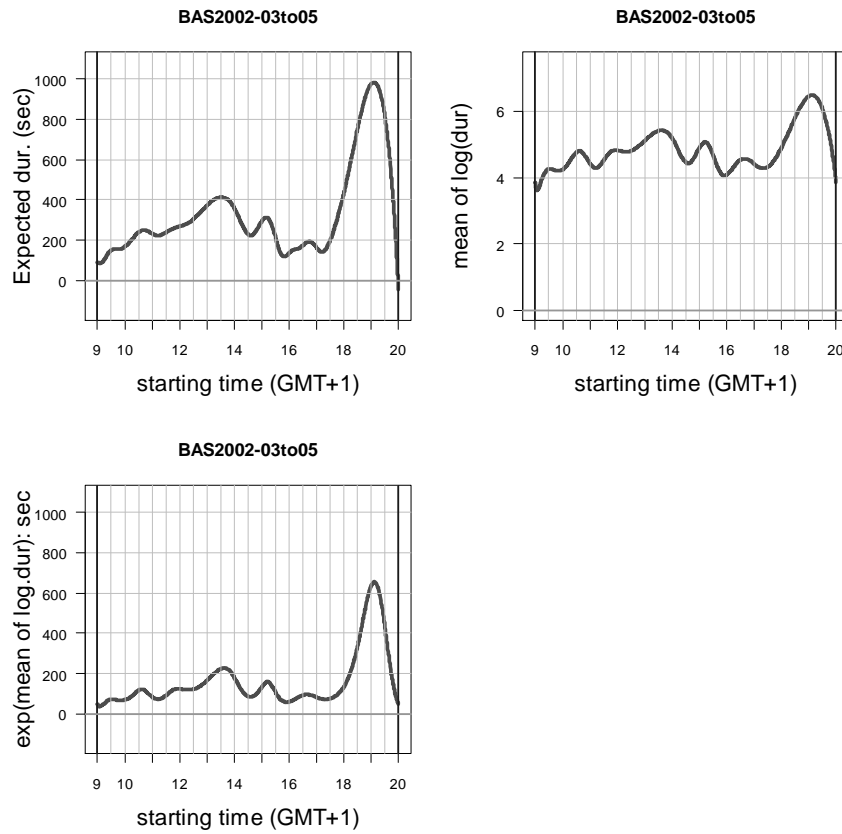
Figure 2.7: Nonparametric estimate of mean price duration for BASF (top left), mean of log-duration (top right) and exponent of mean of log-duration (bottom left) as a function of time of last significant price change. The durations correspond to small-M price changes for the period March to May in the year 2002.

values when the last significant price change occurs between 19:30 and 20:00. This

is due to the fact that model (2.3) does not impose restrictions on the sign of the

response variable.

One way[20] is to restrict the sign of the response variable in two steps. In the

first step, we regress the log-durations on the times of last significant price change,

i.e.

$$\log(w_{n,d}) = \nu(t_{n-1,d}) + e_{n,d} \quad (\text{for } n = 1, 2, \ldots, n_d), \tag{2.5}$$

where $\nu(t_{n-1,d})$ denotes the mean of the log-duration when the time of last price

change is $t_{n-1,d}$. Then, in the second step, we transform the values predicted by

the estimated smoothing function from the log-scale back to the original scale (in

seconds) by exponentiating, i.e.

$$\hat{m}(t_{n-1,d}) = \exp(\hat{\nu}(t_{n-1,d})). \tag{2.6}$$

Finally, the deseasonalised durations that started at $t_{n-1,d}$ are computed by Equa-

tion (2.4).

Applying regression model (2.5) to the price durations of BAS over the period

March-May 2002, the estimated smoothing spline function is shown by the bottom

left plot in Figure 2.7. Note that the shape of the sign-restricted smoothing

spline function is essentially the same as that of the unrestricted smoothing spline

function (top left plot). But, the conditional mean values given by the sign-

---

[20] Another possibility would be to determine the number of knots for the smoothing function
by "generalised" cross-validation method. This is, however, not recommended, because the
observed durations sequence is serially correlated.

restricted function are smaller over the entire range of starting times than the conditional mean values given by the unrestricted function. This observation demonstrates that the log-transformation of durations introduced a bias into the estimated smoothing spline function. Nevertheless, the deseasonalised durations for all three shares were constructed by assuming that the seasonal component from the observed durations is given by the sign-restricted regression model (2.5).

In Appendix B. we display the mean functions (estimated before and after the application of transformation) for sm-M price durations for Allianz (Figure 4.11), E.ON (Figure 4.12) and BASF (Figure 4.13) in the years 2002-2004.

2.4.2 Analysis of adjusted durations

We deseasonalised the observed price durations, corresponding to sm-M price changes in March to May of the years 2002-2004, of the Allianz, the E.ON and the BASF shares. In each year, the intraday seasonal component was estimated using the sign-restricted regression model (2.5). We analysed each of the nine sequences of deseasonalised price durations by examining the plot of its smooth, the plot of daily averages, its descriptive statistics and its ACF.

In the left column of Figure 2.8, we plotted a smooth, computed by the **R** function `lowess`, of the deseasonalised price durations for Allianz in each year. We note that the sample mean for each of the three years is much larger than 1 (it is 2.753 for March to May 2002, 3.426 for March to May 2003 and 3.592 for
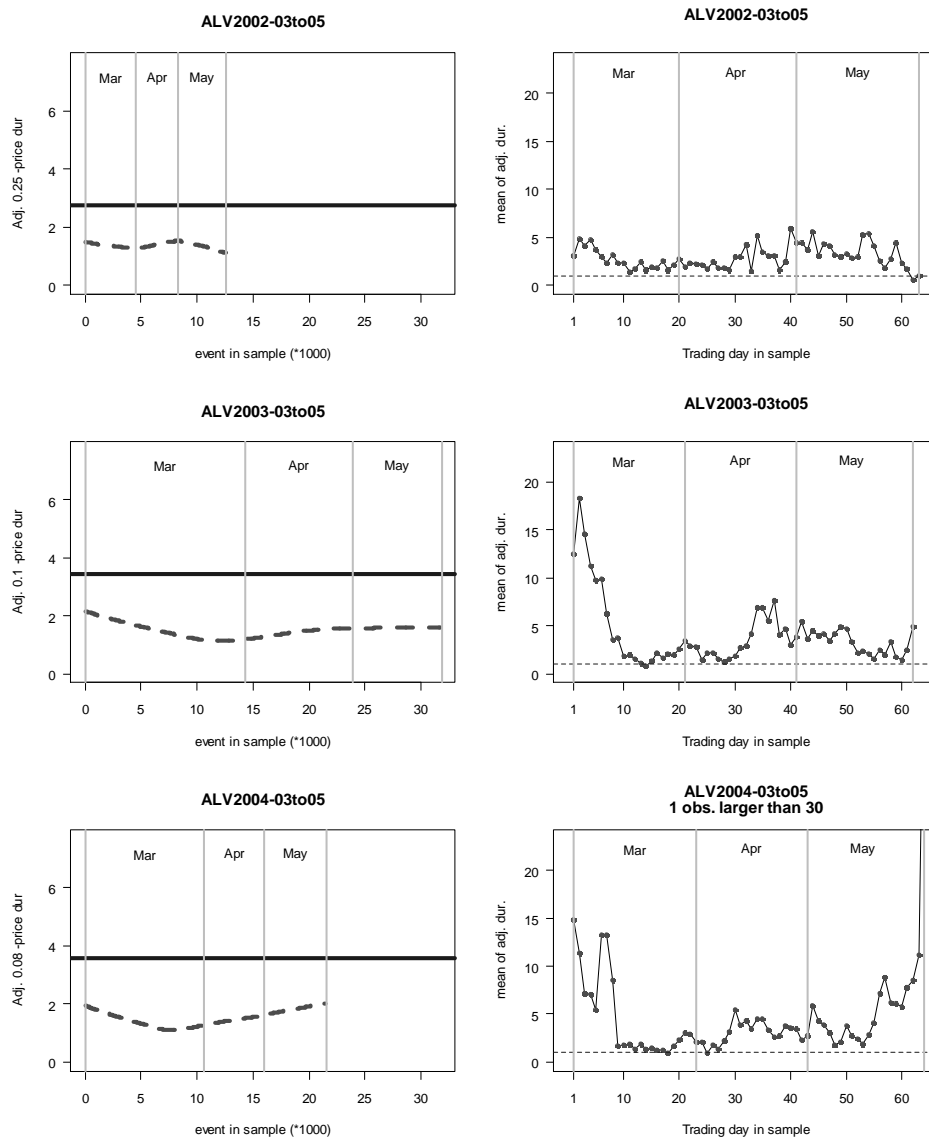
Figure 2.8: (Seasonally-) adjusted price durations of Allianz: mean durations, estimated nonparametrically, as a function of number of significant price changes (dotted line, left column) and daily average (right column). The durations correspond to small-M price changes for the period March to May in years 2002, 2003 and 2004. In each plot of left column, sample mean of durations in each year is indicated by the solid horizontal line.

March to May 2004). The difference from 1 is due to the bias of the estimated smoothing spline function. We also note that the mean duration function is not constant over the number of price changes in each of the three years. In the right column of Figure 2.8, the plot of the daily average of the deseasonalised price durations of Allianz shows considerable variation in each of the three years. We have also presented, for each of the three years, the plot of a smooth and the plot of daily average of the deseasonalised price durations of the E.ON (Figure 4.14) and the BASF (Figure 4.15) shares in Appendix B.. In each of these figures we also note that the mean function is not constant over the number of price changes and that the daily average varies considerably in each year.

Table 2.10 gives the descriptive statistics of the deseasonalised price duration for Allianz, E.ON and BASF. For each share we find that the smallest deseason-alised duration (fourth column) in each year is zero. We also find in each year that the sample mean (column 8) is much larger than 1 and that the deseasonalised price durations is overdispersed (column 9) relative to its mean.

The ACF of the deseasonalised price durations of the Allianz share, in the years 2002-2004, are displayed in Figure 2.9. The price durations in each year correspond to sm-M price changes in March, April and May. We find in each year that the deseasonalised price durations has a significantly large positive first order autocorrelation, and the size of the autocorrelation decreases very slowly with increasing order of the lag. In fact the autocorrelation remains significantly

ALV

year	sm-M	#(obs.)	min.	med.	max.	s.d.	mean	$\frac{\text{s.d.}}{\text{mean}}$
2002	0.25	12554	0.00	1.20	116.32	4.91	2.75	1.78
2003	0.10	31802	0.00	1.20	266.24	7.41	3.43	2.16
2004	0.08	21487	0.00	1.18	423.49	9.15	3.59	2.55

EOA

year	sm-M	#(obs.)	min.	med.	max.	s.d.	mean	$\frac{\text{s.d.}}{\text{mean}}$
2002	0.06	9262	0.00	1.17	76.62	4.50	2.58	1.74
2003	0.06	18303	0.00	1.14	104.34	4.42	2.58	1.72
2004	0.04	20653	0.00	1.20	206.50	6.04	2.95	2.05

BAS

year	sm-M	#(obs.)	min.	med.	max.	s.d.	mean	$\frac{\text{s.d.}}{\text{mean}}$
2002	0.05	10430	0.00	1.18	69.40	4.55	2.69	1.69
2003	0.05	20902	0.00	1.14	152.21	4.82	2.62	1.84
2004	0.04	13582	0.00	1.20	306.28	7.25	3.15	2.30

Table 2.10: Descriptive statistics of adjusted (non-negativity of duration preserved) small-M price durations of Allianz, E.ON and BASF for the period March to May, 2002-2004. The eight columns are: size of small-M (Euro), length, smallest, median, largest, standard deviation, average and relative dispersion.
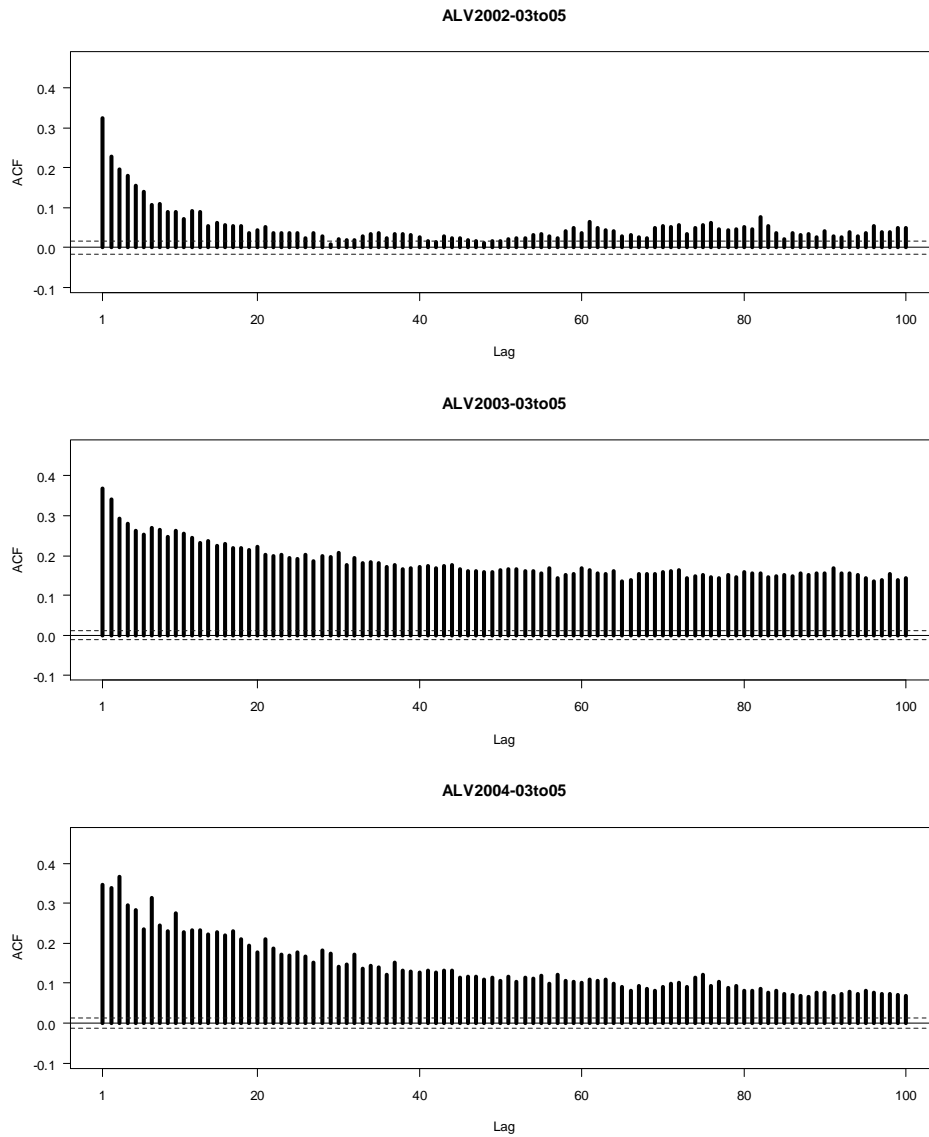
Figure 2.9: ACF of (seasonally-) adjusted price durations of Allianz corresponding to small-M price changes for the period March to May in years 2002, 2003 and 2004.

positive even at lag 100.

We have also presented the ACF of the deseasonalised price durations, in the years 2002-2004, of the E.ON share (Figure 4.16) and of the BASF share (Figure 4.17) in Appendix B.. For each of the shares we again note that in each year the first order autocorrelation of the deseasonalised durations is a large positive value and that the autocorrelation remains positive until at lag 100. The slowly decaying autocorrelation indicates that the variation of the mean duration function we found in each of Figures 2.8 to 4.15 should be removed from the deseasonalised durations.

2.4.3 Standardisation of adjusted durations

One way to remove the variation of the mean duration function is to standardise the daily mean of the deseasonalised durations. In each year with D ($D = 1, 2, \ldots$) trading days we compute the standardised price durations in trading day d ($d = 1, 2, \ldots, D$), denoted by $\mathbf{z}_d = (z_{n,d})$, by dividing the deseasonalised durations by its sample mean on day d, i.e.

$$z_{n,d} = \frac{a_{n,d}}{\bar{a}_d}, \tag{2.7}$$

where $\bar{a}_d$ denotes the sample mean of $(a_{n,d})$ and n_d denotes the number of significant price changes in trading day d. By construction, the standardised durations in trading day d have a mean value of exactly one.

2.4.4   Analysis of standardised durations

We compute the standardised sm-M price durations for Allianz, E.ON and BASF. Table 2.11 gives the descriptive statistics of the standardised series for each share. In this table we note that, in each of the nine series, the overdispersion relative to the standard exponential distribution remains considerable (see column 9).

Figure 2.10 displays the sample ACFs of the standardised series for Allianz. In this figure we note that, for each of the three periods, the positive serial correlation in the series remains strong. In Appendix B. we present the sample ACFs of the standardised series for E.ON (Figure 4.18) and for BASF (Figure 4.19). In both of these figures we note again the strong positive serial correlation in all series.

2.4.5   Properties of standardised durations

We have computed the standardised price durations of the Allianz, the E.ON and the BASF shares, where the durations of each share correspond to sm-M price changes in March to May of the years 2002-2004. We found three features of the standardised price durations from our analysis of each of the nine standardised duration sequences. Firstly, the trend of the standardised durations is roughly constant over the number of price changes in each year. This indicates that the standardised price durations are at least mean stationary. Secondly, the distribution of the standardised durations is overdispersed (i.e. the sample standard

ALV

year	*sm-M*	*#(obs.)*	*min.*	*med.*	*max.*	*s.d.*	*mean*	$\frac{\text{s.d.}}{\text{mean}}$
2002	0.25	12554	0.00	0.47	26.36	1.59	1.00	1.59
2003	0.10	31802	0.00	0.48	30.09	1.58	1.00	1.58
2004	0.08	21487	0.00	0.44	31.35	1.63	1.00	1.63

EOA

year	*sm-M*	*#(obs.)*	*min.*	*med.*	*max.*	*s.d.*	*mean*	$\frac{\text{s.d.}}{\text{mean}}$
2002	0.06	9262	0.00	0.50	30.72	1.57	1.00	1.57
2003	0.06	18303	0.00	0.52	23.59	1.51	1.00	1.51
2004	0.04	20653	0.00	0.48	32.54	1.55	1.00	1.55

BAS

year	*sm-M*	*#(obs.)*	*min.*	*med.*	*max.*	*s.d.*	*mean*	$\frac{\text{s.d.}}{\text{mean}}$
2002	0.05	10430	0.00	0.48	26.36	1.55	1.00	1.55
2003	0.05	20902	0.00	0.52	42.28	1.52	1.00	1.52
2004	0.04	13582	0.00	0.49	37.51	1.54	1.00	1.54

Table 2.11: Descriptive statistics of standardised small-M price durations of Allianz, E.ON and BASF for the period March to May, 2002-2004. The eight columns are: size of small-M (Euro), length, smallest, median, largest, standard deviation, average and relative dispersion.
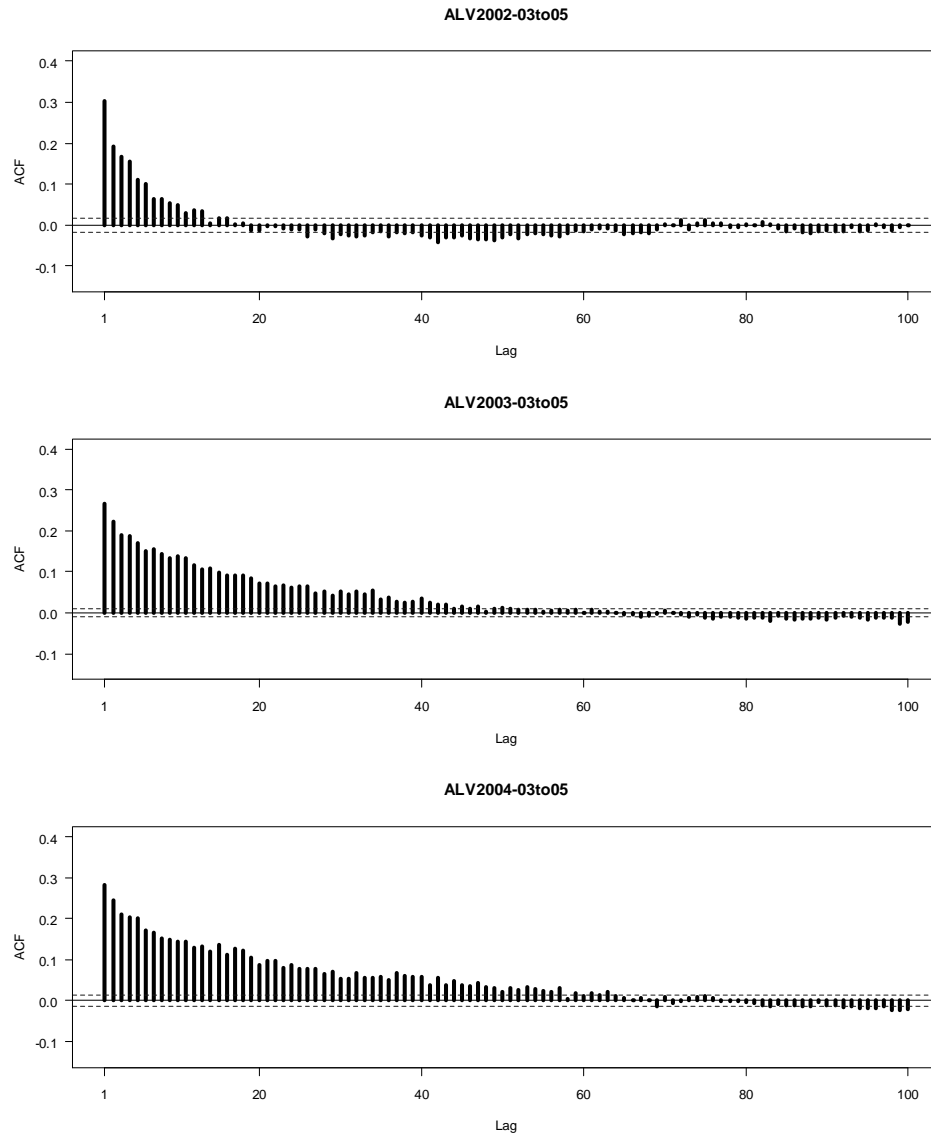
Figure 2.10: ACF of standardised price durations of Allianz corresponding to small-M price changes for the period March to May in years 2002, 2003 and 2004.

deviation exceeds the sample mean). Finally, there is a significantly large positive first order autocorrelation. As the order of the lag increases, the size of this positive autocorrelation decreases slowly towards zero. This indicates the presence of positive serial correlation in the standardised duration sequence (i.e. long duration tend to follow long durations and short duration tends to follow the short durations).

2.5   The ACD models

The standardised durations **z** are a series of nonnegative continuous-valued observations. We assume that these series are also covariance stationary. Engle and Russell (1997, 1998) proposed the class of models, which they called the *autoregressive conditional duration* (ACD) models, for **z**. Their models can accommodate both overdispersion (see, for example, Hautsch, 2004, p. 81) and serial dependence in **z**.

An ACD model for $\{Z_n : n = 1, \ldots, N\}$ consists of two components: firstly, means of durations Z_n, denoted by μ_n, and secondly, random errors, denoted by ϵ_n. The model assumes that the components are combined multiplicatively, i.e. it has the general form:

$$Z_n = \mu_n \cdot \epsilon_n \quad (\text{for } n = 1, \ldots, N). \tag{2.8}$$

In an ACD model the errors are defined by the ratio

$$\epsilon_n = \frac{Z_n}{\mu_n} \quad (\text{for } n = 1, \ldots, N). \tag{2.9}$$

The model assumes that these errors are independent random variables with a common probability distribution, denoted by $g_\epsilon(e)$, and that the mean of each error is 1.

2.5.1 Specification

In order to specify the type of ACD model we apply we need to postulate a model for μ_n and a family of parametric distributions for the error.

The baseline model for μ_n, used in Engle and Russell (1997, p. 191; 1998, p. 1133), is a linear function of p (for $p = 1, 2, \ldots$) lagged durations, $Z_{n-1}, \ldots, Z_{n-p}$, and q (for $q = 1, 2, \ldots$) preceding means, $\mu_{n-1}, \ldots, \mu_{n-q}$. An ACD model with μ_n having this general form is given the abbreviation ACD(p, q). We refer to μ_n as the conditional mean given the history, since it is a weighted average of the preceding observations. To determine the constraints that apply to the parameters of a model for μ_n we consider the special case of ACD$(1, 1)$, i.e.

$$\mu_n = \omega + \alpha Z_{n-1} + \beta \mu_{n-1} \quad (\text{for } n = 2, \ldots, N). \tag{2.10}$$

In Equation (2.10) we require that $\omega > 0$ and $\alpha, \beta \geq 0$ to ensure that the conditional means take on strictly positive values. The unconditional mean for this ACD model is given by[21]:

$$E(Z_n) = \frac{\omega}{1 - (\alpha + \beta)}. \tag{2.11}$$

---

[21]This follows by taking the expection of both sides of Equation (2.10) and by applying the covariance stationarity of $\{Z_n\}$.

By imposing the additional constraint $(\alpha + \beta) < 1$ we ensure the existence of $E(Z_n)$ in Equation (2.11). The recursion (2.10) is initialised by setting μ_1 equal to the unconditional mean. For a survey and classification of several other specifications for μ_n, see Hautsch (2004, Section 5.3).

For the error component of an ACD model we may use, as a model for ϵ, any parametric families of distributions with support defined on the positive real line. As the baseline model Engle and Russell (1997, p. 191; 1998, p. 1133) used the standard exponential distribution, which is given by the density function

$$g_\epsilon(e) = \exp(-e) \quad \text{(for } e \geq 0\text{)}. \tag{2.12}$$

We refer to ACD models with this distribution for the error as the *exponential ACD* (EACD) models.

In order to accommodate the overdispersion, families of distributions at various levels of generalisation of the exponential are used in the literature[22]. To use any one of these families for the error component in the ACD model we select the sub-family that satisfies the following conditions. Firstly, it includes the standard exponential distribution as a special case. This we achieve by fixing the value of one of the several parameters, at which the standard exponential can be obtained. Secondly, the sub-family that satisfies condition 1 has a mean of 1. This we obtain by dividing the distribution concerned by its mean.

---

[22]See Hautsch (2004, Section 5.2.3) for a brief account.

2.5.2 The likelihood

Suppose the N consecutively observed durations $z_1, \ldots, z_N$ are generated by an ACD model with conditional mean μ_n and error described by probability distribution $g_e(\epsilon)$. Define $\mathbf{z}^{(n)} = (z_1, z_2, \ldots, z_n)$. Then, the *exact* likelihood of observing $\{\mathbf{z}^{(N)}\}$ under the specified ACD model is given by the (joint) distribution of $\{Z_n\}$ evaluated at $\{\mathbf{z}^{(N)}\}$. A complication that arises when we want to evaluate this is that we need to provide the distribution of the first few observations, which is not known to us.

We can, however, approximate the exact likelihood instead. One way of doing so is to take the first few observations as known and set the first few conditional means equal to their unconditional mean. For computational convenience, the likelihood L_N we maximise is the probability density of

$$\{Z_{u+1} = z_{u+1}, Z_{u+2} = z_{u+2}, \ldots, Z_N = z_N\}$$

conditioned on the first u observations. For μ_n a function of p lagged durations and q lagged conditional means, we take u to be the larger of p and q.

In the case of the ACD$(1, 1)$ models, the conditional likelihood to be maximised is

$$L_N = f(z_2, \ldots, z_N \mid z_1) = \prod_{n=2}^{N} f\left(z_n \mid \mathbf{z}^{(n-1)}\right). \tag{2.13}$$

The conditional distributions in the likelihood (2.13) are determined by substituting the specified conditional mean and probability distribution describing the

error in Equation (2.13), i.e. by

$$f\left(z_n|\mathbf{z}^{(n-1)}\right) = \frac{1}{\mu_n}g_e\left(\frac{z_n}{\mu_n}\right). \tag{2.14}$$

We use the $EACD\,(1,1)$ model as the basic ACD model. That is, an ACD model with conditional mean μ_n a linear function of one lagged duration Z_{n-1} and one lagged conditional mean μ_{n-1}, and the error described by a standard exponential distribution. The likelihood of observing $\left\{\mathbf{z}^{(N)}\right\}$ under this model is given by Equation (2.13), with conditional distributions following from Equation (2.14)

$$f\left(z_n|\mathbf{z}^{(n-1)}\right) = \frac{1}{\mu_n}\exp\left(-\frac{z_n}{\mu_n}\right) \tag{2.15}$$

and the recursion for the conditional means (μ_n) is initialised by setting $\mu_1 = E\left(Z_1\right)$ and $\mu_2 = \omega + \alpha z_1 + \beta\mu_1$.

In this chapter the family of Burr distributions is the most general model we used for the error component in the ACD model. This was first used by Grammig and Maurer (2000, Section 3.2), in which they call the class of models they obtained Burr ACD (BACD). For this class of ACD models we demonstrate how one selects the sub-family to use for the error component.

Consider the ACD$(1,1)$ model in which the distribution for the error is defined by Equation (A.1) in Grammig and Maurer (2000, p. 35)

$$g_X\left(x|\mu,\kappa,\sigma^2\right) = \frac{\mu\kappa x^{\kappa-1}}{\left[1+\sigma^2\mu x^\kappa\right]^{1+\left(\frac{1}{\sigma^2}\right)}}, \tag{2.16}$$

where its support is the set of all strictly positive real numbers (i.e. $x > 0$) and the

constraints on its parameters are that they must have strictly positive real values.

The mean of the Burr distribution (2.16) follows by setting $j = 1$ in Equation

(A.4) of Grammig and Maurer (2000, p. 36)

$$E\left(X\right) = \mu^{-\frac{1}{\kappa}} \left[\frac{\Gamma\left(1 + \frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{1}{\sigma^2} - \frac{1}{\kappa}\right)}{(\sigma^2)^{1 + \frac{1}{\kappa}} \cdot \Gamma\left(1 + \frac{1}{\sigma^2}\right)} \right], \tag{2.17}$$

where $\Gamma\left(\cdot\right)$ denotes the gamma function. The additional constraint on the pa-

rameter σ^2 is that it must be strictly less than the value of the parameter κ.

This ensures that $\Gamma\left(\frac{1}{\sigma^2} - \frac{1}{\kappa}\right)$ in Equation (2.17), and hence $E\left(X\right)$, is defined. We

represent the family of Burr distributions by $X \sim Burr\left(\mu, \kappa, \sigma^2\right)$.

In the sub-family of Burr distributions that we use, we set the parameter μ

equal to 1. This ensures that the Burr distribution approximates the standard

exponential, if κ equals 1 and σ^2 approaches 0. To obtain the distribution that

has a mean of 1, we follow the literature by dividing the conditional mean μ_n by

the mean of $Burr\left(1, \kappa, \sigma^2\right)$[23] instead. We refer to this quantity as the "scaled"

conditional mean, and it is denoted by ϕ_n.

The likelihood of the $BACD\left(1, 1\right)$ model is given by Equation (2.13) in which

the conditional densities, following from Equation (2.14), are given by

$$f\left(z_n | \mathbf{z}^{(n-1)}; \boldsymbol{\theta}\right) = \frac{\left(\phi_n^{-\kappa}\right) \kappa z_n^{\kappa - 1}}{\left[1 + \sigma^2 \left(\phi_n^{-\kappa}\right) z_n^{\kappa}\right]^{1 + \left(\frac{1}{\sigma^2}\right)}} \quad \text{(for } n = 2, \ldots, N\text{)}. \tag{2.18}$$

The recursion for the conditional means $\left(\mu_n\right)$, and hence $\left(\phi_n\right)$, is initialised by

setting $\mu_1 = E\left(Z_1\right)$ and $\mu_2 = \omega + \alpha z_1 + \beta \mu_1$. The parameters of the conditional

---

[23]Note that this is strictly positive.

densities (2.18) are $\boldsymbol{\theta} = (\omega, \alpha, \beta, \kappa, \sigma^2)'$, where ω, α and β are the parameters that apply to the conditional mean function, and κ and σ^2 are the parameters that apply to the Burr distribution. The conditional distributions for the Weibull ACD models and the generalized gamma ACD models of Lunde (1999, p. 5) are given in Table VI of Hujer et al. (2002, p. 38).

2.6   The Markov switching ACD models

In the literature the standard ACD models were modified or generalised, in numerous different ways, in order to arrive at more flexible models for various types of waiting times (See Chapter 1). Hujer et al. (2002) proposed the Markov switching ACD (MSACD) models in which both components of an ACD model are allowed to depend on the state of an unobserved Markov process. In this way their models can accommodate abrupt changes in the mean of Z_n in addition to the overdispersion.

In this section we define and introduce notation for the MSACD models. We discuss parameter estimation, model selection and model checking. Appendix B. outlines a small (though computer-intensive) simulation study that we used to check the estimation software.

2.6.1 Description

The *Markov switching ACD* (MSACD) model $\{Z_n : n = 1, 2, \ldots\}$ consists of two parts: firstly, an unobserved state process $\{C_n\}$ is assumed to be a Markov chain, and secondly, an observed state-dependent process $\{Z_n\}$ such that the conditional distribution of Z_n given the preceding observations depends only on the current state C_n. Define $\mathbf{C}^{(n)} = (C_1, C_2, \ldots, C_n)$ and $\mathbf{Z}^{(n)} = (Z_1, Z_2, \ldots, Z_n)$. The dependence structure of the model can be summarised by:

$$\Pr\left(C_n | \mathbf{C}^{(n-1)}\right) = \Pr\left(C_n | C_{n-1}\right) \quad (\text{for } n = 2, 3, \ldots) \tag{2.19}$$

$$f\left(z_n | \mathbf{Z}^{(n-1)}, \mathbf{C}^{(n)}\right) = f\left(z_n | \mathbf{Z}^{(n-1)}, C_n\right) \quad (\text{for } n = 2, 3, \ldots). \tag{2.20}$$

For the observed process the ACD model (2.8) is assumed, in which the components depend on the Markov chain as follows. Firstly, the conditional mean of Z_n given $\mathbf{Z}^{(n-1)}$ depends only on the current state C_n but not on the preceding states (this follows immediately from Eq. (2.20)). Let $\mu_{n,j}$ denote this conditional mean, given C_n is in state j (for $j = 1, \ldots, m$). We use a model, like (2.10), for $\mu_{n,j}$:

$$E\left(Z_n | \mathbf{Z}^{(n-1)}, C_n = j\right) = \mu_{n,j}. \tag{2.21}$$

Secondly, we define $\frac{Z_n}{\mu_{n,j}}$ as the error if C_n is in state j. Its distribution depends on the current state but not on the preceding states or observations, i.e.

$$g\left(\frac{z_n}{\mu_{n,j}} \middle| \mathbf{Z}^{(n-1)}, \mathbf{C}^{(n)}\right) = g\left(\frac{z_n}{\mu_{n,j}} \middle| C_n\right). \tag{2.22}$$

Any distribution we use for the error component in an ACD model, such as that

given by Equation (2.12) may be used as the error distribution in state j. We assume that the error, conditional on preceding observations, has a mean of 1, i.e.

$$E\left(\frac{z_n}{\mu_{n,j}}\,\middle|\,\mathbf{Z}^{(n-1)}\right) = 1.$$

Given C_n, the conditional distribution of an ACD model, such as that given by Equation (2.15), can be specified in Equation (2.20).

In this chapter we assume that $\{C_n\}$ in the MSACD models is a stationary irreducible Markov chain (ZM, p. 18), with *(one-step) transition probability matrix* (t.p.m.) $\mathbf{\Gamma}$ and stationary distribution $\boldsymbol{\delta}$. Denote the entries of $\mathbf{\Gamma}$ by $\gamma_{ij} = \Pr\left(C_n = j | C_{n-1} = i\right)$ with row sums equal to one, i.e.

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix},$$

where m is the *number of states* of the Markov chain, and $\boldsymbol{\delta}$ is a row vector satisfying $\boldsymbol{\delta}\mathbf{\Gamma} = \boldsymbol{\delta}$ and $\boldsymbol{\delta}\mathbf{1}' = 1$. If the Markov chain $\{C_n\}$ has m states then $\{Z_n\}$ is called an $m-state$ *MSACD model*.

In our definition of the MSACD model one can use a different type of ACD model for each state. In this chapter we concentrate on the case that one type of ACD model is used for all states.

The parameters of the MSACD model can be divided into two groups: the parameters of the Markov chain and the parameters of the specified ACD model when the Markov chain is in state j at duration Z_n. The Markov chain is com-

pletely determined by its t.p.m. $\boldsymbol{\Gamma}$, thus we refer to the vector collecting the $(m^2 - m)$ off-diagonal elements of $\boldsymbol{\Gamma}$, denoted by $\boldsymbol{\theta_\Gamma}$, as the *parameter vector of the Markov chain.* Given state j, we collect the parameters of the ACD model in a vector denoted by $\boldsymbol{\theta}_j$. Each element in this vector is subscripted by j to denote its state-dependence. We refer to $\boldsymbol{\theta}_j$ as the *state-dependent parameter vector* and to its elements as the *state-dependent parameters* of the MSACD model. For example, the state-dependent parameter vector for a Burr ACD$(1,1)$ model is $\boldsymbol{\theta}_j = \left(\omega_j, \alpha_j, \beta_j, \kappa_j, \sigma_j^2 \right)'$, where ω_j denotes the constant term in the model for $\mu_{n,j}$. In the case that the model for $\mu_{n,j}$ contains more than one lagged duration and/or lagged state-dependent conditional mean, we denote the coefficient of the $k-$th (for $k = 1, \ldots, p$) lagged duration term by α_{jk} and the coefficient of the $l-$th (for $l = 1, \ldots, q$) lagged state-dependent conditional mean term by β_{jl}. For example, the state-dependent parameter vector of an EACD$(2,2)$ model is $\boldsymbol{\theta}_j = \left(\omega_j, \alpha_{j1}, \alpha_{j2}, \beta_{j1}, \beta_{j2} \right)'$, where α_{j2} denotes the coefficient of the second lagged duration term. Note that the parameters in $\boldsymbol{\theta}_j$ can be further divided into two groups: those that apply to the model for $\mu_{n,j}$ and to the model for the error. Denote by $\boldsymbol{\theta}$ the vector collecting $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$ and $\boldsymbol{\theta_\Gamma}$. We then refer to $\boldsymbol{\theta}$ as the *parameters of the MSACD model.*

The MSACD model generates the observations as follows: at the n-th duration the Markov chain selects a parameter vector $\boldsymbol{\theta}_j$, and then the conditional distribution given $\boldsymbol{\theta}_j$ generates the observation z_n.

We now introduce notation and terminology for particular parts of the MSACD model. Define

$$f_{n|n-1,j}\left(z\right) = f\left(z_n|\mathbf{Z}^{(n-1)}, C_n = j\right) \quad \text{(for } j = 1, \ldots, m\text{)}.$$

Then, given state j, we denote the distribution of Z_n conditional on preceding observations by $f_{n|n-1,j}$. For the error we define

$$g_j\left(\epsilon\right) = g_{e_n}\left(\epsilon|C_n = j\right) \quad \text{(for } j = 1, \ldots, m\text{)}.$$

Then given state j, the probability distribution of the error is denoted by g_j. At the n-th duration, we refer to the m conditional distributions $f_{n|n-1,j}$ as the *state-dependent conditional distributions*, to the m conditional means $\mu_{n,j}$ as the *state-dependent conditional means*, and to the m distributions g_j as the *state-dependent error distributions* of the MSACD model.

2.6.2 Model fitting

We specify a MSACD model in three steps. In the first step, we specify the number of states of the Markov chain. In the next step, we specify the type of lagged conditional mean to be included in the model for $\mu_{n,j}$. The state-dependent conditional means $\mu_{n-1,j}, \mu_{n-2,j}, \ldots$ are used in this chapter. An alternative that we can use is the state-independent conditional means $\mu_{n-1}, \mu_{n-2}, \ldots$ (see the discussion in Hujer et al., 2002, p. 10). In the last step, we specify the conditional distribution of an ACD model to be used as the state-dependent conditional distribution of the MSACD model.

We call $\{Z_n\}$ an *exponential MSACD* model in the case that the exponential is specified for the state-dependent error distributions. We denote by $\mathrm{MSACD}(p, q)$ the case that the state-dependent conditional means $\mu_{n,j}$ have an $\mathrm{ACD}(p, q)$ specification. For example, a two-state exponential $\mathrm{MSACD}(1, 1)$ model is a MSACD model with t.p.m.

$$\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix},$$

the conditional mean in state j defined by

$$\mu_{1,j} = \frac{\omega_j}{1 - \alpha_j - \beta_j}$$

$$\mu_{n,j} = \omega_j + \alpha_j Z_{n-1} + \beta_j \mu_{n-1,j} \quad (\text{for } n = 2, 3, \ldots)$$

and the state-dependent conditional distributions defined by

$$f_{n|n-1,j}(z) = \frac{1}{\mu_{n,j}} \exp\left(-\frac{z_n}{\mu_{n,j}}\right) \quad (\text{for } j = 1, 2).$$

The likelihood. Assume that N consecutive observations $z_1, \ldots, z_N$ are generated by an $m-$state MSACD model with t.p.m. $\boldsymbol{\Gamma}$ and initial distribution $\boldsymbol{\delta}$ for its Markov chain, and conditional distribution of the specified ACD model as the state-dependent conditional density $f_{n|n-1,j}(z)$. Define $\mathbf{z}^{(n)} = (z_1, z_2, \ldots, z_n)$. The likelihood L_N of the MSACD model, conditioned on the first u observations, is given by

$$\begin{aligned} L_N &= f\left(z_{u+1}, \ldots, z_N \,|\, z_1, \ldots, z_u\right) \\ &= \prod_{n=u+1}^{N} f\left(z_n | \mathbf{z}^{(n-1)}\right). \end{aligned} \tag{2.23}$$

A derivation of the explicit formula for the likelihood (2.23) is given in Appendix B., pages 210-213.

In Equation (2.23) the conditional distribution of Z_n is a mixture conditional distribution

$$
\begin{aligned}
f\left(z_n|\mathbf{z}^{(n-1)}\right) &= \sum_{j=1}^{m} f\left(z_n, C_n = j \mid \mathbf{z}^{(n-1)}\right) \\
&= \sum_{j=1}^{m} \Pr\left(C_n = j \mid \mathbf{z}^{(n-1)}\right) \cdot f_{n|n-1,j}\left(z\right),
\end{aligned}
\tag{2.24}
$$

where the state-dependent conditional density $f_{n|n-1,j}\left(z\right)$ is the $j-$th component distribution and the probability that C_n is in state j, conditioned on the previous observations $\mathbf{z}^{(n-1)}$, denoted by $p_{n|n-1,j}$, is the probability weight assigned to the $j-$th component distribution. In this chapter we refer to the conditional distribution (2.24) as the conditional distribution of the MSACD model[24].

For events $n = (u+1)$ through to $n = N$, where u is the number of observations the likelihood is conditioned upon, the quantities $p_{n|n-1,j}$ and $f_{n|n-1,j}\left(z\right)$ required to evaluate the conditional distribution of Z_n are computed by Hamilton's (1994, p. 692) algorithm. This algorithm recursively computes the distribution of C_n conditioned on $\mathbf{z}^{(n-1)}$ using a set of two equations (for $j = 1, 2, \ldots, m$). Define $\mathbf{z}^{(0)}$ as the empty set. The first equation (his Equation 22.4.5), which can

---

[24]Hujer et al. (2002, p.9) refer to conditional distribution (2.24) as the "marginal density," or as the "incomplete data density" in their description of the EM algorithm for MSACD models. However, Equation (2.24) is the marginal for the joint distribution of Z_n and C_n conditional on the preceding observations.

be described as the updating equation, is

$$\Pr\left(C_n = j\,|\,\mathbf{z}^{(n)}\right) = \frac{f\left(z_n\,|\,C_n, \mathbf{z}^{(n-1)}\right)\Pr\left(C_n\,|\,\mathbf{z}^{(n-1)}\right)}{f\left(z_n\,|\,\mathbf{z}^{(n-1)}\right)} \tag{2.25}$$

and the second equation (his Equation 22.4.6), which can be described as the prediction equation, is

$$\Pr\left(C_{n+1} = j\,|\,\mathbf{z}^{(n)}\right) = \sum_{i=1}^{m}\Pr\left(C_{n+1} = j\,|\,C_n = i\right)\Pr\left(C_n = i\,|\,\mathbf{z}^{(n)}\right). \tag{2.26}$$

A derivation of Equations (2.25) and (2.26) is given in Appendix B., pages 207-210. Note that the denominator of the expression on the right-hand side of Equation (2.25) is the conditional distribution of Z_n

$$f\left(z_n\,|\,\mathbf{z}^{(n-1)}\right) = \sum_{j=1}^{m} f\left(z_n\,|\,C_n = j, \mathbf{z}^{(n-1)}\right)\Pr\left(C_n = j\,|\,\mathbf{z}^{(n-1)}\right), \tag{2.27}$$

which represents the contribution of z_n to the likelihood of $\{z_n\}$.

Now, we substitute Equation (2.27) into Equation (2.25) and then rewrite the resulting equation in terms of the notations introduced earlier, the updating equation takes on the form

$$p_{n|n,j} = \frac{p_{n|n-1,j}\cdot f_{n|n-1,j}(z)}{\sum_{s=1}^{m} p_{n|n-1}^{(s)}\cdot f_{n|n-1,s}(z)} \quad (\text{for } j = 1, 2, \ldots, m). \tag{2.28}$$

Similarly, rewriting Equation (2.26) in terms of the notations introduced earlier; the prediction equation takes on the form

$$p_{n+1|n,j} = \sum_{i=1}^{m} \gamma_{ij} \cdot p_{n|n,i} \quad (\text{for } j = 1, 2, \ldots, m). \tag{2.29}$$

Equations (2.28) and (2.29) represent the set of equations given in Hujer et al. (2002, p. 10).

From Equation (2.24), it is easy to show that the marginal conditional mean is a probability-weighted average of the state-dependent conditional means, i.e.

$$\mu_n = \sum_{j=1}^{m} \Pr\left(C_n = j \mid \mathbf{z}^{(n-1)}\right) \cdot \mu_{n,j}. \tag{2.30}$$

Hujer et al. (2002, p. 39) further show that the distribution of the error, conditional on the preceding observations, is a mixture, i.e.

$$g\left(\epsilon_n \mid \mathbf{z}^{(n-1)}\right) = \sum_{j=1}^{m} \Pr\left(C_n = j \mid \mathbf{z}^{(n-1)}\right) \cdot g\left(\epsilon_n \mid C_n = j\right). \tag{2.31}$$

From Equation (2.31), some properties of the error process for the MSACD model can be described as follows. Firstly, the errors of the MSACD model are not independent, since the distribution of the error depends on previous observations. Secondly, the assumption that the conditional error given preceding observations has a mean of one is satisfied, since the mean of the state-dependent error distribution is one by construction. Finally, the variance of the error changes from duration to duration. For a sketched proof of properties two and three, see Hujer et al. (2002, p. 39).

In general, the series of observations to which we fit an MSACD model consists of durations series in successive trading days. Here we assume that the component series are independent of each other, since we recompute the durations in each day (see Section 2.2.1). Thus the likelihood of the observations is a product of the likelihoods of the components.

Maximisation of the likelihood to estimate parameters. The parameters are estimated by maximising the likelihood of the MSACD model. For

this purpose Hujer et al. (2002) used both a numerical maximisation algorithm and the *Expectation-Maximisation* (EM) algorithm. In order to minimise the programming effort for each modification of the model (ZM, p. 72), we estimate the parameters by *direct numerical maximisation* (DNM) of the likelihood[25]. To estimate the parameters this way several problems need to be addressed. The main problems are the constraints on the parameters and the existence of multiple local maxima in the likelihood function.

Constraints on the parameters. Since elements of the t.p.m. are probabilities and the ACD model is a strictly positive-valued process, maximisation of the likelihood of a MSACD model is a constrained optimisation problem. The observed duration series is typically long and the estimated off-diagonal elements of the t.p.m. tend to lie close to the lower boundary value of zero, thus maximisation using a constrained optimizer can be slow (ZM, p. 47). However, in order to use **R**'s unconstrained optimizer `nlm` for maximisation of the likelihood, the optimization problem needs to be first converted to an unconstrained one.

In general there are three groups of constraints: those that apply to the parameters of the state-dependent conditional means, those that apply to the parameters of the Markov chain and those that apply to the parameters of the state-dependent error distributions[26]. In the case of a Burr MSACD with state-

---

[25]Hujer et al. (2002, p. 22) also used DNM when the state-dependent conditional mean depends also on lagged state-independent conditional mean. However, they preferred the EM algorithm for other specifications of the state-dependent conditional mean (Hujer et al., 2002, p. 13).

[26]A state-dependent error distribution that is a standard exponential does not have any parameters to be estimated.

dependent conditional means that are a linear function of one lagged duration and one lagged state-dependent conditional mean, the constraints for each group of parameters are:

1. The constant term ω_i in the state-dependent conditional mean must be strictly positive, for $i = 1, \ldots, m$. The coefficient of both the lagged duration term α_i and lagged state-dependent conditional mean term β_i in the state-dependent conditional mean must be nonnegative and sum of the coefficients must be strictly less than one, for $i = 1, \ldots, m$.

2. The parameters of the state-dependent error distribution κ_i and σ_i^2 must be strictly positive and $\sigma_i^2 < \kappa_i$, for $i = 1, \ldots, m$.

3. The rows of the t.p.m. $\boldsymbol{\Gamma}$ must sum to 1, and all the parameters γ_{ij} must be in the interval $[0, 1]$.

The implementation difficulties caused by the constraints on the parameters of this MSACD model can be avoided by reparameterising the model. The t.p.m. $\boldsymbol{\Gamma}$ is reparameterised as described in ZM (pp. 48-49). As an illustration, starting values for parameters γ_{ij} of a two-state Markov chain are reparameterised to the "working parameters" by the following transformations

$$\tau_{ij} = \log\left(\frac{\gamma_{ij}}{\gamma_{ii}}\right) \quad (\text{for } i \neq j).$$

The estimates of the working parameters returned by the unconstrained optimizer

are reparameterised back to the "natural parameters" by the following transformations

$$\hat{\rho}_{ij} = \begin{cases} \exp\left(\hat{\tau}_{ij}\right) & \text{(for } i \neq j) \\ \\ 1 & \text{(for } i = j) \end{cases}$$

$$\hat{\gamma}_{ij} = \frac{\hat{\rho}_{ij}}{\hat{\rho}_{i1} + \hat{\rho}_{i2}}.$$

The constraints on parameters of state-dependent conditional means may be divided into two types: the strictly positive constraints on the constant terms ω_i, and the nonnegative elements that sum to less than one constraints on the coefficients of lagged duration term α_i and of lagged state-dependent conditional mean β_i. The transformation of the parameter ω_i is relatively simple. Starting value for the constant term is transformed to the unconstrained parameter by $W_i = \log\left(\omega_i\right)$, for $i = 1, \ldots, m$. Estimates returned by the unconstrained optimizer is transformed back to the constrained parameter by $\hat{\omega}_i = \exp\left(\hat{W}_i\right)$. The constraints on the parameters α_i and β_i (for $i = 1, \ldots, m$) are similar to those on the parameters of $\mathbf{\Gamma}$, except that the sum of α_i and β_i cannot equal to one. Thus, the transformations of the parameters α_i and β_i use a slight modification of the reparameterisation of the $\mathbf{\Gamma}$ as described in ZM (pp. 48-49). In state i we define an additional parameter, denoted by γ, by

$$\gamma_i = 1 - \left(\alpha_i + \beta_i\right).$$

In this definition we assume that the value of β_i is likely to be close to 1[27]. Then,

---

[27]This can be interpreted as the existence of strong persistence in the price durations, a finding of Hautsch (2004, p. 125) in his application.

place β_i followed by α_i and γ_i in a vector $\mathbf{d}_i$ defined by

$$\mathbf{d}_i = (d_{1,i}, d_{2,i}, d_{3,i})'.$$

so the parameters β_i, α_i and γ_i are renamed as $d_{1,i}$, $d_{2,i}$, $d_{3,i}$ respectively. The starting values for the constrained parameters α_i and β_i are transformed to the unconstrained real numbers $D_{v,i}$ defined by

$$D_{k-1,i} = \log\left(\tfrac{d_{k,i}}{d_{1,i}}\right) \quad (\text{for } k = 2, 3) \tag{2.32}$$

The estimates returned by the unconstrained optimizer are transformed back to the constrained parameters by

$$\hat{d}_{k,i} = \begin{cases} \frac{1}{1+\sum_{v=1}^{2}\exp\left(\hat{D}_{v,i}\right)} & (\text{for } k = 1) \\[2ex] \frac{\exp\left(\hat{D}_{k-1,i}\right)}{1+\sum_{v=1}^{2}\exp\left(\hat{D}_{v,i}\right)} & (\text{for } k = 2, 3) \end{cases} \tag{2.33}$$

The values of $\hat{d}_{1,i}$, $\hat{d}_{2,i}$ and $\hat{d}_{3,i}$ represent $\hat{\beta}_i$, $\hat{\alpha}_i$ and $\hat{\gamma}_i$ respectively. For the more general case that the state-dependent conditional mean is a linear function of p lagged durations $X_{n-1}, \ldots, X_{n-p}$ and q lagged state-dependent conditional means $\mu_{n-1,i}, \ldots, \mu_{n-q,i}$, there will be a total of $r = p + q + 1$ parameters (for $p, q \in \mathbb{N}$). The additional parameter becomes

$$\gamma_i = 1 - \left(\sum_{l=1}^{p}\alpha_{n-l,i} + \sum_{l=1}^{q}\beta_{n-l,i}\right),$$

and the vector of constrained parameters generalises to

$$\mathbf{d}_i = (d_{1,i}, \ldots, d_{q,i}, d_{q+1,i}, \ldots, d_{q+p,i}, d_{r,i},)'$$

so the parameters $\beta_{1,i}, \ldots, \beta_{q,i}, \alpha_{1,i}, \ldots, \alpha_{p,i}$ and γ_i are renamed as $d_{1,i}, \ldots, d_{r,i}$ respectively. Transformations to the unconstrained real numbers defined by Equation (2.32) now holds for $k = 2, \ldots, r$ and transformation back to the constrained parameters defined by Equation (2.33) now generalises to

$$
\hat{d}_{k,i} = \begin{cases} \frac{1}{1+\sum_{v=1}^{r-1} \exp(\hat{D}_{v,i})} & \text{(for } k = 1) \\ \frac{\exp(\hat{D}_{k-1,i})}{1+\sum_{v=1}^{r-1} \exp(\hat{D}_{v,i})} & \text{(for } k = 2, \ldots, r). \end{cases}
$$

The transformation of the state-dependent Burr distribution parameters κ_i and σ_i^2 proceed as follows, for $i = 1, \ldots, m$. Define constrained parameter vectors $\boldsymbol{\Delta}_1 = (\sigma_1^2, \ldots, \sigma_m^2)'$ and $\boldsymbol{\Delta}_2 = (\kappa_1 - \sigma_1^2, \ldots, \kappa_m - \sigma_m^2)'$. Then, starting values for $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ are transformed to unconstrained real numbers by $D_1 = \log(\boldsymbol{\Delta}_1)$ and $D_2 = \log(\boldsymbol{\Delta}_2)$. Estimates returned by the unconstrained optimizer are transformed back to the constrained parameter vectors by $\hat{\boldsymbol{\Delta}}_1 = \exp(\hat{D}_1)$ and $\hat{\boldsymbol{\Delta}}_2 = \exp(\hat{D}_2)$ from which the estimated state-dependent Burr distribution parameters are calculated by $\hat{\boldsymbol{\kappa}} = \hat{\boldsymbol{\Delta}}_1 + \hat{\boldsymbol{\Delta}}_2$ and $\hat{\boldsymbol{\sigma}}^2 = \hat{\boldsymbol{\Delta}}_1$.

Number of parameters requiring estimation. The likelihood of an MSACD model, as given by Equation (2.23), is a function of the parameters as well as two groups of quantities: those to start the recursions for the state-dependent conditional means and those to start the recursion for the conditional distribution of C_n given preceding observations. We reduced the number of quantities requiring estimation as follows. In the case that the conditional mean in state j has an $ACD(p,q)$ specification. We start the recursion by setting each of

the max (p, q) number of quantities to the unconditional mean (as given by Equation (2.11)). For example, in state j, the conditional mean in an $MSACD(1, 1)$ model is

$$\mu_{1,j} = E\left(Z_1 | C_1 = j\right) = \frac{\omega_j}{1 - \alpha_j - \beta_j} \quad \text{(for } j = 1, \ldots, m\text{)}.$$

For $p_{1|0,j} = \Pr\left(C_1 = j | \mathbf{z}^{(0)}\right)$, which is needed to start the recursion represented by Equations (2.28) and (2.29), we use

$$\mathbf{p}_{1|0} = \left(\Pr\left(C_1 = 1\right), \ldots, \Pr\left(C_1 = m\right)\right) = \boldsymbol{\delta},$$

where $\boldsymbol{\delta}$ is the stationary distribution of the Markov chain. Applying expressions in both cases to the likelihood, it is now a function of the parameters only.

A total of k independent parameters is required to be estimated. The k here is the sum of: m times $(1 + p + q)$ parameters for the state-dependent conditional mean, m times the number of parameters for the state-dependent error distribution and $(m^2 - m)$ parameters for the Markov chain[28]. Thus, one needs to provide k starting values for the numerical maximizer to locate a local maximum.

The most general MSACD model we consider in this chapter has two states, the Burr distributions for state-dependent errors and state-dependent conditional means that depend on lagged duration and lagged state-dependent conditional mean. We can treat this twelve parameter model as the result of progressive extensions of the baseline model (i.e. the model which assumes that the observa-

---

[28]Thus, for example, a two-state Burr MSACD(1, 1) model has a total of twelve parameters to be estimated, of which six come from the ACD(1, 1) specification for the state-dependent conditional means, four from the state-dependent Burr distributions and the remaining two parameters from the Markov chain.

m	Error	$\boldsymbol{\Gamma}$	$\boldsymbol{\theta}_\epsilon$	$\boldsymbol{\omega}$	k
1	Exponential	0	0	1	1
1	Burr	0	2	1	3
2	Exponential	2	0	2	4
2	Burr	2	2	2	6

Table 2.12: Number of parameters of MSACD models with no additional dependence on past observations.

tions are realisations of independent exponential random variables with a common mean). In the first extension we generalise the model for the error from the exponential to the Burr distribution. Given each distribution we then increase the number of states of the Markov chain from one to two. The types of MSACD models considered in the first two extensions are presented in Table 2.12. In the final extension we allow the means in the one-state models and the state-dependent means in the two-state models to depend on preceding observations. The types of models considered in the final extension are presented in Table 2.13. In order to fit the model at each level of generalisation we select starting values by assigning parameter estimates of model at the preceding level and providing values for the additional parameters.

Multiple local maxima in the likelihood. The likelihood of an MSACD model is a function of the parameters and it can have several local maxima. Our goal is to find the global maximum. Depending on the set of starting values used,

m	Error	$\boldsymbol{\Gamma}$	$\boldsymbol{\theta_\epsilon}$	$\boldsymbol{\omega}$	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	k
1	Exponential	0	0	1	1	0	2
1	Exponential	0	0	1	1	1	3
1	Burr	0	2	1	1	0	4
1	Burr	0	2	1	1	1	5
2	Exponential	2	0	2	2	0	6
2	Exponential	2	0	2	2	2	8
2	Burr	2	4	2	2	0	10
2	Burr	2	4	2	2	2	12

Table 2.13: Number of parameters of MSACD models with additional dependence on past observations.

the numerical maximizer can easily identify a local, but not the global, maximum. This problem applies also to estimation by the EM algorithm.

In order to improve the chances of finding the global maximum we use a range of starting values and then see whether the same maximum is reached each time (Hujer et al., 2002, p. 12; ZM, Section 3.4.2). For their EM algorithm Leroux and Puterman (1992, p. 550 and 553) generated many sets of starting values and then chose the set at which the largest maximised likelihood value was identified. We apply this idea to estimate the parameters of the MSACD models in three steps. In the first step, we generate several sets of plausible starting values and then we use each set to evaluate the likelihood. In the second step, we arrange these likelihood values in decreasing order of magnitude and select from these those starting values that correspond to the U (e.g., 20) largest likelihood values. In the last step, we use this subset of starting values for the maximisation, and see whether the same maximum is identified in each case. We now apply our method to maximise the likelihood of N consecutive observations assumed to be generated by a model which has k parameters.

For the $i-$th parameter we assign b_i (for $i = 1, 2, \ldots, k$) distinct values[29]. The range of these values is assumed to contain the estimate of this parameter. We generate a total of B sets of starting values by using the **R** function `expand.grid`[30]. The computational effort (in terms of time) required to maximise the likelihood

---

[29]We assign at least three separate values for each parameter.
[30]The B here represents the product of the b_i.

of a long series of observations is considerable, especially if the model has a large number of parameters. Using all sets of starting values that we generated for the maximisation is not feasible, since B becomes very large when k is large. One way to overcome this problem is to select a portion of the B sets with which we have a large chance of locating the global maximum. We do so by first evaluating the likelihood at each of the B sets and then we use r (for $r = 1, 2, \ldots$) largest values of likelihood as the criterion to select our substantially smaller number of sets of starting values. We refer to the r sets obtained this way as our "best" set, and these are the sets we use for the maximisation[31]. By setting this method as an algorithm we have an automatic procedure to select a set of starting values for a MSACD model.

Assignment of starting values. We started by fitting the $m-$state exponential MSACD$(0, 0)$ model. The parameters for this model consist of the m constant terms in the state-dependent conditional mean $\omega_1, \ldots, \omega_m$, and the $(m^2 - m)$ off-diagonal elements of the t.p.m.[32]. To fit a two-state model we assigned a common starting value of $\left(\frac{0.1}{m-1}\right)$ to all off-diagonal elements of the t.p.m.[33] and for each ω_i we assigned the following eleven values: 0.5, 1, 2, ..., 10. We shall see later that by including lagged terms in the state-dependent means the estimates of ω_i we obtained there are considerably small in relation to the range of

---

[31] We assume here that by initialising the iterations at a best starting values the numerical maximiser has a better chance to locate the global maximum than that at inferior starting values.

[32] For $m = 1$, this is just the single element 1.

[33] This implies that initially the Markov chain is equally likely to be in each state (ZM, Section 3.4.2)

starting values we used here.

For each parameter in addition to those for the model at the preceding level of generalisation we assigned starting values as follows[34]. For a given m the parameters of the Burr MSACD$(0,0)$ model consist of those for: the Markov chain, the state-dependent conditional means, and the state-dependent error distributions. We assigned the estimated t.p.m. and estimates of the constant terms in the state-dependent conditional means of the exponential MSACD$(0,0)$ as starting values for parameters of the Burr MSACD$(0,0)$ model. In order to generate starting values we have to assign values for each parameter of a state-dependent Burr distribution (κ_i, σ_i^2). By trial-and-error we found that the ranges of the following values for κ_i: 0.5, 1, 1.5, 2, 3, …,6 and for σ_i^2: 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4 contained $\hat{\kappa}_i$ and $\hat{\sigma}_i^2$, respectively.

For a given m we then extended the state-dependent conditional mean in an exponential and a Burr MSACD$(0,0)$ model to that depending on a lagged duration (i.e. an MSACD$(1,0)$ specification) and then also depending on a lagged state-dependent conditional mean (i.e. an MSACD$(1,1)$ specification). We assigned parameter estimates of an exponential MSACD$(0,0)$ model as the starting values for parameters of an exponential MSACD$(1,0)$ model. In order to generate starting values we have to assign values for each α_i. By trial-and-error we found that the range of the following values: 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, …, 0.5

---

[34]The number of values that we can assign is a compromise between the degree of accuracy and computational feasibility.

contained $\hat{\alpha}_i$.

Next we assigned the estimated t.p.m. of the exponential MSACD$(0,0)$ model as starting values for the t.p.m. in an exponential MSACD$(1,1)$ model. In order to generate starting values we have to assign values for each parameter of a state-dependent conditional mean $(\omega_i, \alpha_i, \beta_i)$. By trial-and-error we found that the ranges of the following values for each ω_i: 0.01, 0.05, 0.1, for each α_i: 0.01, 0.05, 0.1, 0.3, 0.5, 0.7 and for each β_i: 0.1, 0.3, 0.5, 0.7, 0.8, 0.9 contained $\hat{\omega}_i$, $\hat{\alpha}_i$ and $\hat{\beta}_i$, respectively. Starting values for the parameters of the Burr MSACD$(1,0)$ model and of the Burr MSACD$(1,1)$ model were chosen in a similar way.

Standard errors. Parametric bootstrap can be used to estimate the standard errors of the maximum likelihood estimators in the MSACD models as is done for the HMMs (see ZM, p. 55). On account of the considerable computational effort required (ZM, p. 53) we refrain from doing so in our applications.

2.6.3 Model selection and checking

In order to select an appropriate model and then to assess its out-of-sample fit, we divided an observations series into two sub-samples. We fitted the models to the first sub-sample (termed the *calibration sample*) and we examined a fitted model by using the second (termed the *validation sample*). In our applications the calibration sample covers roughly the first two-thirds of the sample period.

Model selection. After we fitted various types of MSACD model to the

calibration sample we need some criterion for model comparison. Following Hujer et al. (2002, p. 22) we use the Bayesian information criterion (BIC) for model selection.

Model checking. In the out-of-sample period we applied Diebold et al.'s (1998) method to assess the fit of the selected model. In their method the probability integral transform (PIT) of each observation is defined as

$$u_n = \Pr\left(Z_n \leq z_n \middle| \mathbf{z}^{(n-1)} \right).$$

The PITs are based on the conditional distribution (under the model) of Z_n given the preceding observations. If the selected model is correct, Diebold et al. (1998, pp. 867-868) showed that the series of PITs $\{u_n\}$ are independent and identically distributed with common distribution $U(0,1)$[35].

We assessed the distributional assumption as follows. Assume that the model is correct. Then in a histogram of the PITs the frequency that the values of u_n occurs in a bin is described by the binomial distribution. For this distribution the total number of independent Bernoulli trials is the length of validation sample and the constant probability of success is the bin width. Using the binomial as the null distribution we performed a hypothesis test in two ways. Firstly, we construct, at each bin, a 95% confidence band for the histogram estimator of the density value[36]. An estimate that lies outside this confidence band indicates that the selected model may not be adequate for the out-of-sample period. Secondly,

---

[35] The basic idea, however, dates back to at least Rosenblatt (1952).
[36] The construction uses a normal approximation to the binomial distribution.

we can carry out a χ^2 goodness-of-fit test[37].

We assessed the independence assumption as follows. Firstly, we plot the sample ACF for $\{u_n\}$ at lags of up to 50[38]. Secondly, we test the null hypothesis that the first fifty autocorrelations are zero[39] by computing the Ljung-Box test statistics[40]. A rejection of the null indicates that the selected model is inadequate to describe the serial correlation in the validation sample. The pattern of the autocorrelation coefficients may suggest the nature of this inadequacy.

2.7   Modelling price durations

Our objective is to assess whether the MSACD model can be used successfully to represent the times between price changes of a specified amount. For this purpose we used the transaction data for ALV, EOA and BAS. The periods covered are the months of March to May in the years 2002-2004.

The sm-M price duration is the variable we model. Its computation is described in Section 2.2. We applied the models to nine series of price durations. Price durations series have the following three features (see Section 2.3.5). Firstly, they are overdispersed relative to the exponential distribution. Secondly, they are positively serially correlated. Finally, they display intraday seasonal variation.

---

[37] For the calculation of the test statistics, the observed cell frequency is the observed frequency of u_n having value in a bin, the expected cell frequency is the expected value of the binomial distribution. Under the null hypothesis that $\{u_n\}$ is independent and identically $U(0,1)$, the test statistics has an asymptotic chi-squared distribution with $m-1$ degrees of freedom, where m denotes the number of bins in the histogram.

[38] As used in Table II of Hujer et al (2002, p. 23)

[39] Under the null hypothesis, the test statistics has a chi-squared distribution with fifty degrees of freedom and the hypothesis test is carried out at the usual 5% significance level.

[40] This is computed by the **R** function `Box.test`.

2.7.1 Preprocessing of observed durations

We found in Section 2.3.5 that the series of price durations display variation in day-to-day average and intraday seasonality. To standardise the mean we pre-processed the observations in two stages. In the first stage we used the method proposed by Engle and Russell (1997 and 1998) to remove the seasonal variation. In their method they fit a spline function to the entire series of observations and then they divide the series by the fitted function[41]. In the second stage we standardised the mean on each day by dividing the seasonally-adjusted series for each day by its average. The series we obtained here have an average of 1. In order to assess the influence of day-to-day variation of the mean on the forecast performance of the models, we used first the adjusted durations and then the standardised durations to examine the model.

2.7.2 Model fitting, selection and checking

We fitted a variety of MSACD models to the preprocessed series. The models we used can be divided into two groups: those that do not assume dependence at the observation level and those that assume additional dependence at the observation level. Within each group, we used models in the order of increasing number of parameters: one-state models followed by two-state models, models with ex-

---

[41] The application of Engle and Russell's method has the consequence that the average of the deseasonalised series is 1 only if it is taken over the entire period.

What is done in the literature is this. The deseasonalised series is split into an in-sample and an out-of-sample period. In this case the average for neither sub-periods are 1 (see, for example, Table I on Hujer et al., 2002, p. 22).

ponentially distributed errors followed by those with the Burr, and models with dependence on lagged duration only followed by those with dependence on both lagged duration and lagged conditional mean. In the last group we used models that assume no dependence at the observation level and then we used models that assume additional dependence at the observation level.

Parameter estimation by maximum likelihood. For each type of model used, we estimated its parameters by the method of maximum likelihood. We used direct numeric optimisation to maximise the incomplete-data likelihood. This is carried out by using **R** (version 2.9.0)'s unconstrained nonlinear optimisation routine `nlm`. To apply `nlm` we transform the parameters to its unconstrained equivalent, as described in Section 2.6.2. The starting values we uses was chosen from thirty possible sets of values as described in Section 2.6.2.

Model selection and checking. For each sequence of price durations, we compared the fitted models as described in Section 2.6.3. We checked the forecast performance of the selected model by using the preprocessed durations in the validation sample. This is described in Section 2.6.3.

2.7.3   Model fitted to fixed sample

In this application we used the price durations for March to May in the year 2004. For each of the three shares the observations over the entire three-month

period are deseasonalised[42].

Fixed sample of adjusted durations. We splitted the seasonally-adjusted series into two parts. The first part covered the first two months (i.e. the months of March and April), with which we used to fit the models. The second part covered the month of may, with which we used to check the selected model.

Table 2.14 presents the estimates of parameters in the model selected for each share. We note that firstly, the same model specification (i.e. a two-state Burr MSACD model) was chosen for each of the three fixed calibration samples of adjusted durations. Secondly, there is a noticeable difference between estimated "persistence" parameter in state 2, $\hat{\beta}_2$, for each of the three fixed calibration samples. In particular, $\hat{\beta}_2$ of the model chosen from calibration sample of adjusted durations for E.ON share is close to zero. Thirdly, $\hat{\sigma}_2$ of the model chosen from calibration sample of adjusted durations for ALV share is noticeably smaller (0.028) than those for EOA (0.187) and BAS (0.194). Lastly, $\hat{\gamma}_{12}$ of the model chosen from calibration sample of adjusted durations for ALV share is noticeably larger than those for EOA (0.030) and BAS (0.034).

Using the model chosen, we plot the histogram and the ACF of the forecast UPRs computed from the standardised durations of the ALV (Figure 2.11), the EOA (Figure 2.12) and the BAS (Figure 2.13) shares in May of 2004. In each of the three histograms, we plotted the 95% confidence interval (dotted lines) for

---

[42]The intraday seasonal component is estimated by a smoothing spline function with sixteen interior knots (correspond roughly to knots positioned at 30-minute intervals). To ensure that the spline function is non-negative we apply Equation (2.5) and then Equation (2.6).

$\hat{\boldsymbol{\theta}}$	BAS	EOA	ALV
$100 \cdot \hat{\omega}_1$	3.3	1.7	0.7
$10 \cdot \hat{\alpha}_1$	1.19	0.72	0.74
$\hat{\beta}_1$	0.857	0.915	0.922
$\hat{\kappa}_1$	1.087	0.975	1.015
$\hat{\sigma}_1^2$	0.231	0.165	0.167
$100 \cdot \hat{\omega}_2$	0.7	6.7	0.3
$10 \cdot \hat{\alpha}_2$	0.16	7.71	5.34
$\hat{\beta}_2$	0.821	0.002	0.442
$\hat{\kappa}_2$	0.880	0.850	0.807
$\hat{\sigma}_2^2$	0.194	0.187	0.028
$\hat{\gamma}_{12}$	0.034	0.030	0.100
$\hat{\gamma}_{21}$	0.251	0.189	0.263
code	1	1	1

Table 2.14: Adjusted price durations of BASF, E.ON and Allianz: MSACD model chosen for March to April, 2004.
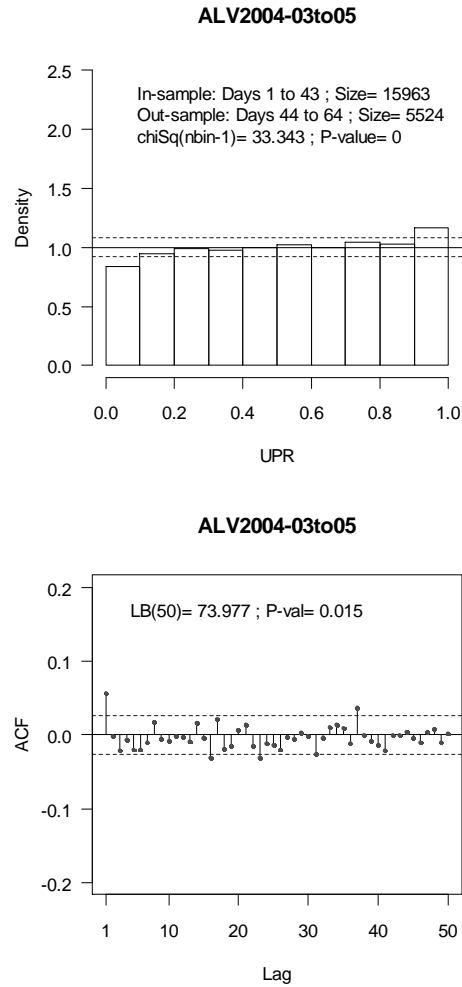
Figure 2.11: Fixed sample of adjusted price durations of Allianz in March to May, 2004: forecast UPRs of the two-state Burr MSACD(1,1) model. Histogram of observed UPRs (top) and ACF of observed UPRs (bottom).

EOA2004-03to05

In-sample: Days 1 to 43 ; Size= 14542
Out-sample: Days 44 to 64 ; Size= 6111
chiSq(nbin-1)= 44.685 ; P-value= 0

EOA2004-03to05

LB(50)= 49.088 ; P-val= 0.51
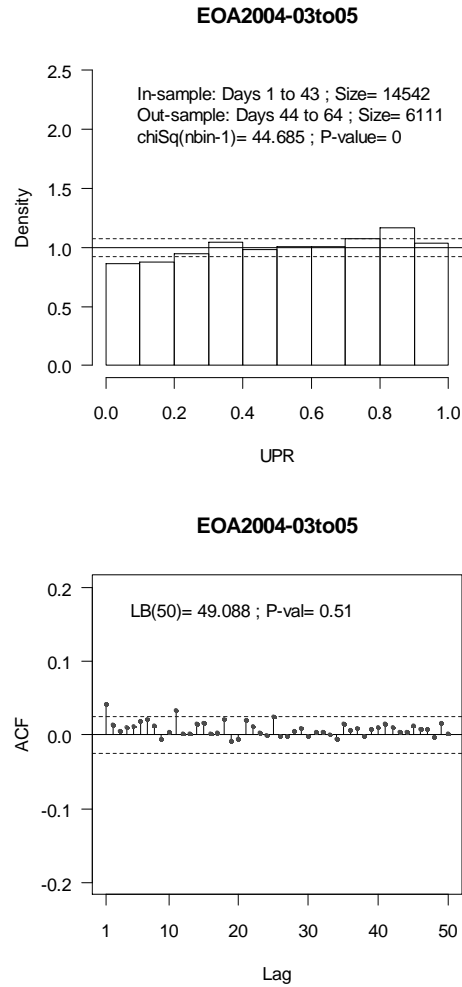
Figure 2.12: Fixed sample of adjusted price durations of E.ON in March to May, 2004: forecast UPRs of the two-state Burr MSACD(1,1) model. Histogram of observed UPRs (top) and ACF of observed UPRs (bottom).

BAS2004-03to05

In-sample: Days 1 to 43 ; Size= 9995
Out-sample: Days 44 to 64 ; Size= 3587
chiSq(nbin-1)= 28.559 ; P-value= 0.001

Density

UPR

BAS2004-03to05

LB(50)= 72.101 ; P-val= 0.022

ACF

Lag
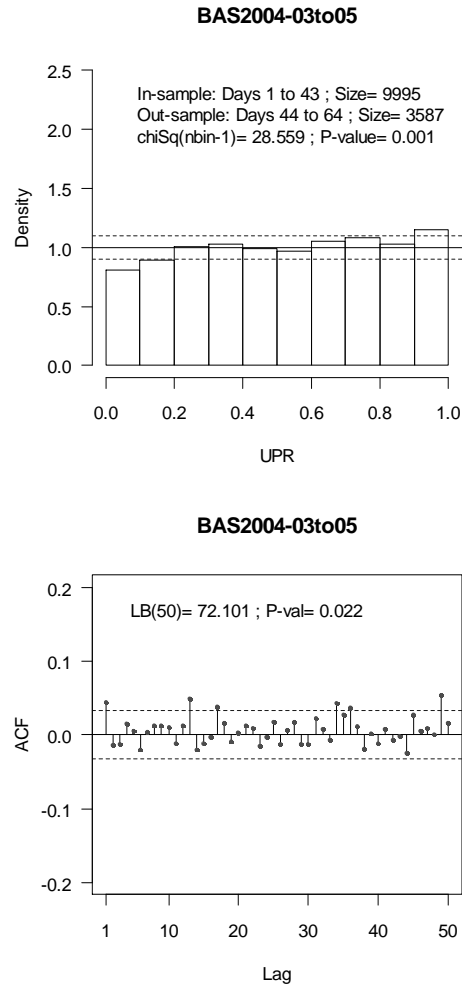
Figure 2.13: Fixed sample of adjusted price durations of BASF in March to May, 2004: forecast UPRs of the two-state Burr MSACD(1,1) model. Histogram of observed UPRs (top) and ACF of observed UPRs (bottom).

the estimated density function. In each of the three sample ACFs we plotted an approximate 95% confidence interval (dotted line) for the ACF of independent and identically distributed random variables.

From Figures 2.11 to 2.13 we note the following. Firstly, the confidence intervals for the estimated heights of the histogram are quite narrow due to the relatively long length of adjusted durations in the validation sample (about 5500 observations for ALV, 6000 for EOA and 3500 for BAS). Secondly, under each of the three chosen models, the distribution of the forecast UPRs are significantly different from a standard uniform distribution: p-values of the chi-squared statistics (with nine degrees of freedom) are all close to zero. A closer examination of the three histograms shows the general tendency for the departure: less forecast UPRs observed for very short durations (i.e. those that fell in the first two bins) and more forecast UPRs observed for the very long durations (i.e. those that fell in the last two bins) than those expected under the standard uniform. Finally, under the MSACD model chosen from adjusted durations for ALV and BAS, the observed Ljung-Box test statistics are significant at the 5% level.

Fixed sample of standardised durations. For each share, we used the standardised price durations in the first two months (i.e. March and April 2004) to choose the MSACD models and then the standardised price durations in the third month (i.e. May 2004) to check the chosen model. The model chosen for the "fixed" calibration samples of each share is given in Table 2.15. We note that the

$\hat{\boldsymbol{\theta}}$	BAS	EOA	ALV
$100 \cdot \hat{\omega}_1$	8.4	9.6	4.3
$10 \cdot \hat{\alpha}_1$	1.43	6.60	1.25
$\hat{\beta}_1$	0.793	0.000	0.844
$\hat{\kappa}_1$	1.082	0.857	0.992
$\hat{\sigma}_1^2$	0.233	0.212	0.178
$100 \cdot \hat{\omega}_2$	1.0	5.6	5.6
$10 \cdot \hat{\alpha}_2$	0.14	0.85	8.39
$\hat{\beta}_2$	0.805	0.869	0.000
$\hat{\kappa}_2$	0.870	0.984	0.849
$\hat{\sigma}_2^2$	0.124	0.170	0.165
$\hat{\gamma}_{12}$	0.034	0.198	0.030
$\hat{\gamma}_{21}$	0.280	0.032	0.193
code	1	1	1

Table 2.15: Standardised price durations of BASF, E.ON and Allianz: MSACD model chosen for March to April, 2004.

same model specification (i.e. a two-state Burr MSACD model) was chosen for each of the three fixed calibration samples of standardised durations. Secondly, there is a noticeable difference between estimated "persistence" parameter, $\hat{\beta}_i$, from each of the three fixed calibration samples. In particular, $\hat{\beta}_1$ of the chosen model for standardised durations of E.ON and $\hat{\beta}_2$ for standardised durations of ALV are zero (to three significant digits). Lastly, $\hat{\gamma}_{12}$ of the chosen model for standardised durations for EOA share is noticeably larger than those for BAS (0.034) and ALV (0.030). Similarly, $\hat{\gamma}_{21}$ for standardised durations of ALV share is noticeably larger than those for BAS (0.280) and EOA (0.032).

Using the model chosen, we plotted the histogram and the ACF of the forecast UPRs computed from the standardised durations of the ALV (Figure 2.14), the EOA (Figure 2.15) and the BAS (Figure 2.16) shares in May of 2004. From Figures 2.14 to 2.16, we note the following. Firstly, the confidence intervals for the estimated heights of the histogram are quite narrow. This indicates that the series over the out-of-sample period is considerably long (roughly 5 500 observations for ALV, 6 000 for EOA and 3 500 for BAS). Secondly, under the model chosen for ALV (top row, Figure 2.14) and for BAS (top row, Figure 2.16), there are insufficient evidence to reject the hypothesis that the forecast UPRs come from a standard uniform distribution: the probability of observed chi-squared statistics (with nine degrees of freedom) are 0.292 for ALV and 0.248 for BAS. However, a closer examination of the histograms for both of these two samples
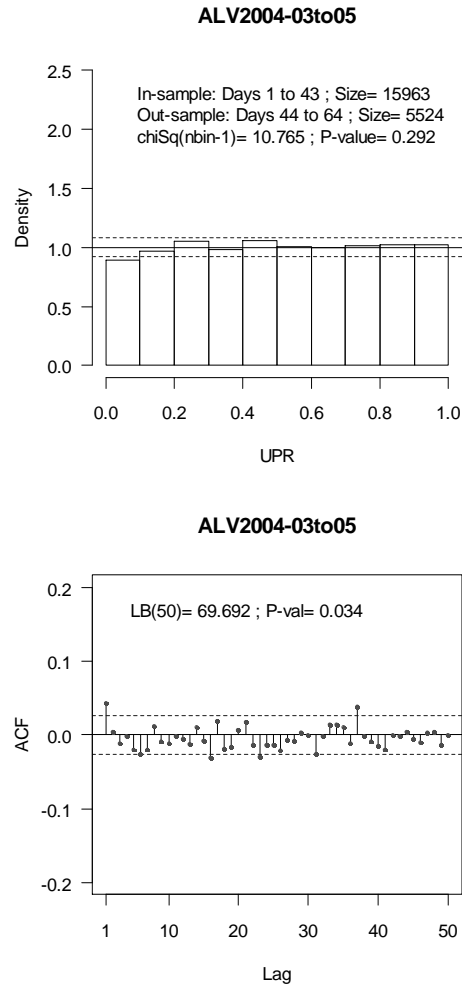
Figure 2.14: Fixed sample of standardised price durations of Allianz in March to May, 2004: forecast UPRs of the two-state Burr MSACD(1,1) model. Histogram of observed UPRs (top) and ACF of observed UPRs (bottom).

EOA2004-03to05

In-sample: Days 1 to 43 ; Size= 14542
Out-sample: Days 44 to 64 ; Size= 6111
chiSq(nbin-1)= 32.909 ; P-value= 0

EOA2004-03to05

LB(50)= 32.161 ; P-val= 0.977
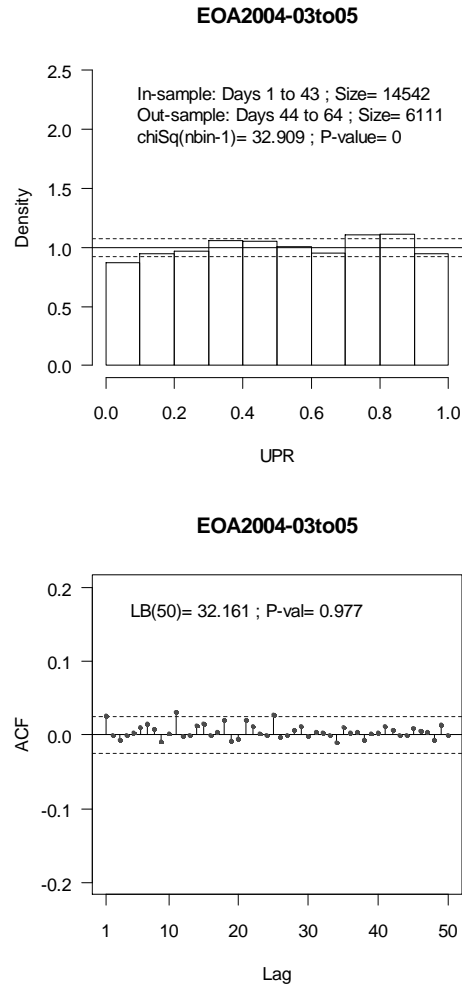
Figure 2.15: Fixed sample of standardised price durations of E.ON in March to May, 2004: forecast UPRs of the two-state Burr MSACD(1,1) model. Histogram of observed UPRs (top) and ACF of observed UPRs (bottom).
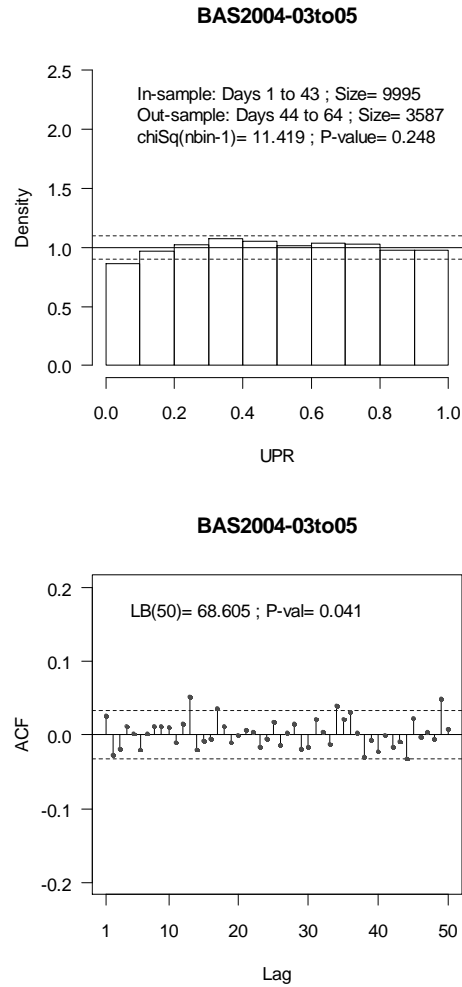
Figure 2.16: Fixed sample of standardised price durations of BASF in March to May, 2004: forecast UPRs of the two-state Burr MSACD(1,1) model. Histogram of observed UPRs (top) and ACF of observed UPRs (bottom).

shows that the forecast UPRs for very short durations (i.e. first bin from the left) are observed significantly less frequently than that expected under the standard uniform. Thirdly, under the model chosen from standardised durations for EOA (top row, Figure 2.15), the distribution of the forecast UPRs is significantly different from a standard uniform: probability of observed chi-squared statistics (with nine degrees of freedom) is zero. A visual inspection of the histogram shows that forecast UPRs for very short durations (i.e. first bin from the left) are observed significantly less frequently and for the very long durations (i.e. second and third bin from the right) are observed significantly more frequently than the standard uniform. Finally, under the MSACD model chosen for ALV (bottom row, Figure 2.14) and BAS (bottom row, Figure 2.16), the probabilities of observed Ljung-Box test statistics are less than 0.05: for ALV, the probability is 0.034 and for BAS, it is 0.041. Thus, there are insufficient evidence to support the hypothesis that the first fifty forecast UPRs are jointly independent.

Discussion. The variation in the level of daily mean has a considerable influence on the forecast performance of the model. By comparing our results from the analysis of forecast UPRs for all three series that we used, we found that, for the out-of-sample period, the selected model provided a better fit to the standardised observations than it does to the deseasonalised ones. We note this, for example, in our application to the price durations for ALV. In the test for uniform distribution the value of the chi-squared statistics we obtained for the

standardised durations (10.765 in top half of Figure 2.14) is substantially lower than that we obtained for the adjusted durations (33.343 in top half of Figure 2.11).

Our assessment here, however, is not based on strictly out-of-sample forecasts. What we have done instead (and what is done in the literature) is this. By preprocessing the series over the entire period we assumed that, for each day in the out-of-sample period, both the mean level and the seasonality were known in advance. In practice neither of these are available to us at the time of forecast. In order to determine whether the model, under such an assumption, fits the out-of-sample observations we need to base our assessment on forecasts in which both the daily mean and seasonality are also forecast.

2.7.4   Model fitted to rolling sample

Our interest here is to make strictly out-of-sample assessment of the forecast performance of the MSACD model. For this purpose we used a technique that we refer to as rolling sample with fixed window. In this technique we recalibrate the selected model regularly and then we use the fitted model after each recalibration to forecast observations over an out-of-sample period. In our applications the rolling sample covers sixteen days in which the in-sample period covers the first fifteen days and the out-of-sample period covers day 16.

We used the sm-M price durations for ALV, EOA and BAS. The periods

covered are March to May in the years 2002-2004. The three consecutive months

(i.e. March to May) cover 63 trading days in the year 2002, 62 days in 2003 and

64 days in 2004.

For each of the nine series investigated, only the first fifteen days were used to

select a model. In each of the subsequent rolling samples only the selected model

was used but it was recalibrated (i.e. the model parameters are reestimated[43]) in

each in-sample period of the rolling sample.

For each calibration sample we preprocessed the observations in two stages (see

Section 2.4.3) and then we fitted the model to the standardised durations. For

each validation sample we preprocessed the observations by using the seasonal

component fitted to the preceding fifteen days and the forecast for the level of

mean of adjusted durations[44]. We then computed the forecast pseudo-residual

of each observation in the standardised series so obtained. By examining these

residuals we assessed the forecast performance of the fitted model and we were

able to identify observations that are extreme relative to the model.

There are 48 rolling samples in the year 2002, 47 in 2003 and 49 in 2004.

For each calibration sample we used the termination code returned by `nlm` to

determine whether the optimiser has located a local maximum. We considered

---

[43] As starting values for the parameters we use the estimates of the parameters in the model fitted to the preceding calibration sample.

[44] The estimate use the value in trading day 15 of the exponential filter:

$$N_d = \rho \bar{a}_d + (1 - \rho) N_{d-1}$$

for $d = 2, 3, \cdots, 15$ and $N_1 = \bar{a}_1$. The weight $\rho \in [0, 1]$ is chosen by minimising the one-step-ahead forecast MSE.

the model as successfully fitted when a code of 1 is returned. For each validation sample we test whether the forecast UPRs are distributed $U(0,1)$ and whether they are independent by performing the tests as described in Section 2.6.3.

Comparison of models. Tables 2.16 to 2.18 present the estimates of parameters in the models selected for BAS, EOA and ALV for the first fifteen trading days in March, 2002-2004. From our model selection exercise for BAS we note the following. The optimisation method converged for all models we fitted to series for March 2003 and March 2004. But, for March 2002, the method diverged when we fitted the two-state Burr MSACD$(0,0)$ and two-state Burr MSACD$(1,1)$ models. In both these cases the likelihood values returned by the optimiser are not local maxima, and thus the best model for March 2002 was inconclusive.

From Table 2.16, we first compare the models selected for March 2002 and March 2003. We see that both selected models have a two-state Burr MSACD specification. However, the state-dependent component models have an ACD$(0,0)$ specification in March 2002, but an ACD$(1,1)$ specification in March 2003. Thus, a direct comparison of the state-dependent component processes chosen is not possible. However, the estimated probability of being in state 2 conditional on the Markov chain being in state 1 at the previous event has increased from 0.270 in March 2002 to 0.319 in March 2003. On the other hand, the estimated probability of C_n in state 1 conditional on C_{n-1} in state two has decreased from 0.113 in March 2002 to 0.033 in March 2003. In the model selected for March 2004 we see that

Year	**2002**	**2003**	**2004**
$100 \cdot \hat{\omega}_1$	30.7	13.6	7.2
$10 \cdot \hat{\alpha}_1$		0.48	1.17
$\hat{\beta}_1$		0.058	0.837
$\hat{\kappa}_1$	0.917	1.059	1.140
$\hat{\sigma}_1^2$	0.060	0.622	0.234
$100 \cdot \hat{\omega}_2$	129.1	10.2	2.2
$10 \cdot \hat{\alpha}_2$		1.19	0.23
$\hat{\beta}_2$		0.795	0.774
$\hat{\kappa}_2$	1.256	1.097	0.876
$\hat{\sigma}_2^2$	0.313	0.193	0.087
$\hat{\gamma}_{12}$	0.270	0.319	0.080
$\hat{\gamma}_{21}$	0.113	0.033	0.332
code	1	1	1
prop. fit	20/48	11/47	44/49
prop. sig.	14/48	4/47	7/49

Table 2.16: Standardised price durations of BASF: model chosen in trading days 1-15 of the years 2002-2004. Last three rows: termination code (code), fits to calibration sample (prop. fit), significant chi-squared statistics in validation sample (prop. sig.).

Year	**2002**	**2003**	**2004**
$100 \cdot \hat{\omega}_1$	17.8	3.6	12.3
$10 \cdot \hat{\alpha}_1$	1.07	0.72	6.66
$\hat{\beta}_1$	0.741	0.901	0.001
$\hat{\kappa}_1$	1.183	1.094	0.846
$\hat{\sigma}_1^2$	0.357	0.202	0.240
$100 \cdot \hat{\omega}_2$	15.3	13.9	6.7
$10 \cdot \hat{\alpha}_2$	5.91	4.80	1.03
$\hat{\beta}_2$	0.008	0.000	0.844
$\hat{\kappa}_2$	0.775	0.856	1.002
$\hat{\sigma}_2^2$	0.006	0.026	0.207
$\hat{\gamma}_{12}$	0.054	0.048	0.231
$\hat{\gamma}_{21}$	0.274	0.239	0.047
code	2	1	1
prop. fit	16/48*	8/47*	27/49*
prop. sig.	5/48	5/47	4/49

Table 2.17: Standardised price durations of E.ON: model chosen in trading days 1-15 of the years 2002-2004.

Year	**2002**	**2003**	**2004**
$100 \cdot \hat{\omega}_1$	9.9	1.9	5.5
$10 \cdot \hat{\alpha}_1$	0.95	0.27	8.70
$\hat{\beta}_1$	0.824	0.856	0.009
$\hat{\kappa}_1$	1.070	0.843	0.856
$\hat{\sigma}_1^2$	0.232	0.006	0.029
$100 \cdot \hat{\omega}_2$	8.2	3.8	4.2
$10 \cdot \hat{\alpha}_2$	4.23	0.70	1.26
$\hat{\beta}_2$	0.045	0.913	0.844
$\hat{\kappa}_2$	0.926	1.137	0.973
$\hat{\sigma}_2^2$	0.006	0.221	0.185
$\hat{\gamma}_{12}$	0.039	0.267	0.216
$\hat{\gamma}_{21}$	0.238	0.106	0.048
code	3	2	1
prop. fit	7/48	0/47	39/49
prop. sig.	7/48	7/47	7/49

Table 2.18: Standardised price durations of Allianz: model chosen in trading days 1-15 of the years 2002-2004.

the estimated coefficient of the lagged conditional mean in state 1, $\hat{\beta}_1$, is 0.837, and in state 2, $\hat{\beta}_2$, is 0.774. Thus, in both states, the dependence of the current conditional mean on the lagged conditional mean is strong (i.e. value of $\hat{\beta}_i$ is close to 1). This is evidence to support strong persistence in the standardised duration series even after allowing for two states.

From our model selection exercise for EOA we note that the optimisation method converged for all models we fitted to the series for March 2003. But, for March 2002, the method diverged for all two-state Burr MSACD specifications we fitted, thus the best model for March 2002 was inconclusive. From Table 2.17, the findings on the models chosen for standardised durations of EOA are as follows. Firstly, similar to findings made from Table 2.16, over the first fifteen trading days of March in each of the years 2002, 2003 and 2004, a two-state Burr ACD model was chosen. Secondly, an $ACD(1,1)$ specification has been chosen for the state-dependent component models in both March of 2002 and March of 2003. However, there is a noticeable difference between the component models chosen in March of 2002 and that chosen in March of 2003. In particular, the persistence of the component model in state 1, as indicated by $\hat{\beta}_1$, has increased from 0.741 in March 2002 to 0.901 in March 2003. The level of the component model in state 1, as indicated by $\hat{\omega}_1$, has decreased from 0.178 in March 2002 to 0.036 in March 2003. The estimated Markov chain in March 2002 ($\hat{\gamma}_{12} = 0.054$ and $\hat{\gamma}_{21} = 0.274$) is roughly the same as that in March 2003 ($\hat{\gamma}_{12} = 0.048$ and $\hat{\gamma}_{21} = 0.239$).

From Table 2.17, the findings on the model chosen for EOA in March 2004 are as follows. The component model in state 1 has negligible persistence ($\hat{\beta}_1 = 0.001$) but has a strong persistence in state 2 ($\hat{\beta}_2 = 0.844$). This is evidence that the level of persistence in the standardised duration is noticeably different between the two states.

Before the discussion of Table 2.18, we note the following from the model selection process. The estimation of all MSACD models converged in March 2002, but the estimation of the two-state Burr MSACD$(0, 0)$ did not satisfy the convergence criteria in March 2003. Thus, the best model in March 2002 was inconclusive. From Table 2.18, the findings on the models chosen for standardised durations of ALV are as follows. Firstly, similar to the findings made from Table 2.16, over the first fifteen trading days of March in each of the years 2002-2004, a two-state Burr ACD model was chosen. Secondly, an ACD$(1, 1)$ specification has been chosen for the state-dependent component models in both March of 2002 and March of 2003. Again, there was a noticeable difference between the component models chosen in March of 2002 and that chosen in March of 2003. The persistence of the component model in state 2, as indicated by $\hat{\beta}_2$, has increased from 0.045 in March 2002 to 0.913 in March 2003. The level of the component model in state 1, as indicated by $\hat{\omega}_1$, has decreased from 0.099 in March 2002 to 0.019 in March 2003. While in state 2, $\hat{\omega}_2$ decreased from 0.082 in March 2002 to 0.038 in March 2003. The estimated Markov chain in March 2002 indicates that the probability

of remaining in state 1 ($\hat{\gamma}_{12} = 0.039$) is higher than that of remaining in state 2. In March 2003, there was instead a higher probability of remaining in state 2 ($\hat{\gamma}_{21} = 0.106$). Thus, there was also a noticeable difference between the estimated Markov chains.

From Table 2.18, the findings on the model chosen for ALV in March 2004 are as follows. The component model in state 1 has negligible persistence ($\hat{\beta}_1 = 0.009$) but has a strong persistence in state 2 ($\hat{\beta}_2 = 0.844$). This is evidence that the level of persistence in the standardised duration was noticeably different between the two states.

Model checking: a case study. Of the nine sequences of price durations we used, the series for BAS for the year 2004 was the only one for which the model is adequate. In order to demonstrate the application of the model to the rolling samples, we discuss the results obtained in this application.

In the first rolling sample the two-state Burr MSACD$(1, 1)$ model was chosen. This model was recalibrated in each of the subsequent 48 rolling samples. We checked whether the numerical method used to maximise the likelihood converged in each rolling sample as follows. Firstly, we examine the termination code returned by `nlm` for the calibration sample. In the top half of Figure 2.17 we plotted the termination code for calibration samples 1-49. We note that the optimiser has located a local maximum (termination code of 1) in all but calibration samples 30-34. Secondly, we examine the number of iterations performed by `nlm`
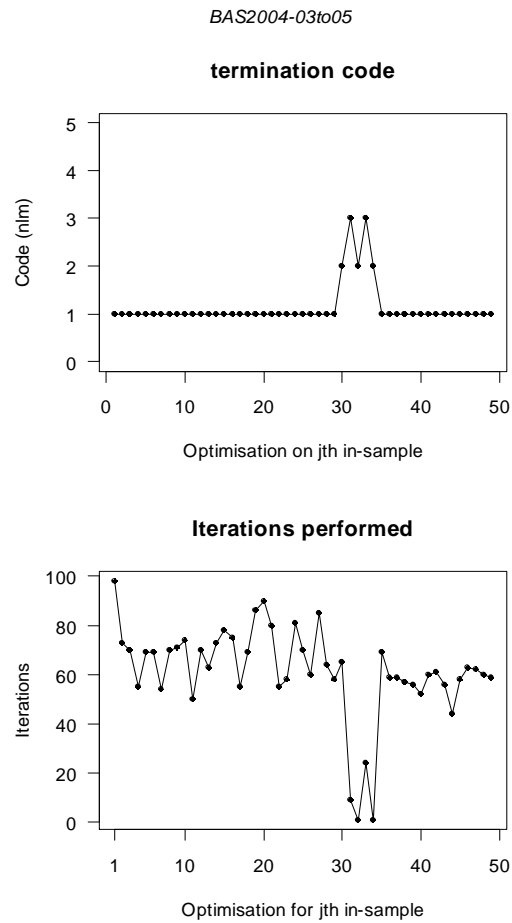
Figure 2.17: Standardised price durations of BASF in March to May, 2004: calibration of chosen model in samples 1-48. Termination code returned by `nlm` (top) and the number of iterations performed by `nlm` (bottom).

before the procedure terminated. In the bottom half of Figure 2.17 we plotted the number of iterations for calibration samples 1-49. In calibration samples 30-34 we note that the procedure terminated after only a few iterations. This indicates that different starting values are needed.

Using the fitted models in rolling samples 1-49 we plot the estimates of each parameter. Figure 2.18 displays the plots of the estimates of parameters in the state-dependent conditional distributions and Figure 2.19 displays the plots of the estimates of parameters of the Markov chain. In Figure 2.18 we note the following. Firstly, the estimates of β_i and σ_i^2 in state 1 and state 2 "cross" each other[45]. Secondly, in the case of $\hat{\beta}_i$, the variation in its value from one rolling sample to the next is considerable. Thirdly, when the optimisation did not converge in one rolling sample, then the optimisations usually do not converge in all of the subsequent rolling samples. We note this, for example, in the plot of $\hat{\beta}_i$ where in state 2 the values after rolling sample 34 vary only a little. These observations suggest that the series of standardised durations was still nonstationary[46]. The plots of estimates of parameters of the Markov chain show similar properties and thus are not presented here.

For each rolling sample we examined the UPRs of the observations in the validation sample. To determine whether these pseudo-residuals are $U(0,1)$ we applied the chi-squared test. A histogram of the values of the test statistics for

---

[45]This is not due to label switching.

[46]This suggests that the variation(s) in the level of daily mean and/or the intraday seasonal component are/is considerable.
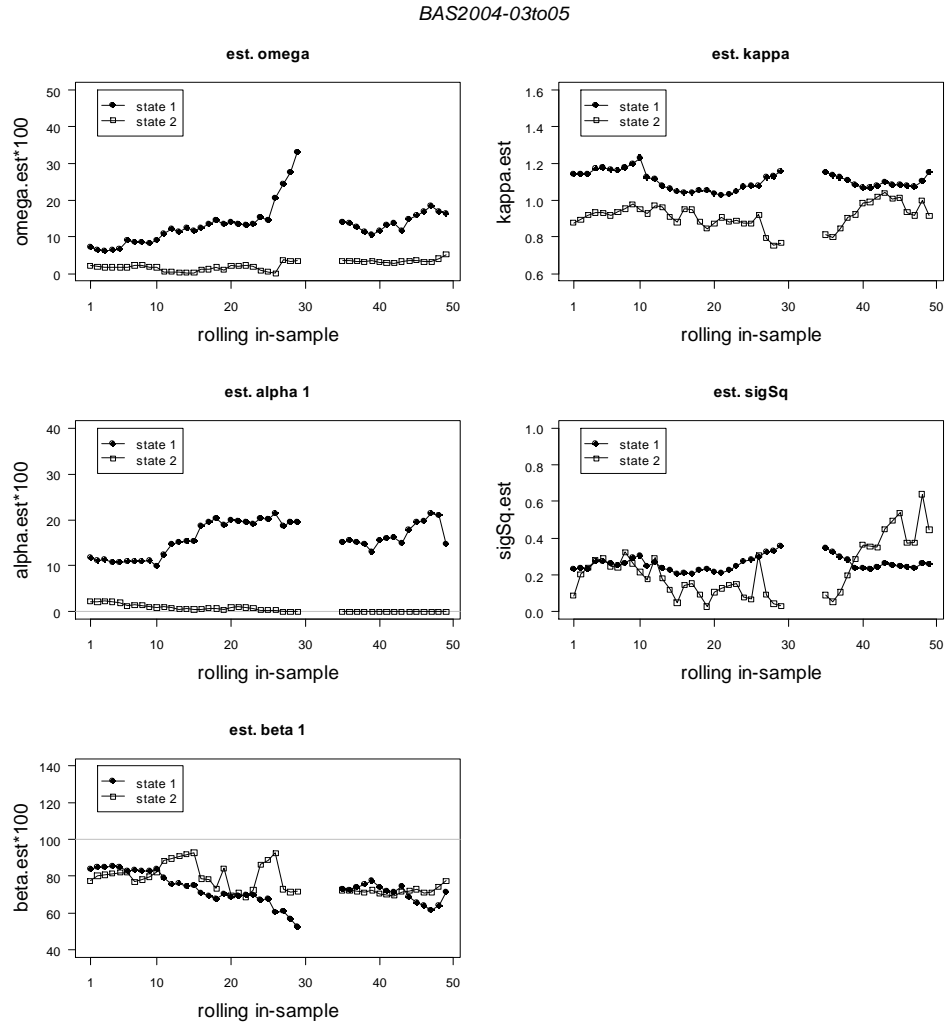
Figure 2.18: Standardised price durations of BASF in March to May, 2004: model fitted in calibration samples 1-48. State-dependent parameters of conditional mean (left column) and conditional distribution of duration.
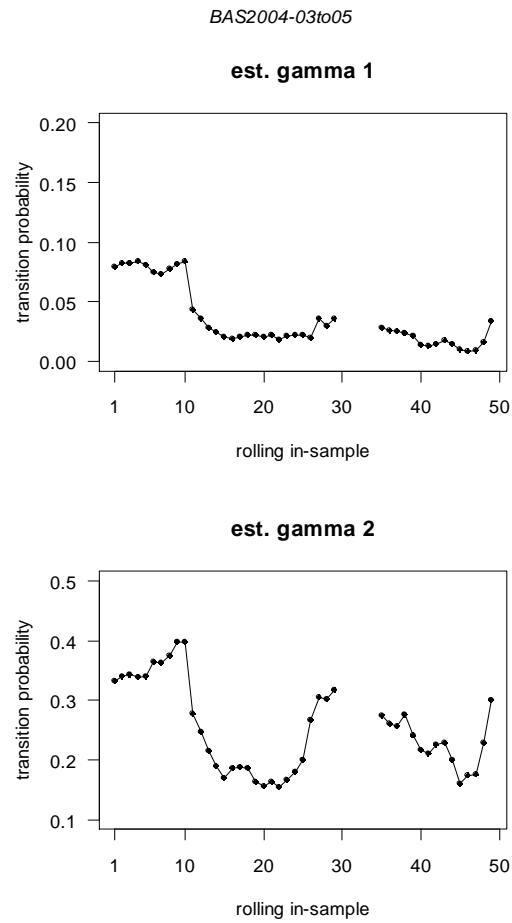
Figure 2.19: Standardised price durations of BASF in March to May, 2004: model fitted in calibration samples 1-48. Off-diagonal element in row 1 (top) and in row 2 (bottom) of the t.p.m..

validation samples 1-49 is presented in the top half of Figure 2.20. We note that two of these values exceed even 37 (i.e. the 99.9% percentile of the null distribution[47]). These allow us to identify the trading days on which the fitted model was not valid. We also checked whether the UPRs are independent of each other. For this purpose we applied the Ljung-Box test. A histogram of the values of the test statistics is presented in the bottom half of Figure 2.20. Again, the model is not valid for trading days on which the value exceed the 95-th percentile of the null distribution.

In the period March-May, 2004, the seven trading days on which the model was not valid are days 22, 25, 27, 42, 44, 45 and 48. In order to identify the source of the problem we assessed the fit of the model for each of these days by means of forecast pseudo residuals, as defined on page 97 of ZM. Their index plots are presented in Figure 2.21. In all of these plots we note that the means of the residuals series are not stationary (notably those for days 25 and 48). This suggests nonstationarity in the quantities that we obtained after we standardised the means for the seven days concerned. In all the plots in Figure 2.21 we see forecast pseudo residuals that are extreme. This indicates that the observations concerned are possible outliers relative to their predecessors. Figures 2.22 and 2.23 display the histograms and qq-plots of the forecast pseudo residuals. For all days we note that they deviate strikingly from the standard normal distribution.

Since we used the forecasts of intra-daily seasonal component and the daily

---

[47]That is, chi-squared with nine degrees of freedom.

BAS2004-03to05

Histogram of GOF test stats
H0: UPR is U(0,1)

7 of 49 obs. are sig. at 5% level

2 obs. are larger than 37

value of test stats

Histogram of Ljung-Box test stats
H0: First 50 ACs are zero

1 of 49 obs. are missing
7 of 48 obs. are sig. at 5% level
2 obs. are larger than 138

value of test stats
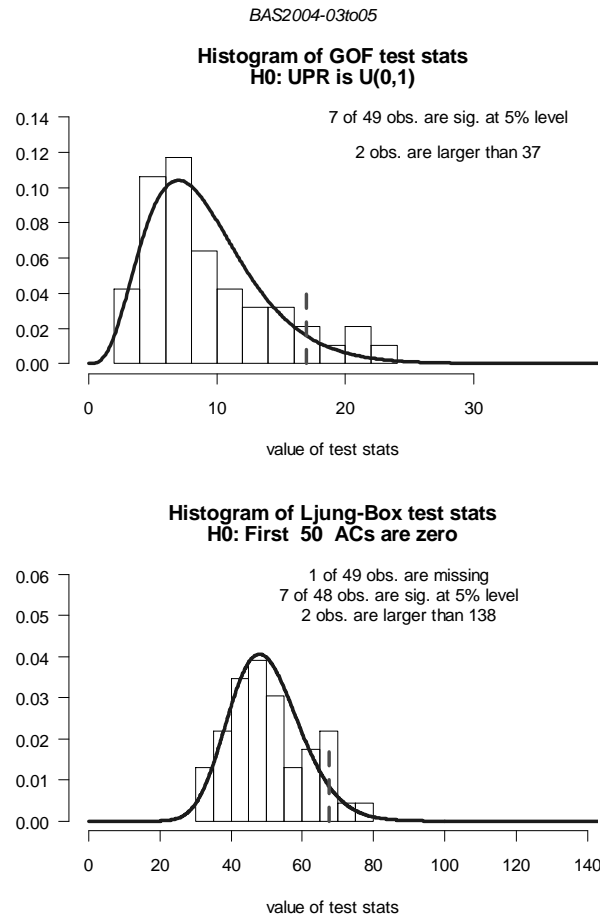
Figure 2.20: Observed price durations of BASF in March to May, 2004: distribution of chi-squared test statistics (top) and of Ljung-Box test statistics (bottom) of forecast UPRs. Solid line depicts chi-squared distribution with 9 d.o.f. (top) and chi-squared distribution with 50 d.o.f. (bottom). Vertical dashed line represents the 95th-percentile of the distribution.
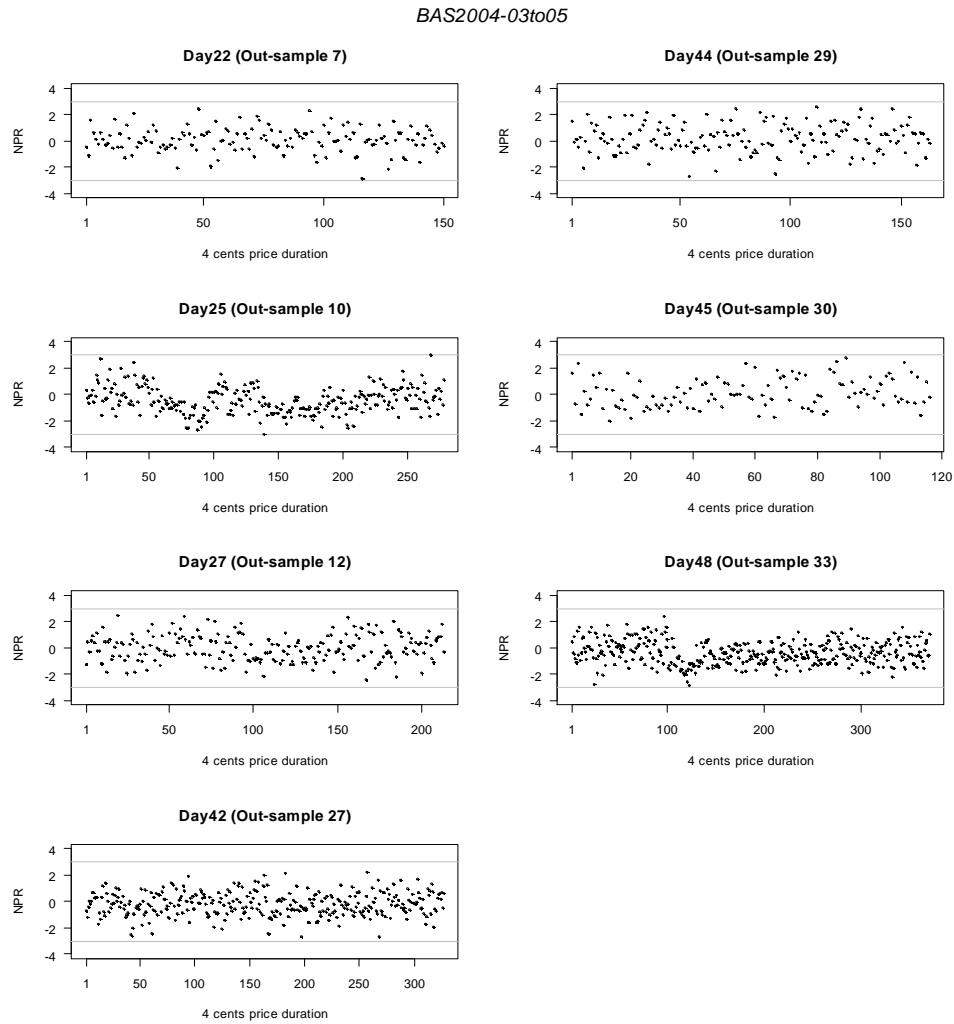
Figure 2.21: Forecast pseudo residuals for BAS in March-May, 2004: its index plots for the seven trading days in which the fitted model is not valid.
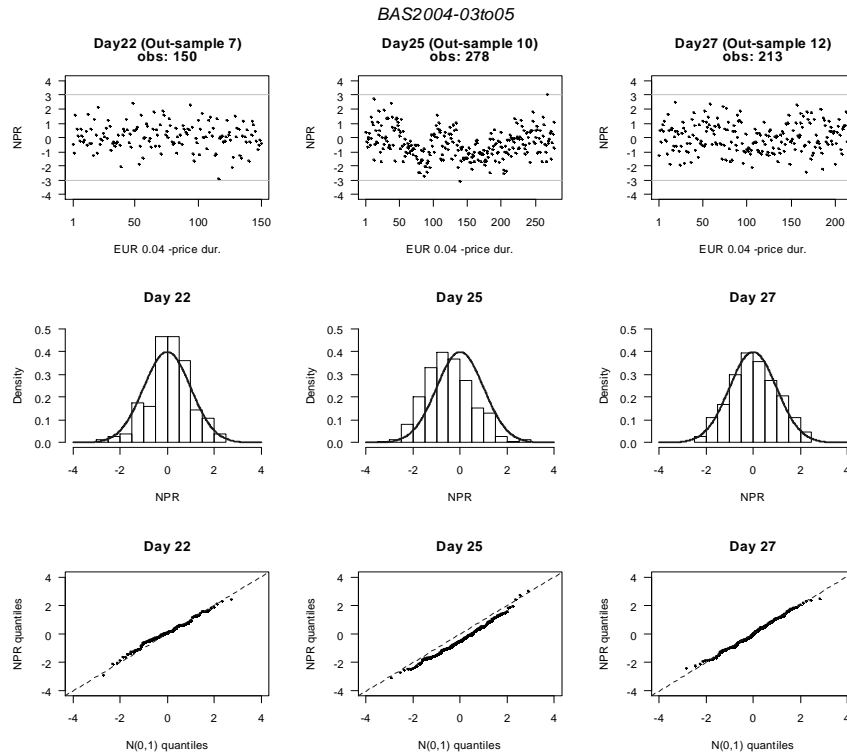
Figure 2.22: Forecast pseudo residuals for BAS in March-May, 2004: its plots for days 7, 10 and 12. Index plot (top row), histogram (middle row) and qq-plot (bottom row) are shown. Solid line depicts standard normal distribution (middle row) and dashed line depicts equality of sample and theoretcal quantiles.
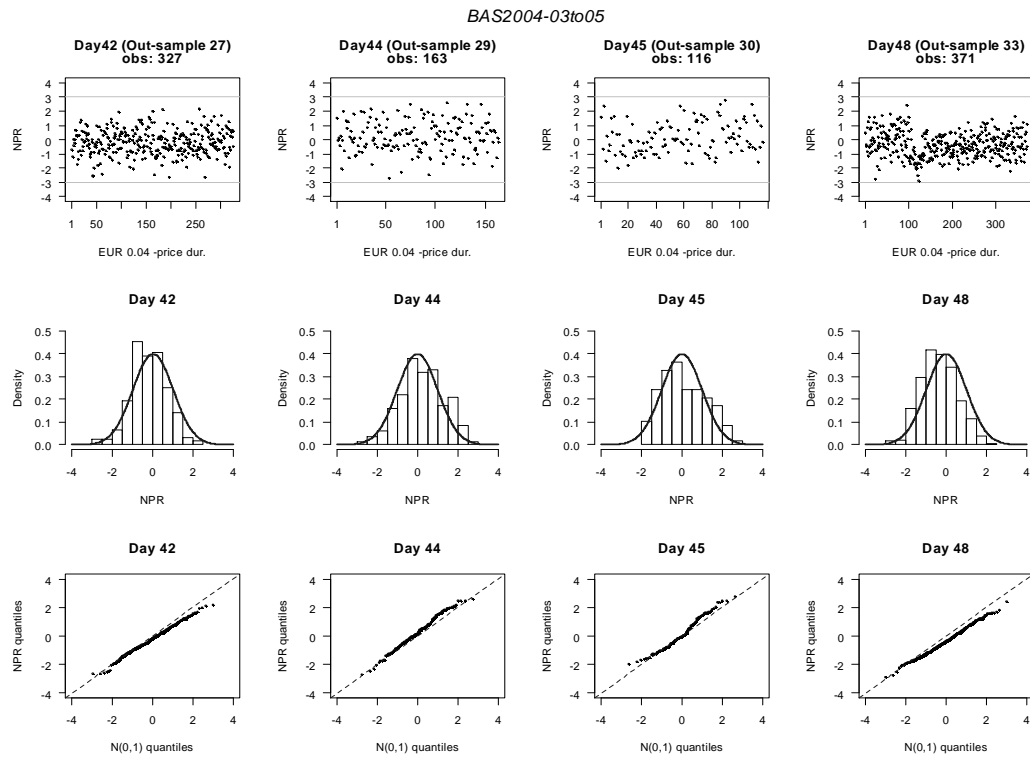
Figure 2.23: Forecast pseudo residuals for BASF in March-May, 2004: its plots for days 27, 29, 30 and 33.

mean to standardise the mean for the seven days concerned, we now assess the forecast performance of the model we used for each component. For this purpose the measure we used for the seasonal component is the sum-of-squares of prediction error[48], and for the mean level is the square of forecast error[49]. Figure 2.24 displays each of the two measures for validation samples 1-49. We note that both measures display considerable variation. This suggests that both the seasonal variation and the mean level are unstable. Furthermore, we found that the model generally performed poorly for days in which the seasonal variation observed deviates strikingly from its forecast. This suggests that variation in the seasonality has a strong detrimental effect on the accuracy of the model's forecasts.

We now examine observations that are extreme relative to the model. We found that large observations generally occurred in periods during which trading was inactive or the price varied slowly. The small observations, on the other hand, occurred in periods of rapid trading and substantial changes in the price.

Problems encountered. With the exception of the durations series for BAS in the year 2004, we encountered several problems in our applications. The main problems are as follows. Firstly, the appropriateness of the selected model was doubtful. In fitting the models to the calibration sample in the first rolling sample, we noticed convergence problems for some of the twelve types of models that we used. The likelihood values achieved in these cases are not necessarily

---

[48]We define this as the difference between the observation and the prediction provided by the spline function.

[49]We define this as the difference between the observed level and the level forecasted.
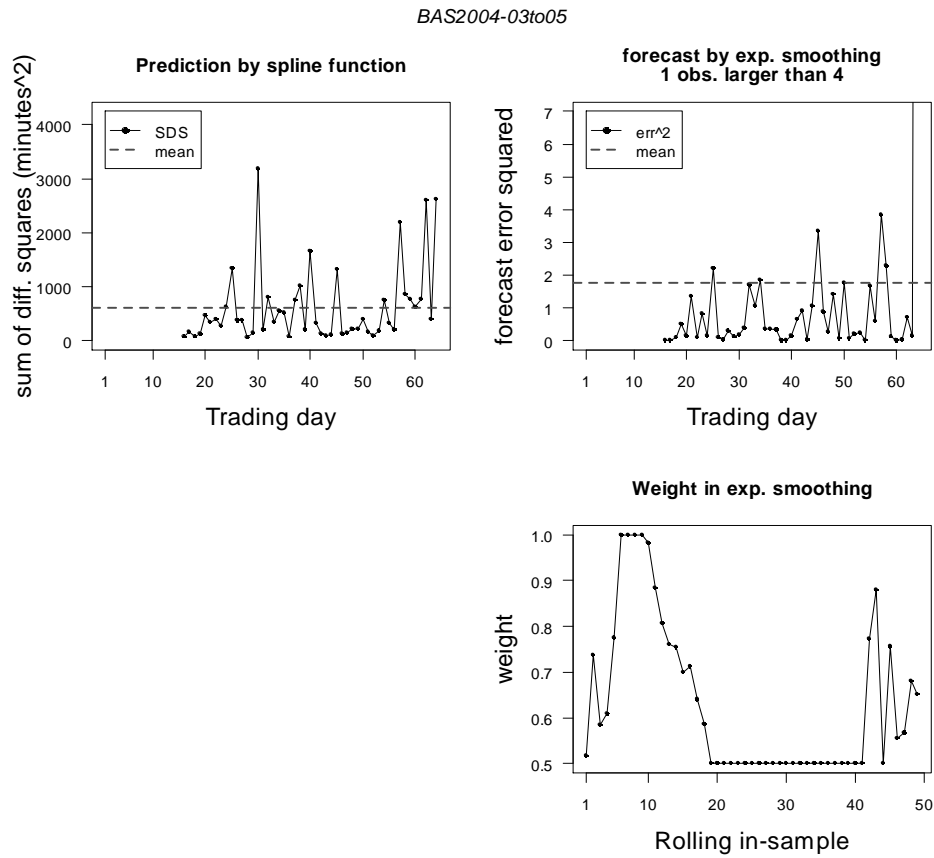
Figure 2.24: Validation samples for BAS in March-May, 2004: prediction accuracy of spline function (left column) and forecast accuracy of exponential filter (top right) for days 16-64. The parameter used in the exponential filter (bottom right) for calibration samples 1-49.

even local maxima. The models fitted are therefore not comparable by a model selection criteria. Secondly, the forecast performance of the selected model was poor in the subsequent rolling samples. For trading days in which the model performed markedly poorly (as indicated by its NPRs) we analysed the model. We found that the daily mean forecasts differ considerably from the observed values. Furthermore, the NPRs for these days still exhibit seasonal variation.

2.8 Summary and discussion

In this chapter it is the forecast performance of the MSACD model that we assessed. To achieve this we employed the rolling sample with fixed window in which we removed the forecasts for the nonstationary components in an out-of-sample period before we examined the model. Our approach differs from those that are commonly described in the literature in three important respects:

1. The selected model was assessed according to its forecast performance. Hautsch (2004, p. 106), for example, validated the fitted model on the calibration sample, i.e. the observations to which the model was fitted is used again for checking.

2. The forecast performance was assessed strictly out-of-sample. In the literature, for example Grammig and Maurer (2000, p. 28) and Hujer and Vuletić (2007, p. 644), the authors first standardised the mean of the entire observation series. They then split the resultant series into two sub-samples.

The proposed models were calibrated on the first sub-sample and then the fitted models were validated on the second. The application of such method has two consequences. Firstly, the averages of observations in both the calibration and in the validation samples are not 1. Secondly, it assumes that the intraday seasonality in the validation period is available at the time of forecast.

3. The forecasts (under the fitted model) were computed over a one-day period, and the model was recalibrated daily. In the literature, the model was calibrated once and then used to compute forecasts over a period of several days. This approach has the main weakness that the estimates of the parameters do not describe the observations that have become available after the time at which the model was fitted.

We found that the forecast performance of the MSACD model was in general poor. The source of the problem is that the instability of the seasonal pattern and the variation in the levels of daily mean were both considerable. This is reflected in the estimates of the parameters where considerable variation was displayed. The task of forecasting these components can be simplified. One way of doing so is to model observations at fixed time intervals instead and then to allow seasonality in the model we use.

CHAPTER 3

AN HMM FOR THE INTRADAY RETURNS

In Chapter 2 of this dissertation we applied the MSACD model to the price durations series. The instability of the intraday seasonality and the considerable variation of the daily average duration created the main difficulties when we attempted to forecast the price durations.

In this chapter we model the intraday returns. We calculated the returns at regular time intervals, and thus the number of returns observed in each trading day is a constant. In contrast the number of price durations varied from one trading day to the next; it has to be forecasted as does the intraday seasonality, in order to apply methods that are designed to model the standardised durations[1]. Two features of the intraday returns series are the occurrences of zero return[2] and the presence of intraday seasonality[3]. Both features need to be accommodated by the model.

One way to accommodate the occurrences of zero return in our model is as

---

[1]In the current literature it is assumed that the durations in the validation sample have already been adjusted for intraday seasonality. For the purpose of forecasting out-of-sample, however, this assumption does not hold, because neither the intraday seasonal variation nor the number of transactions in the validation sample are known in advance. Thus, we need to forecast both variables.

[2]i.e. no share price change in the time interval. The frequency of such events increases when the share price is observed at a higher frequency during the trading day.

[3]This feature is well-documented (see, for example, Andersen and Bollerslev, 1997).

follows. We model the returns using two components; the first models the presence or the absence of price change and the second models the level of return observed when a price change does occur. Each component of the univariate process is modelled separately[4]. For each component of the return we employ the predicted volume to explain the seasonality, since our aim is to forecast the intraday return volatility. Our objectives in this chapter are firstly, to find a relationship between the returns and the predicted volumes series, secondly, if such a relation exists, to build a model for the returns, and lastly, to apply this model to forecast the return volatility.

In order to build the models for the returns and then to assess whether they can be used successfully to forecast the return volatility, the transaction data for eight shares traded on the XETRA were investigated. We computed the intraday returns as $100 \cdot \log \left(\frac{p_t}{p_{t-1}} \right)$, where p_t is the share price at the end of intraday time interval t. The models were applied to the returns on the following shares: Allianz (ALV), E.ON (EOA), Badische Anilin- und Soda-Fabrik (BAS), Lufthansa (LHA), Touristik Union International (TUI), Deutsche Bank (DBK), DaimlerChrysler (DCX) and Bayer Schering (SCH). The data investigated cover the following two-month periods in the year 2004: March-April, May-June, August-September, and November-December.

---

[4]This model is different from that of Engle and Russell (2005), who represented the transaction price series and the series of duration until the transaction as a bivariate process.

3.1 Analysis of the intraday returns

In this section we investigate the relationship between intraday return and intraday volume traded. We use data for three actively traded shares for this; these are (in decreasing level of activity): ALV, EOA and BAS. The data cover four separate months in the year 2004: March, May, August and November.

In each of the trading days investigated we observed the price and the volume traded at regular intervals. Let p_t denote the price on intraday interval t and v_t denote the volume traded on intraday interval t. We define p_t as the last recorded price in interval t and we calculate v_t as the sum of shares traded in interval t. In our investigations we examined the pairs of observations $\{(V_t, r_t)\}$, where $V_t = \log(v_t)$ and r_t denotes the return on intraday time interval t. In the case of zero volume, V_t was defined as zero[5]. In what follows we refer to V_t as "the volume on interval t," or simply as "volume."

We performed five separate investigations in this section. We began each of our investigations by analysing the data for ALV. We chose to start with the data for the month of May, since the variation in daily returns is moderate. In that month we analysed the data twice: we examined first the data for three separate days and then for the entire month. In our daily analysis of data we used the quartiles of the daily total numbers of intraday trades to select the days as follows: firstly, we analysed the day for which the total number of trades is at the

---

[5] For investigations in this section, however, we deleted observations for intervals over which no trade took place.

median and then we analysed the days for which the total numbers are at the first

and the third quartiles. We continued our investigation by analysing the data for

the other three months. We concluded our investigation by analysing the data for

the remaining two shares for each of the four months.

3.1.1 Dependence of the squared return on volume

We first investigate the relationship between squared return, denoted by r_t^2

and volume V_t. We used the cross correlation function to measure the relationship

between these two series, which we now define.

Let (X_t, Y_t) be a stationary bivariate time series. The cross covariance function

at lag k is defined as $\gamma_{XY}(k) = Cov(X_t, Y_{t+k})$ (for $k = 0, \pm1, \pm2, \cdots$). The

cross correlation function at lag k is then defined as $\rho_{XY}(k) = \frac{\gamma_{XY}(k)}{\sigma_X \cdot \sigma_Y}$, where

$\sigma_X = \sqrt{Var(X_t)}$ and $\sigma_Y = \sqrt{Var(Y_t)}$.

Let $\{(x_t, y_t)\}$ denote n pairs of observations. A sample estimate of $\gamma_{XY}(k)$ at

lag k, given in Box et al. (1994, p. 411, Eq. 11.1.4), is:

$$\hat{\gamma}_{xy}(k) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}) & (\text{for } k = 0, 1, 2, \cdots) \\ \frac{1}{n} \sum_{t=1}^{n+k} (y_t - \bar{y})(x_{t-k} - \bar{x}) & (\text{for } k = 0, -1, -2, \cdots), \end{cases}$$

where $\bar{x}$, $\bar{y}$ denote the sample means of $\{x_t\}$ and $\{y_t\}$, respectively. A sample

cross correlation function at lag k is then defined as $\hat{\rho}_{xy}(k) = \frac{\hat{\gamma}_{xy}(k)}{s_i \cdot s_o}$, where $s_x =$

$\sqrt{\hat{\gamma}_{xx}(0)}$ and $s_y = \sqrt{\hat{\gamma}_{yy}(0)}$. In order to obtain a useful estimate of $\hat{\rho}_{xy}(k)$, a

minimum of 50 pairs of observations should be used (see the example in Box et

al., 1994, p. 412). We used the **R** function `ccf` to compute $\hat{\rho}_{xy}(k)^6$.

We begin our investigation by analysing the bivariate time series of V_t and r_t^2 for ALV on 19 May 2004. The total trades on this day is the median of the daily totals of trades in the month of May. Figure 3.1 displays $\{V_t\}$, $\{r_t^2\}$ and the sample cross correlation function for these two series. Both series are calculated at observation intervals of thirty seconds, one minute and five minutes. In Figure 3.1 we note the following four features. Firstly, there is a clear intraday-seasonality in both $\{V_t\}$ and $\{r_t^2\}$: the level of each series reaches a high over the first and the last thirds of the day, and it drops to a low over the second third. Secondly, there is a positive relation between $\{V_t\}$ and $\{r_t^2\}$, since both series attain their highest and lowest levels over the same period of the day. Thirdly, the cross correlation coefficient at lag zero is considerably larger than those at all other lags. There are significant cross correlation coefficients at negative lags (i.e. the dependence of r_t^2 depend on V_t) and at positive lags (i.e. the dependence of V_t on r_t^2). The sample cross correlation function does not decrease exponentially towards zero with increasing lags[7]. Finally, longer sampling intervals lead to less significant cross correlation coefficients. These four features apply in general to all three

---

[6]Let $\rho_{xy}(k)$ denote the sample cross correlation function at lag k that is defined in the **R** function `ccf`. Using the expression for $\hat{\gamma}_{xy}(k)$ the formula for $\rho_{xy}(k)$ can be written as:

$$\rho_{xy}(k) = \begin{cases} \hat{\gamma}_{yx}(k) & (\text{for } k = 0, 1, 2, \cdots) \\ \hat{\gamma}_{xy}(-k) & (\text{for } k = \cdots, -2, -1, 0). \end{cases}$$

[7]We are interested in the cross correlation at lag zero. In examining the cross correlation functions for our applications we have taken into account the fact that the volume series are not stationary.
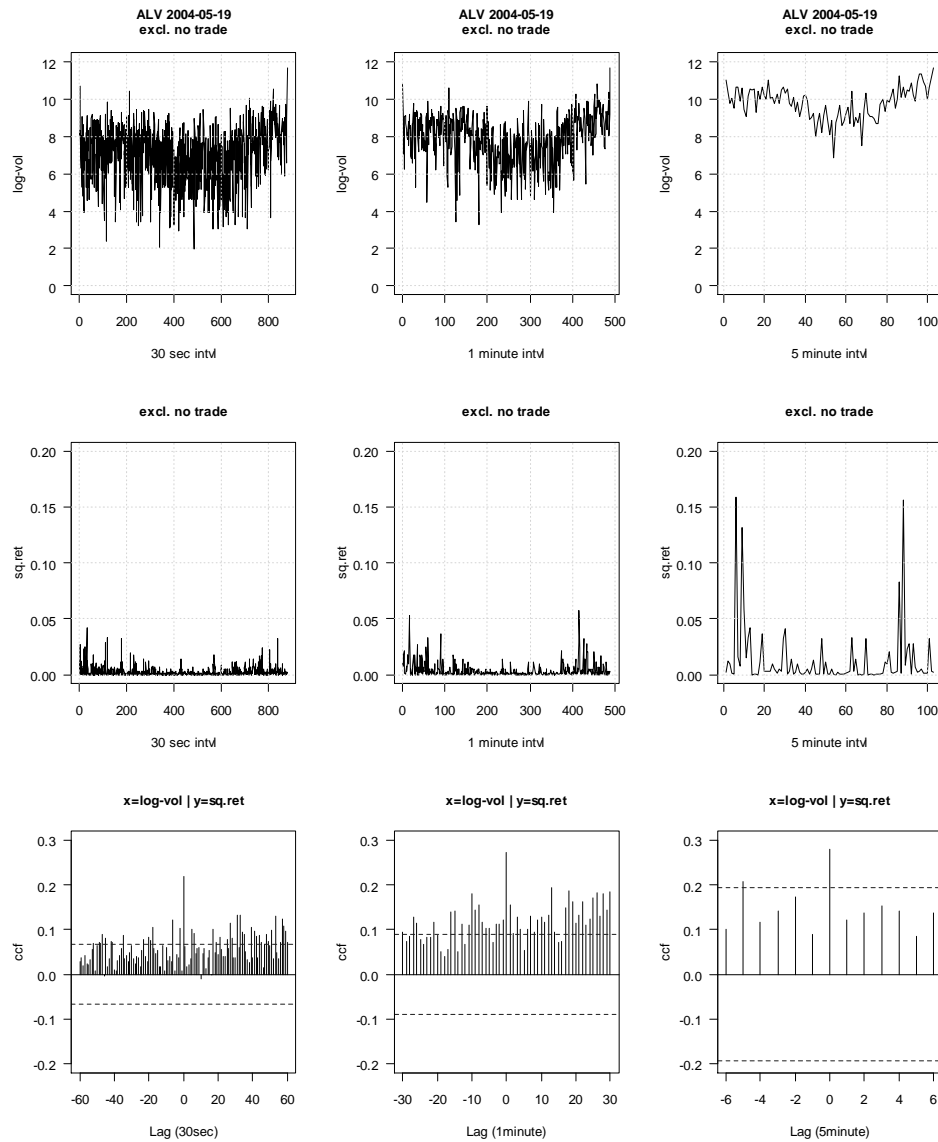
Figure 3.1: ALV in 2004-05-19: log-volumes (top), squared returns (middle) and cross-correlation function (bottom). Observation intervals used: 30 seconds (left column), one minute (middle column) and five minute (right column).

shares for each of the four months we examined.

In our assessment of the cross correlation between $\{V_t\}$ and $\{r_t^2\}$ the two main problems are as follows. Firstly, we need an objective way to select the length of the sampling interval, since it affects the sample cross correlation function (see Section 3.1.2). Secondly, in order to construct a forecast model for $\{r_t^2\}$ we need to have the volume information available at the time of the forecast (see Section 3.1.3).

3.1.2 Selection of observation interval

We need to select a regular interval to compute the observations, because we saw in Section 3.1.1 that the observation interval used affects the size of the cross correlation at lag zero. We then apply the selected observation interval to compute all observations.

Let x_t be the binary variable that represents the occurrence $(x_t = 1)$ or the absence $(x_t = 0)$ of a price change on intraday interval t. For each bivariate time series $\{(V_t, r_t)\}$ we first compute the binary series from the returns, where $r_t \neq 0$ represents that a price change has occurred. We then calculate $\{x_t\}$ for a variety of observation intervals. We select the fixed interval such that in the binary series calculated at this interval roughly 50% of the observations have zero value. We exclude here binary series in which the proportion of zero observations is less than 25%, as well as intervals that are shorter than 15 seconds.

Figure 3.2 displays the proportions of no price change for Allianz that we obtained daily by using observation intervals of fifteen and thirty seconds for each of the months March, May, August and November in the year 2004. We note in this figure that when we computed the intraday returns at intervals of thirty seconds the daily proportions that we obtained for each of the months fluctuate closer to 0.25 than those that we obtained when a 15-second interval was used. However, by using the 30-second interval, the daily proportions for all of the other shares we used fluctuate closer to 0.5 than those for Allianz. On account of this we used the 30-second interval to compute all observations.

3.1.3 Prediction of the volume

In Section 3.1.1 we saw that the intraday volume series $\{V_t\}$ display a clear seasonal variation. In order to forecast returns using volumes as a covariate, we need to have V_t available at the time of forecast. Here we are interested in making a 30-second-ahead forecast. One way to obtain a prediction for V_t, denoted by N_{t-1}, is by the exponential filter:

$$N_t = \alpha V_t + (1 - \alpha) \, N_{t-1} \quad (\text{for } t = 2, 3, \ldots, T), \tag{3.1}$$

where $\alpha \in [0, 1]$ denotes the weight[8] on the current observation. We start this procedure by setting $N_1 = V_1$. In what follows we refer to N_{t-1} as the predicted volume.

---

[8]This is also called the "smoothing parameter."
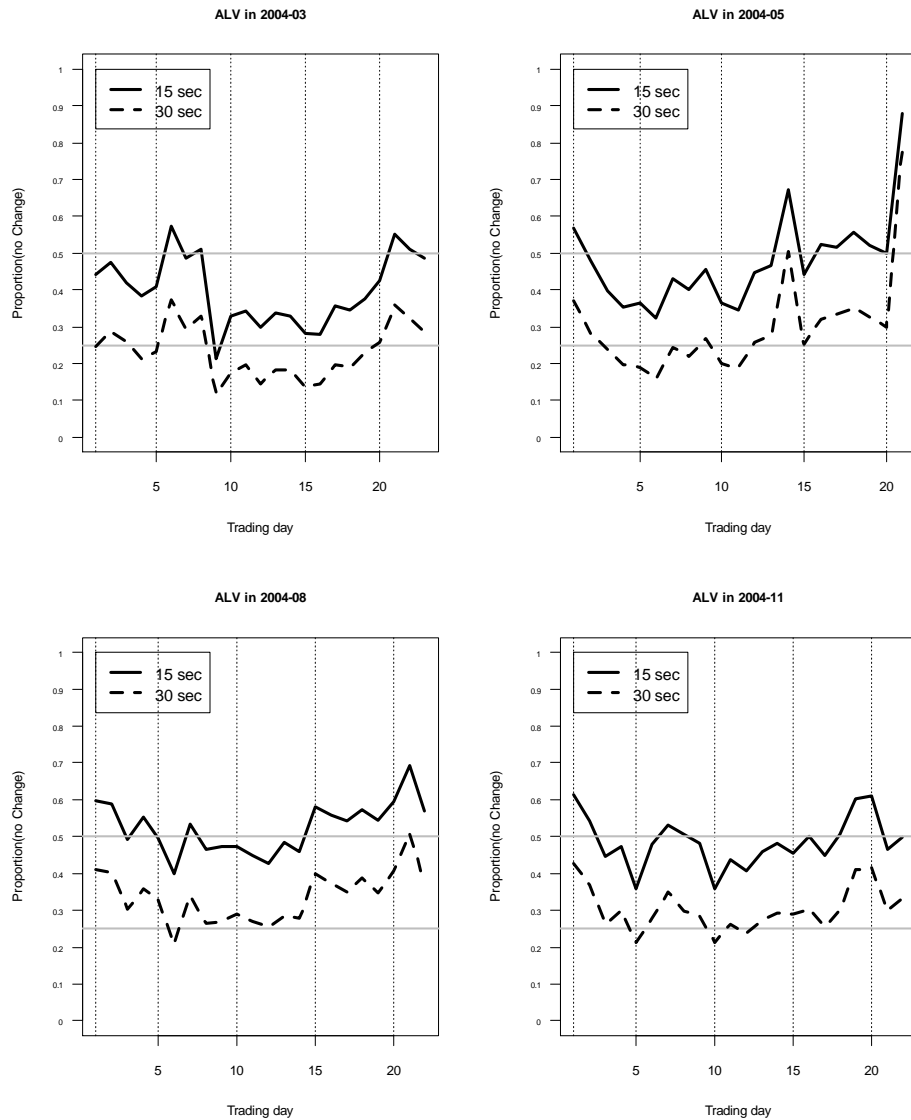
Figure 3.2: Price change indicators of Allianz in March, May, August and November, 2004: daily proportion of no price change when price change indicators are calculated at different observation interval.

In Equation (3.1) we note that α is a parameter that needs to be supplied. Our aim now is to investigate the relation between return and lagged volume, thus we need to determine which value of α best suits this purpose. We used the following method to select α. We first compute, at the selected sampling interval, the binary series that represents the occurrences of price change $\{x_t\}$. We then compute, at the same sampling interval, the series of predicted volumes $\{N_t\}$ for a variety of values of α. For each $\{N_t\}$ computed at an α we fit a logistic regression of the binary variable on the predicted volume[9]:

$$logit\left[\Pr\left(X_t = 1|N_{t-1}\right)\right] = \beta_0 + \beta_1 N_{t-1}.$$

We select the α that leads to the lowest AIC.

We applied the exponential filter, using the selected α, to the volumes series for each of the three shares investigated. Figures 3.3 and 3.4 display, respectively, the α selected daily in each of the four months investigated, for volume series observed at 15$-$ and at 30$-$ second intervals for ALV. Comparing Figure 3.3 with Figure 3.4 we note that, at an observation interval of 30 seconds, an α of 0.05 is a reasonable fixed value to compute predicted volumes for ALV for all four months investigated.

---

[9]This is computed by the **R** function `glm`.
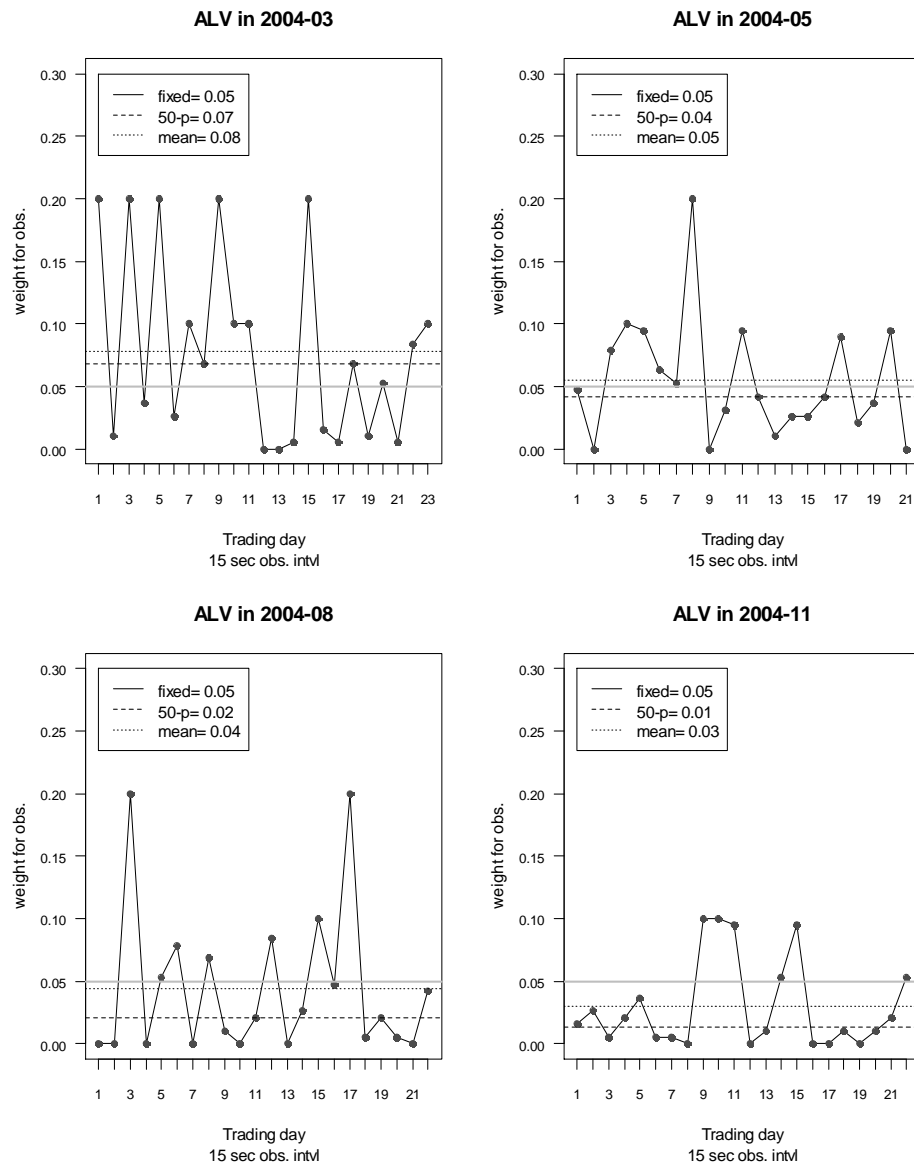
Figure 3.3: Allianz in March, May, August and November, 2004: weight for observation used by daily best logistic regression of price change indicator on predicted log-volume. Observations at 15 second interval.
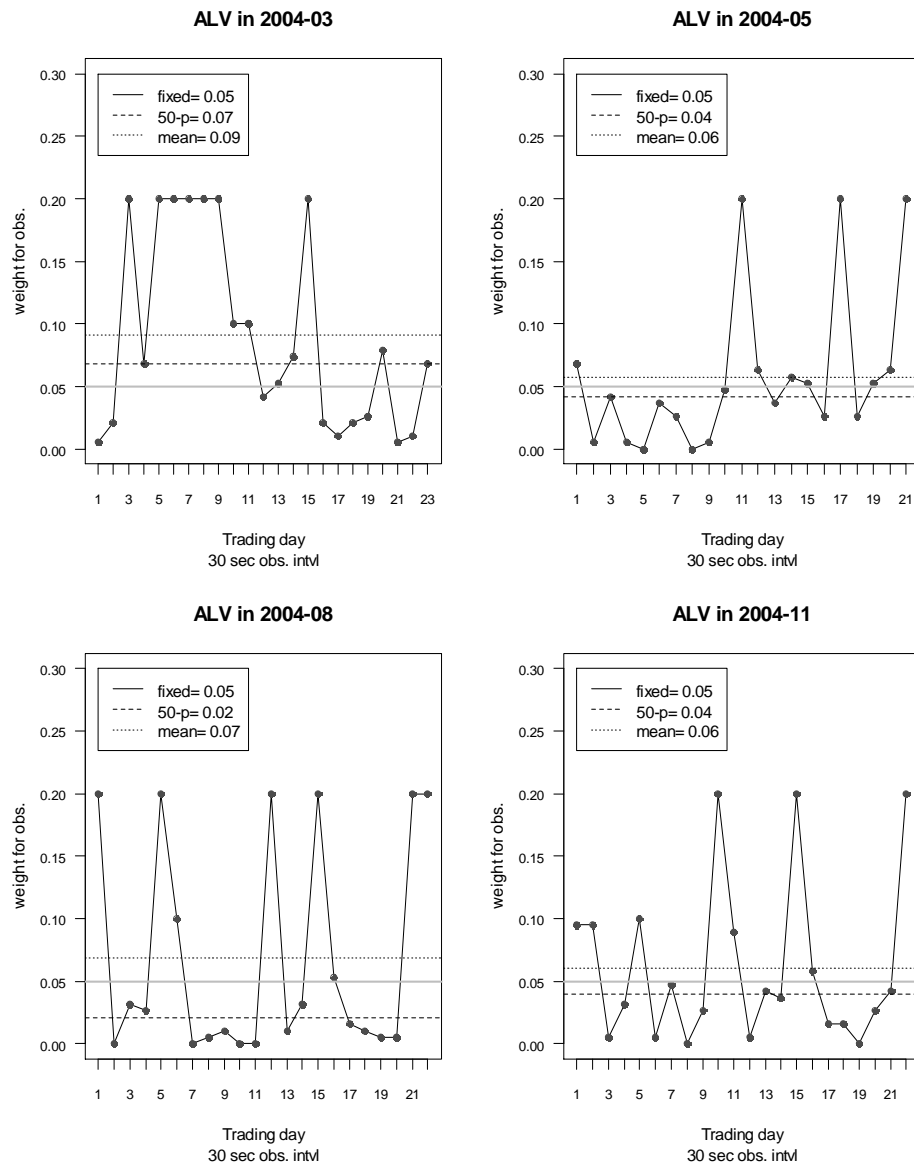
Figure 3.4: Allianz in March, May, August and November, 2004: weight for observation used by daily best logistic regression of price change indicator on predicted log-volume. Observations at 30 second interval.

3.1.4 Occurrence of price change

We now investigate the relation between predicted volume N_{t-1} and the binary variable representing price change occurrence on interval t, denoted by X_t. To illustrate this relationship we present here our analysis of the intraday bivariate time series of N_{t-1} and X_t for ALV in March, May, August and November of the year 2004. We computed the volumes and the returns at 30 second intervals, and we computed the predicted volumes by the exponential filter (using an α of 0.05).

For each of the trading days we investigated we fitted a logistic regression of X_t on N_{t-1} to the observations[10]. Figure 3.5 shows, for each trading day, the probability of price change (under the fitted model) as a function of N_{t-1} for ALV in each of the four months. We note the following features in this figure. Firstly, in each of the four months shown the probability of price change at interval t in each trading day is an increasing function of the predicted volume at interval t. Secondly, in each of the four months, the curves for daily probability of price change function lie close to each other, i.e. they are reasonably stable across the trading days. Finally, in each of the four months, the daily curves all have positive slope and they can be divided roughly into two groups: a group with steeper-slopes and the other with flatter slopes.

---

[10]Using the **R** function `glm`.
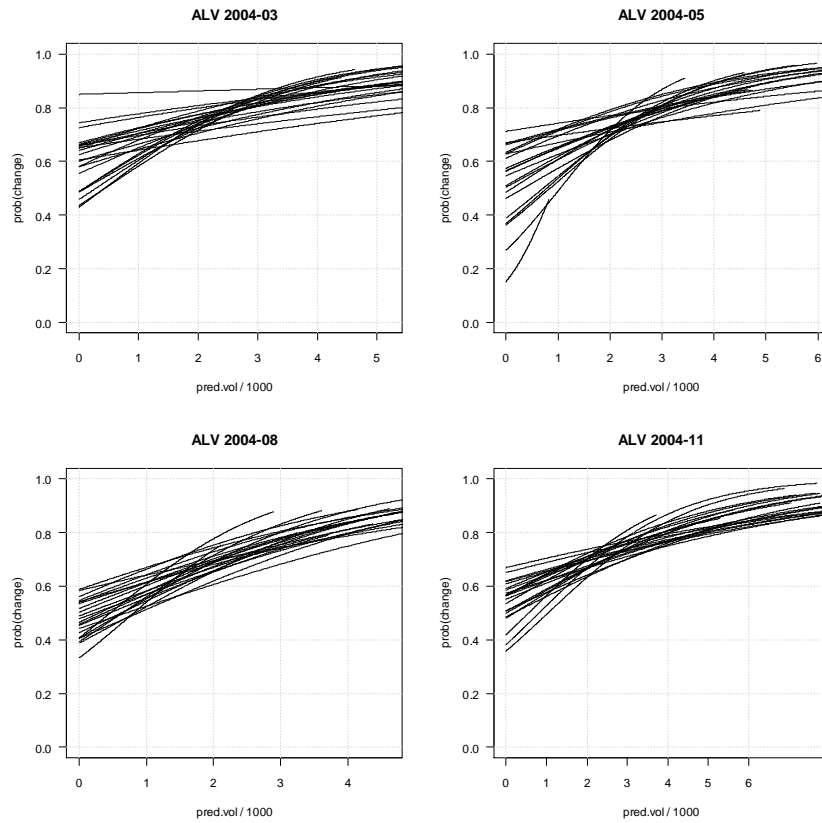
Figure 3.5: Allianz in March, May, August and November, 2004: daily probability of price change as a function of predicted volume. In computing the estimates we did not exclude the outliers.

3.1.5 Variance of conditional return given price change

We investigate here the relation between predicted volume N_{t-1} and squared conditional return given price change has occurred on interval t, denoted by M_t^2. Here, we used the squared return as an approximation of volatility. To illustrate this relationship we present here our analysis of the intraday bivariate time series of N_{t-1} and M_t^2 for ALV on each day of the first week of May in the year 2004.

The top row of Figure 3.6 displays a smooth of the scatter plot of $\{(N_{t-1}, M_t^2)\}$ for each trading day. The bottom row of Figure 3.6 displays the plot of density estimate of predicted volume for each trading day[11]. In this figure we note the following. Firstly, the curves vary considerably from one day to the next. This indicates that the relationship between return squared and volume is not stable. Secondly, on each day, the nonparametric regression line contains two relative extrema. This indicates that the volatility for interval t, conditional on the occurrence of price change, depends on N_{t-1}, and that the volatility alternates between an increasing and a decreasing function of N_{t-1}.

3.2 Forecasting the volatility of intraday returns

From the previous section, we saw in both the plot of probability of price change as a function of lagged volume, and the plot of squared return as a function of lagged volume that a mixture model could provide a better description of their

---

[11]We used the **R** function `density` (Gaussian kernel and bandwidth given by Silverman's "rule-of-thumb") to compute the density estimates. These plots indicate the range of values over which the predicted volumes are most concentrated.
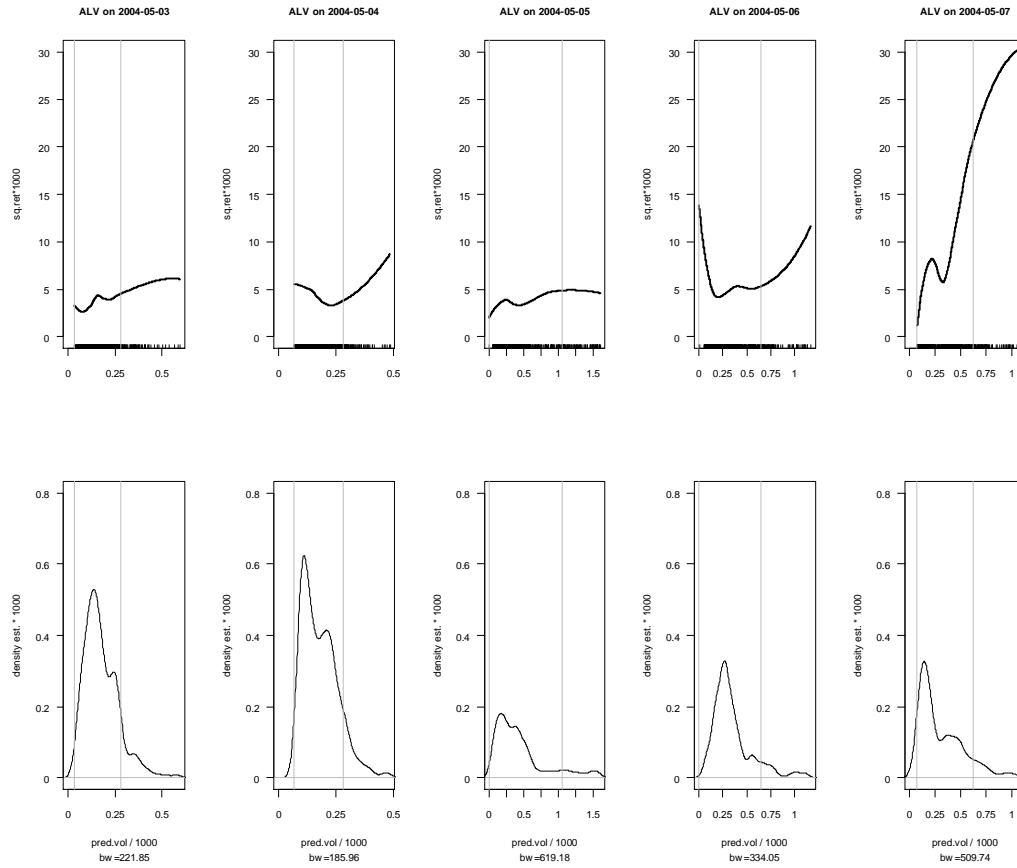
Figure 3.6: Allianz in each trading day of week 1 of May, 2004. Local polynomial regression, computed by the **R** function `loess`, of conditional squared return given price change on predicted volume (top row). Density estimate, computed by the **R** function `density`, of predicted volume (bottom row). Vertical lines indicate the smallest and the ninetieth-percentile of predicted volume respectively.

relationships than that provided by a generalised linear model with the logistic and the log-link function respectively. For this purpose, a particular type of dependent mixture model called the hidden Markov model (HMM) is used as a starting point. In particular, we use the class of HMMs incorporating a covariate to describe the relationships mentioned above. The main reference for HMM is Zucchini and MacDonald (2009), which we denote by ZM in what follows.

3.2.1   Hidden Markov models

An HMM $\{S_t : t = 1, 2, \ldots\}$ consists of two components. The first component is an unobserved random process, denoted by $C_1, C_2, \ldots$, which is assumed to be a (discrete-time) Markov chain. This means that the dependence structure of $\{C_t\}$ satisfies the Markov property (ZM, p. 16)

$$\Pr\left(C_t | C_{t-1}, \ldots, C_1\right) = \Pr\left(C_t | C_{t-1}\right). \tag{3.2}$$

The second component is an observed random process, denoted by $S_1, S_2, \ldots$, which is assumed to have the following dependence structure

$$\Pr\left(S_t | S_{t-1}, \ldots, S_1, C_t, \ldots, C_1\right) = \Pr\left(S_t | C_t\right), \tag{3.3}$$

i.e. the distribution of S_t depends only on the realisation of C_t. If C_t is assumed to take on the values $\{1, 2, \ldots, m\}$ then $\{S_t\}$ is called an *m−state HMM*. We assume that the conditional distribution $\Pr\left(S_t | C_t\right)$ comes from a family of parametric distributions. In the case that this distribution is the Poisson $\{S_t\}$ is termed a *Poisson-HMM*.

For the HMM investigated in this chapter we make three additional assumptions on the Markov chain $\{C_t\}$. Firstly, $\{C_t\}$ is a homogeneous Markov chain with (one-step) transition probability matrix (t.p.m.)

$$
\Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix},
$$

where the transition probabilities are defined by $\gamma_{ij} = \Pr\left(C_t = j | C_{t-1} = i\right)$, for $i, j = 1, \ldots, m$. Secondly, the Markov chain is stationary with stationary distribution $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_m)$, where $\delta_i = \Pr\left(C_t = i\right)$. Lastly, the Markov chain is irreducible[12].

We now introduce notation for discrete and continuous observations. In the case of discrete-valued observations we define $p_i\left(s_t\right) = \Pr\left(S_t = s_t | C_t = i\right)$, where $i = 1, \ldots, m$ are the states of the Markov chain. Thus $p_i\left(\cdot\right)$ is the probability *mass* function for S_t whose index is the realisation of the Markov chain at time t. In the case of continuous-valued observations $p_i\left(\cdot\right)$ is the probability *density* function for S_t whose index is given by the outcome of C_t.

An HMM generates an observation at time t as follows: the unobserved random process selects the parameter θ_i, and then the distribution with this parameter generates the observation. In this sense ZM (p. 30) termed $\{C_t\}$ as the "parameter process" and $\{S_t\}$ as the "state-dependent process." Following from

---

[12] A more detailed discussion of stationarity and irreducibility properties of the Markov chain is given in ZM (p.16-18).

these terminologies we refer to p_i as the *state-dependent distributions* (see ZM, p. 32).

Model specification. The particular class of basic HMM to be applied is specified firstly, by selecting the family of parametric distributions for the state-dependent distributions in the model and secondly, by selecting the number of states m. As an illustration a two-state Poisson-HMM is an HMM with t.p.m.

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix},$$

and the state-dependent distributions $p_i(s_t) = \frac{\lambda_i^{s_t} \exp(-\lambda_i s_t)}{s_t!}$, where $s_t = 0, 1, \ldots$.

The likelihood. Suppose an observations series of size T, denoted by $s_1, \ldots, s_T$, was generated by an m−state HMM with initial distribution $\boldsymbol{\delta}$ and t.p.m. $\boldsymbol{\Gamma}$ for its Markov chain, and p_i as their state-dependent probability distributions. Then the probability of the observed series is

$$\Pr(S_1 = s_1, \ldots, S_T = s_T; \boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ denotes the vector with parameters for the model as its elements. This probability is regarded as a function of $\boldsymbol{\theta}$ which is called the *likelihood* of the m−state HMM; we denote this by L_T. Define the diagonal matrix

$$\mathbf{P}(s_t) = \begin{pmatrix} p_1(s_t) & & 0 \\ & \ddots & \\ 0 & & p_m(s_t) \end{pmatrix}.$$

The formula for L_T (given in ZM, p. 37-38) is

$$L_T = \boldsymbol{\delta} \cdot \mathbf{P}\left(s_1\right) \cdot \boldsymbol{\Gamma}\mathbf{P}\left(s_2\right) \cdot \dots \cdot \boldsymbol{\Gamma}\mathbf{P}\left(s_T\right) \cdot \mathbf{1}', \tag{3.4}$$

where $\mathbf{1}'$ is a column vector of m 1's. In the case that the initial distribution of the Markov chain is assumed to be its stationary distribution, Equation (3.4) can be written as

$$L_T = \boldsymbol{\delta} \cdot \boldsymbol{\Gamma}\mathbf{P}\left(s_1\right) \cdot \boldsymbol{\Gamma}\mathbf{P}\left(s_2\right) \cdot \dots \cdot \boldsymbol{\Gamma}\mathbf{P}\left(s_T\right) \cdot \mathbf{1}'. \tag{3.5}$$

A recursive scheme to compute the likelihood is given in ZM (p. 38). Essential to their development is the vector of *forward probabilities*, which is defined, at each time point t, by

$$\begin{aligned}\boldsymbol{\alpha}_1 &= \boldsymbol{\delta} \cdot \mathbf{P}\left(s_1\right)\\ \boldsymbol{\alpha}_t &= \boldsymbol{\delta} \cdot \mathbf{P}\left(s_1\right) \Pi_{k=2}^{t}\boldsymbol{\Gamma}\mathbf{P}\left(s_k\right) \quad \text{(for } t = 2, \dots, T)\end{aligned} \tag{3.6}$$

From this definition, it follows that likelihood (3.4) can be computed in a recursive manner as

$$\begin{aligned}\boldsymbol{\alpha}_1 &= \boldsymbol{\delta} \cdot \mathbf{P}\left(s_1\right)\\ \boldsymbol{\alpha}_t &= \boldsymbol{\alpha}_{t-1} \cdot \boldsymbol{\Gamma}\mathbf{P}\left(s_t\right) \quad \text{(for } t = 2, \dots, T)\\ L_T &= \boldsymbol{\alpha}_T \cdot \mathbf{1}',\end{aligned} \tag{3.7}$$

where, in the case that s_t is missing, $\mathbf{P}\left(\cdot\right)$ is taken to be the identity matrix. In the case that the Markov chain is assumed to be stationary (that is, $\boldsymbol{\delta} = \boldsymbol{\delta}\boldsymbol{\Gamma}$), we can compute the likelihood (3.5) by recursion (3.7) in which the first equation is replaced with $\boldsymbol{\alpha}_0 = \boldsymbol{\delta}$ and the index t in the second equation runs from 1 to T. The computing time needed to evaluate L_T by DNM is proportional to Tm^2 (see

ZM, p. 38). This is linear in the length of the series, and so the computational effort remains modest even for "long" time series.

Parameter estimation. We estimate the parameters of an HMM by maximum likelihood. This is commonly performed by *direct numerical maximisation* (DNM) of the likelihood and/or by the *Expectation-Maximisation* (EM) algorithm[13].

In this chapter we chose DNM without using analytical derivatives to perform maximum likelihood estimation. This is motivated by our need to implement a variety of models for the exploratory modelling of returns. We exploit in particular two characteristics of DNM. Firstly, its implementation causes little difficulty: we require only the forward probabilities to compute the likelihood and we have available an optimisation routine (i.e. `nlm` in **R**), which does not require us to supply derivatives. Secondly, it allows a model to be modified with a minimum of programming effort. A modification requires changes mainly in the code that evaluates the likelihood.

The implementation of DNM of recursively calculated likelihood routinely encountered three main problems, namely numerical underflow, constraints on the parameters, and existence of multiple local maxima in the likelihood (ZM, p. 45). Each problem, together with a strategy to overcome it, is now described.

Parameter estimation: numerical underflow. For discrete observations

---

[13]The relative merits of the DNM and the EM algorithm are discussed in Bulla and Berzel (2008).

generated by an HMM, the values of forward probabilities in $\boldsymbol{\alpha}_t$ become so small with increasing t that they can be numerically rounded to zero. This problem is called *numerical underflow* of the likelihood (see ZM, p. 46).

We employ the strategy described in ZM (p. 46-47) to overcome numerical underflow. The key idea there is to first define a vector of "scaled" forward probabilities at each time point t as

$$\phi_t = \frac{\boldsymbol{\alpha}_t}{w_t},$$

where $w_t = \boldsymbol{\alpha}_t \mathbf{1}'$. Then the recursive likelihood computation can be re-written as

$$\phi_0 = \boldsymbol{\delta}$$

$$l_T = \log\left(L_T\right) = \sum_{t=1}^{T} \log\left[\phi_{t-1}\boldsymbol{\Gamma}\mathbf{P}\left(s_t\right)\mathbf{1}'\right],$$

where $\boldsymbol{\delta}$ is the initial distribution of the Markov chain. Log-likelihood computation using the above procedure "will avoid underflow in many cases" (ZM, p. 47).

Parameter estimation: parameter constraints. Some parameters of HMMs are bounded. For example, each transition probability must take on a value between zero and one. Thus, maximization of the log-likelihood is a constrained optimization problem.

There are at least two problems with the constrained optimisation formulation. Firstly, the constrained optimizer may be slow in locating the optimum (ZM, p. 47). Secondly, the coding required for the constraints leaves the programming of the likelihood unnecessarily complicated. In order to overcome these problems we adopt the strategy described in ZM (p. 48-49). The essential idea there is to

make an appropriate one-to-one transformation of the constrained parameter to an unconstrained parameter with respect to which the likelihood is maximised. The solution returned by the unconstrained maximiser is then transformed back to the corresponding constrained parameter values. The unconstrained parameters are referred to as the "working" parameters and the constrained parameters as the "natural" parameters.

Parameter estimation: existence of multiple local maxima in the likelihood. The likelihood function of an HMM often has several local maxima (ZM, Section 3.4.1). However, there exists no simple method to determine whether the optimum located by the numerical maximiser is indeed the global maximum.

Since the identified local maximum depends on the starting values used, we adopt the strategy of experimenting with a few plausible starting values, to see whether the same maximum is reached each time (ZM, Section 3.4.1). The selection of plausible starting values for HMM parameters is done according to ZM (Section 3.4.2). In the HMMs described later, however, we have to develop other ways to select starting values for the parameters of state-dependent distributions.

Model selection. From among the equally plausible models that have been fitted to the data we need to select objectively the "best" one. This is a model selection problem[14]. In this chapter we use two criteria for model selection. The

---

[14]See Zucchini (2000) for an introductory account.

first is the *Akaike information criterion* (AIC), which is defined by

$$AIC = -2\log\left(L_T\right) + 2p, \tag{3.8}$$

where L_T is the likelihood function of the HMM as given by Equation (3.4) and p is the number of parameters required to be estimated. The second is the *Bayesian information criterion* (BIC), which is defined by

$$BIC = -2\log\left(L_T\right) + p\log\left(T\right), \tag{3.9}$$

where T is the length of the time series. An example of model selection in stationary HMM is given in ZM (p. 90-92).

Model checking. After the best HMM has been selected, we need to decide whether this model is adequate for the out-of-sample period. This is a model checking problem.

We analyse the forecast pseudo-residuals of an HMM to assess the general fit of the selected model in the out-of-sample period[15]. In the case of continuous observations the two versions of forecast pseudo-residuals we analyse are: firstly, the uniform pseudo-residuals (UPRs)[16], defined as

$$u_t = \Pr\left(S_t \leq s_t \middle| \mathbf{S}^{(t-1)} = \mathbf{s}^{(t-1)}\right)$$

and secondly, the normal pseudo-residuals (NPRs), defined as

$$z_t = \Phi^{-1}\left(u_t\right),$$

---

[15] See ZM (Section 6.2) for a description of the use of pseudo-residuals in HMMs. .
[16] This basic concept dates back to at least Rosenblatt (1952).

where $\Phi^{-1}(u_t)$ denotes the distribution function of the standard normal distribution. Both versions are based on the conditional distribution given all preceding observations[17]. This is given by the ratio of the likelihood of the first t observations to that of the first $t-1$[18]. If the selected model is valid, then the UPRs are distributed $U(0,1)$ and the NPRs are distributed as standard normal. We can check the selected model by a visual inspection of the histogram and qq-plot of the pseudo-residuals.

3.2.2   A model to forecast the volatility of intraday returns

Let R_t denote the return on intraday time interval t and X_t the binary variable representing the occurrence of price change on interval t. In our model we define

$$R_t = \begin{cases} M_t & \text{with probability } \pi_t \\ 0 & \text{with probability } (1-\pi_t), \end{cases} \tag{3.10}$$

where π_t denotes the probability of price change occurring on interval t, i.e. $\Pr(X_t = 1)$ and M_t denotes the conditional return given price change occurred on interval t. Equation (3.10) defines a mixed discrete-continuous random variable, since the outcome of X_t is binary and M_t is continuous-valued.

Model specification. We specify our model for $\{R_t : t = 1, 2, \ldots\}$ by specifying a separate model for each component of $\{R_t\}$. In both the model for the binary series $\{X_t : t = 1, 2, \ldots\}$ and the model for the conditional returns

---

[17]This is termed the forecast distribution under the model. Hence, $\{u_t\}$ and $\{z_t\}$ are described as "forecast" pseudo-residuals.

[18]For discrete observations the formula is given in ZM (p. 97, Section 6.2.3). In the continuous case the formula is the same but with probability function replaced by the density function.

$\{M_t : t = 1, 2, \ldots\}$ we use the exponentially smoothed volumes, lagged by 30 seconds, N_{t-1} as a covariate[19]. We refer to N_{t-1} as the predicted volume, because the lagged smoothed volume is the one-step-ahead forecast of V_t. Values of smoothed volume were available for the same period as the observations of both the binary variable and the conditional returns. The use of lagged smoothed volume, as opposed to simultaneous volume, as a covariate is motivated by our need for the covariate to be available at the time of forecast. In this chapter a 30-second-ahead forecast is considered.

For the binary series $\{X_t\}$ the Bernoulli-HMM we consider is as follows. In Section 3.1.4 we saw (in Figure 3.5) that there is a positive dependence of the price change probability on the predicted volume. This motivated us to introduce N_{t-1} into the model via the parameter π_i of the state-dependent Bernoulli distributions. One way of doing so is to assume that, given N_{t-1}, the logit of price change probability in state i is a linear function of N_{t-1} (for example, $logit\,(\pi_{t,i}) = \alpha_{0,i} + \alpha_{1,i}N_{t-1}$). Here we did not make the restriction that a single linear function should apply to all m states. In Section 3.1.4 we also saw that the curves of daily price change probability as a function of N_{t-1}, in each month investigated, can be divided into two groups, where the curves in the one group have slopes that are steeper than those in the other. This observation motivates us to consider a Markov chain with two states in our model for $\{X_t\}$. In what follows we refer to our model for $\{X_t\}$ as the *two-state Bernoulli-HMM incorporating volume.*

---

[19]The exponentially smoothed volumes are computed by Equation (3.1).

For the conditional returns $\{M_t\}$ the normal-HMM we consider is as follows. The state-dependent normal distributions here have mean zero and variance σ_i^2. This variance represents the return volatility in state i. In Section 3.1.5 we saw (in Figure 3.6) that there is a dependence of the squared return on the predicted volume. This motivated us to introduce N_{t-1} into the model via the parameter σ_i^2 of the state-dependent distributions, since we consider squared return as an approximation of return volatility. We did so by assuming that, given N_{t-1}, the logarithm of variance in state i is a linear function of N_{t-1} (for example, $\log\left(\sigma_{t,i}^2\right) = \alpha_{0,i} + \alpha_{1,i} N_{t-1})$ [20]. In Section 3.1.4 we also saw that in general two relative extrema occur in the curves for squared return as a function of N_{t-1}. This observation motivates us to consider a Markov chain with two states in our model for $\{X_t\}$. In what follows we refer to our model for $\{M_t\}$ as the *two-state normal-HMM incorporating volume*.

From Equation (3.10) we see that our model for $\{R_t\}$ consists of two components, where a separate HMM is assumed for each component. We also allow N_{t-1} to influence the parameter of the state-dependent distribution in both classes of HMMs assumed. Thus, we refer to our model for the returns as a *two-component HMM incorporating volume*.

Model fitting. The model we considered for $\{X_t\}$ and for $\{M_t\}$ are both special cases of HMM incorporating a covariate into the state-dependent proba-

---

[20] Again we did not here make the restriction that a single linear function should apply to all m states.

bilities. Thus, the expression for the likelihood of a sequence of T observations

$s_1, \ldots, s_T$ is (see ZM, p. 126)

$$L_T = \boldsymbol{\delta} \cdot \mathbf{P}_1\left(s_1, N_0\right) \cdot \boldsymbol{\Gamma}\mathbf{P}_2\left(s_2, N_1\right) \cdot \cdots \cdot \boldsymbol{\Gamma}\mathbf{P}_T\left(s_T, N_{T-1}\right) \cdot \mathbf{1}', \tag{3.11}$$

where, for convenience, we chose N_0 to be zero[21] and $\mathbf{P}_t\left(s_t, N_{t-1}\right)$ denotes the

diagonal matrix with $i-$th diagonal element $p_{t,i}\left(s_t, N_{t-1}\right)$. The definition of the

various state-dependent distributions are presented in Appendix C., Table 4.3.

Numerical underflow of the likelihood can be avoided by applying the algo-

rithm in ZM (p. 47) with state-dependent probability distributions now depending

on N_{t-1}. In order to avoid constraints on the model parameters we reparameterise

the model. The transformations of the parameters of state-dependent distribu-

tions are given in Appendix C., Table 4.4. We reparameterise the t.p.m. $\boldsymbol{\Gamma}$ as

follows. We first transform the starting values for parameters of the t.p.m. to the

"working" parameters

$$\tau_{ij} = \log\left(\tfrac{\gamma_{ij}}{\gamma_{ii}}\right) \quad \text{(for } i \neq j \text{ and } i, j = 1, 2\text{).}$$

We then transform the solutions from unconstrained optimizer back to the "nat-

ural" parameters

$$\rho_{ij} = \begin{cases} \exp\left(\tau_{ij}\right) & \text{(for } i \neq j\text{)} \\[2mm] 1 & \text{(for } i = j\text{)} \end{cases}$$

$$\gamma_{ij} = \frac{\rho_{ij}}{\rho_{i1} + \rho_{i2}}.$$

---

[21]An alternative to this is to use the average of the series of N_t that covers the preceding trading day.

Model selection and checking. We perform model selection in models for both the binary series and the conditional returns by means of AIC and BIC (see Section 3.2.1 for an account).

We then examine the selected model to assess its suitability for the observations. We describe three techniques for the binary series and one for the conditional returns. In all the techniques we need to find the forecast distributions under the selected model, i.e. the conditional distribution of each observation given all preceding observations. In the case of the binary series the conditional probability $\Pr\left(X_t = x | \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}\right)$ is given by the ratio of the likelihood of the first t observations to that of the first $t-1$ (cf. ZM, p. 97):

$$\Pr\left(X_t = x | \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}\right) = \frac{\boldsymbol{\alpha}_{t-1}\boldsymbol{\Gamma}\mathbf{P}_t\left(x_t, N_{t-1}\right)\mathbf{1}'}{\boldsymbol{\alpha}_{t-1}\mathbf{1}'}. \qquad (3.12)$$

In the case of the conditional returns the conditional density $f\left(m_t | \mathbf{M}^{(t-1)} = \mathbf{m}^{(t-1)}\right)$ is given by Equation 3.12, with the probability function replaced by the density function and the diagonal matrix replaced by $\mathbf{P}_t\left(m_t, N_{t-1}\right)$.

For the binary series we describe three model-checking techniques to assess the suitability of the selected model. The first technique formulates the model checking problem as a test of the null hypothesis

$$H_0: \quad logit\left[E\left(X_t\right)\right] = logit\left(\hat{\pi}_t\right),$$

where $\hat{\pi}_t$ denotes the conditional probability of price change $\{X_t = 1\}$ given the history $\mathbf{X}^{(t-1)}$ computed under the selected model, against an alternative hypoth-

esis

$$H_A: \quad logit\left[E\left(X_t\right)\right] = S\left[logit\left(\hat{\pi}_t\right)\right] \ ,$$

where S is a smoothing spline function[22]. We visually inspect the plot of S against the logit of $\hat{\pi}_t$; a deviation of S from the identity function constitutes evidence against the null hypothesis. This plot is also used to reveal the nature of the departure.

The second model-checking technique assesses the correspondence between the average of forecast probabilities of price change under the selected model and the relative frequency of occurrences of price change. For the validation sample, we group the forecast probabilities of price change (i.e. the conditional probability of price change $\{X_t = 1\}$, given the history $\mathbf{X}^{(t-1)}$) into one of the twenty intervals of the form $[(i-1)\,w, iw)$ where $i = 1, 2, \ldots, 20$. The fixed interval length w used is 0.05. The mid-point of the $i-$th interval is used as the average of the forecast probabilities for this interval. We then plot (for $i = 1, 2, \ldots, 20$) the relative frequency of the occurrences of price change in the $i-$th interval against the average of the forecast probabilities for this interval. The deviation of the plotted pair from the identity function is evidence of inadequacy of the selected model.

The third model-checking technique uses the logistic regression to assess the correspondence between the occurrences of price change and the forecast proba-

---

[22]This technique is developed in Zucchini et al. (2008, p. 810, Section 5).

bilities of price change. We smooth the standardised deviance residuals[23] under the fitted logistic regression by using the method of locally-weighted polynomial regression. Departure of the smoothed[24] residuals from 0 constitutes evidence for the inadequacy of selected model.

For the conditional return given price change occurred we check the selected model by a visual inspection of the plots of the NPR. This is described in Section 3.2.1.

Improvement in forecast accuracy achieved. In this chapter we forecast the volatility of the intraday return 30 seconds ahead.

Volatility itself is not directly observable, thus we use the squared return as its approximation. We consider two models for forecasting: firstly, the exponential smoother as our baseline model and secondly, our two-component HMM incorporating volume. For each of these models we use the mean-squared forecast error (MSFE) as a measure of its forecast accuracy, where the forecast error is given by the difference between an observation and its forecast. In the out-of-sample period we compare the MSFE for the baseline model with that for our model. We define the forecast quality as the improvement in forecast accuracy achieved by our model on the exponential smoother. We then use the ratio of the MSFE of the two-component model to that of the baseline model as a measure of this improvement.

---

[23]It is computed under the fitted model by Equation 5.3 in Collett (1991, p. 123).
[24]The smoothed values are computed by the **R** function `lowess.`

For each of the two models we consider we now give an account of the method used for volatility forecasting. We applied both methods to all 32 samples we investigated in this chapter. For each sample of length T we estimate the model by using the calibration sample of length t_{in} (i.e. observations at interval $t = 1, 2, \ldots, t_{in}$) and we assess the forecast quality of the model fitted by using the validation sample (i.e. observations at interval $t = t_{in} + 1, t_{in} + 2, \ldots, T$).

Let r_t^2 denote the squared return at interval t. In the first method we compute the (one-step-ahead) forecast at interval t, denoted by $\hat{r}_t^2(1)$, by using the exponential smoother

$$\hat{r}_t^2(1) = \alpha \cdot r_t^2 + (1 - \alpha) \cdot \hat{r}_{t-1}^2(1), \tag{3.13}$$

where α is the smoothing parameter. Define

$$\epsilon_{t-1}(1) = r_t^2 - \hat{r}_{t-1}^2(1),$$

where $\epsilon_{t-1}(1)$ denotes the (one-step-ahead) forecast error made on interval $t-1$ by $\hat{r}_{t-1}^2(1)$. The average forecast error made by $\hat{r}_t^2(1)$ for a sequence of n observations, which we term the mean squared forecast error for exponential filter ($MSFE_{\exp}$), is then defined as

$$MSFE_{\exp}(\alpha) = \frac{1}{n-1} \sum_{t=2}^{n} \epsilon_{t-1}^2(1). \tag{3.14}$$

For the validation sample we use the α estimated by using the calibration sample[25] for the exponential smoother. We re-initialise the forecasting procedure daily[26],

---

[25] This was done by minimising the $MSFE_{\exp}$ for the in-sample period. We used the **R** function `optimize` for the minimisation. The smoothing procedure was re-initialised daily by setting the first observation for each day as the initial forecast.

[26] Again by setting the first observation for each day as the initial forecast.

thus the variation in the squared returns across different days is unaccounted for in the forecasts. Thus the performance of the exponential smoother reported here is better than it would have been had this daily re-initialisation not been done.

In the second method we define the volatility forecast as the variance of the (one-step-ahead) forecast distribution for the returns, i.e. the distribution of R_t conditional on all preceding returns $\mathbf{R}^{(t-1)} = \mathbf{r}^{(t-1)}$. In order to derive a formula for this variance we need results for the forecast distribution for both components of the returns. Firstly, the forecast distributions for the binary series are $\Pr\left(X_t = x_t \,\middle|\, \mathbf{X}^{(t-1)}\right)$, which can be computed as a ratio of two likelihoods of a Bernoulli-HMM incorporating volume. Let $\pi_{t-1}(1)$ denote the forecast probability of price change, i.e.

$$\pi_{t-1}(1) = \Pr\left(X_t = 1 \,\middle|\, \mathbf{X}^{(t-1)}\right).$$

Secondly, the forecast distributions for the conditional returns are $f\left(m_t \,\middle|\, \mathbf{M}^{(t-1)}\right)$. We express each of these as a mixture of two state-dependent normal distributions[27], where the weight $w_i^{(t)}(1)$ is the $i-$th element of the vector $\boldsymbol{\phi}_t \boldsymbol{\Gamma}$ under the two-state normal-HMM incorporating volume. In this mixture the $i-$th component distribution is normal with mean 0 and variance $\sigma_{t,i}^2$. Using the result for the $k-$th uncentered moment of a mixture given in ZM (p. 8), we thus conclude that $f\left(m_t \,\middle|\, \mathbf{M}^{(t-1)} = \mathbf{m}^{(t-1)}\right)$ has a mean of 0 and a variance given by

$$\sum_{i=1}^{2} w_i^{(t)}(1) \cdot \sigma_{t,i}^2. \tag{3.15}$$

---

[27]By using Equation (5.5) on ZM, p. 79.

We now give a derivation for our conclusions. In accordance with Equation (3.10) the forecast distribution for $\{R_t\}$ has mean

$$0 \cdot \left(1 - \pi_{t-1}\left(1\right)\right) + E\left(M_t | \mathbf{M}^{(t-1)}\right) \cdot \pi_{t-1}\left(1\right) = 0, \tag{3.16}$$

where $E\left(M_t | \mathbf{M}^{(t-1)}\right)$ denotes the mean of the forecast distribution for the conditional return, and second moment given by

$$0^2 \cdot \left(1 - \pi_{t-1}\left(1\right)\right) + \left[Var\left(M_t | \mathbf{M}^{(t-1)}\right) + E\left(M_t | \mathbf{M}^{(t-1)}\right)^2\right] \cdot \pi_{t-1}\left(1\right)$$

$$= Var\left(M_t | \mathbf{M}^{(t-1)}\right) \cdot \pi_{t-1}\left(1\right), \tag{3.17}$$

where $Var\left(M_t | \mathbf{M}^{(t-1)}\right)$ denotes the variance of the forecast distribution for the conditional returns. Using results (3.16), (3.17) and (3.15) we conclude that the formula for the variance of the forecast distribution for $\{R_t\}$ is

$$\pi_{t-1}\left(1\right) \cdot \left[w_1^{(t)}\left(1\right) \cdot \sigma_{t,1}^2 + w_2^{(t)}\left(1\right) \cdot \sigma_{t,2}^2\right]. \tag{3.18}$$

Let $Var\left(R_t | \mathbf{R}^{(t-1)}\right)$ denote this variance. Define

$$\epsilon_{t-1}\left(1\right) = r_t^2 - Var\left(R_t | \mathbf{R}^{(t-1)}\right),$$

where $\epsilon_{t-1}\left(1\right)$ denotes the (one-step-ahead) forecast error made on interval $t-1$ by the two-component HMM incorporating volume. The average forecast error made by $Var\left(R_t | \mathbf{R}^{(t-1)}\right)$ in a sequence of n observations, termed the mean squared forecast error for HMM ($MSFE_{HMM}$), is then defined as

$$MSFE_{HMM} = \frac{1}{n-1}\sum_{t=2}^{n}\epsilon_{t-1}^2\left(1\right). \tag{3.19}$$

In order to measure the forecast performance for our two-component model relative to that of the exponential smoother, we define

$$RMSFE = \frac{MSFE_{\exp}}{MSFE_{HMM}}, \tag{3.20}$$

where $RMSFE$ is the acronym for ratio of MSFEs. An $RMSFE$ that exceeds 1 indicates that in the out-of-sample period the forecasts derived from our two-component HMM incorporating volume are, on average, more accurate than those computed by the exponential smoother.

Software implementation. In order to develop our software to implement HMMs that incorporate volume into the state-dependent probabilities, several problems need to be addressed. The main problems are the selection of a suitable model for the volumes series, the assessment of the bias of parameter estimators, the likelihood of a series of consecutive intra-daily bivariate series and the likelihood when data are missing. In Appendix C. (on pages 238-240) we describe how to overcome each of these problems.

3.3   Modelling the occurrence of price change

In this section we model the price change occurrence in the transaction data for major components of DAX. To represent the presence or absence of intraday price change we use Bernoulli-HMMs of the kind that allows for predicted volume in the state-dependent probability of price change.

In order to assess whether such models can be used successfully to forecast the

price change patterns, we use the transaction data for the eight shares we chose:
ALV, EOA, BAS, LHA, TUI, DBK, DCX and SCH. For each of the eight shares
we use the data that cover four separate periods, each of length two months in
the year 2004: March and April (period 1), May and June (period 2), August and
September (period 3), and November and December (period 4).

For each of the 32 datasets used we compute (v_t, r_t) at $30-$ second intervals
in each of the trading days, where in the $t-$th intraday interval v_t is the (log-)
volume traded and r_t is the return. This results in 1020 pairs of observations for
each day. For each of the 32 bivariate series we obtain we use observations that
cover the first month to fit our models (i.e. the calibration sample), and those
that cover the subsequent month to check for deficiencies in the fitted model (i.e.
the validation sample).

3.3.1   Models fitted

We consider various models for the bivariate time series (n_{t-1}, x_t), with n_{t-1}
being the prediction[28] of v_t, and x_t being the binary variable that represents the
presence (i.e. r_t is nonzero, which is denoted by 1) or the absence (i.e. r_t is zero,
which is denoted by 0) of price change. We use the logistic regression model,
which is fitted by the **R** function `glm`, as the baseline model.

We then consider three types of two-state Bernoulli-HMMs based on a station-

---

[28]The exponential smoother (as given in Section 3.1.3) with $\alpha = 0.5$ was used to compute the
series of predicted volumes $\{n_t\}$ on a daily basis. As we have defined in Equation (3.11), we set
$N_0 = 0$.

model	$logit\,(\pi_{t,1})$	$logit\,(\pi_{t,2})$
2.0	$\alpha_{0,1} + \alpha_{1,1}N_{t-1}$	$\alpha_{0,2} + \alpha_{1,2}N_{t-1}$
2.1	$\alpha_{0,1} + \alpha_{1,1}N_{t-1}$	$\alpha_{0,2}$
2.3	$\alpha_{0,1}$	$\alpha_{0,2}$

Table 3.1: Two-state Bernoulli-HMMs: models for logit of state-dependent probabilities of price change.

ary Markov chain[29]. In these models $\pi_{t,i}\,(n_{t-1})$ is the probability of price change occurring in intraday period t, which is a function of the predicted volume, if the Markov chain is in state i. Table 3.1 gives the expressions for $logit\,\pi_{t,i}\,(n_{t-1})$ in each type of model.

The likelihood of a two-state Bernoulli-HMM incorporating volume is given by Equation (3.11) in Section 3.2.2, where the state-dependent distributions are Bernoulli and the covariate is N_{t-1}. We fit the two-state Bernoulli-HMMs via the **R** function `nlm` by direct numerical maximisation of the log-likelihood of (n_{t-1}, x_t). The likelihood is maximised with respect to the unconstrained parameters, where the reparameterisation of the t.p.m. and the state-dependent distributions are described in Section 3.2.2.

In order to fit our models to the calibration samples several problems need to be addressed. The main ones are the determination of starting values for

---

[29] A one-state Bernoulli-HMM is just the logistic regression model, which assumes that the x_t's are independent Bernoulli random variables with the logit of mean in intraday interval t equals to $a + bn_{t-1}$. The constant term and the coefficient are the parameters.

the optimisation procedure and choosing the frequency of model calibration. In this subsection we first describe the method we used to determine the starting values, under the heading "Starting values for the iterations". We then discuss models calibrated on a daily basis, under the heading "Models fitted daily". On computational grounds it is desirable to avoid having to recalibrate the model on a daily basis. We therefore consider the alternative where the model is recalibrated less frequently, e.g. four times per month. We end this subsection by discussing "Models calibrated less frequently in a month".

Starting values for the iterations. We first fit the logistic regression model then we fit the two-state models in the order: Model 2.0, 2.1 then 2.3.

In Model 2.0 we use the parameter estimates of the logistic regression model as the starting values for the parameters of the model for *logit* $\pi_{t,1}$. In the model for *logit* $\pi_{t,2}$ we set $\alpha_{0,2} = 0$ and we take the exponentially smoothed volumes $\{n_t\}$ as given. $\pi_{t,2}$ is thus a function of $\alpha_{1,2}$ only. Using a range of values for $\alpha_{1,2}$, we generate various series of $\pi_{t,2}$. The starting value for $\alpha_{1,2}$ corresponds to a generated $\{\pi_{t,2}\}$ in which the proportion of values between 0.5 and 0.6 in $\{\pi_{t,2}\}$ exceeds 70%. In Model 2.1 we use $(\hat{\alpha}_{0,1}, \hat{\alpha}_{1,1})$ and $\hat{\alpha}_{0,2}$ in Model 2.0 as the starting values for $(\alpha_{0,1}, \alpha_{1,1})$ and $\alpha_{0,2}$. In Model 2.3 we use $\hat{\alpha}_{0,1}$ and $\hat{\alpha}_{0,2}$ in Model 2.1 as the starting value for $\alpha_{0,1}$ and $\alpha_{0,2}$.

Our choice of starting values for the parameters of the Markov chain follows the strategy of ZM (p. 50, Section 3.4.2). In Model 2.0 we assign 0.01 as the

starting value for both γ_{12} and γ_{21}. In Model 2.1 we use $\hat{\gamma}_{12}$ and $\hat{\gamma}_{21}$ in Model 2.0 as the starting values for γ_{12} and γ_{21}. In Model 2.3 we use $\hat{\gamma}_{12}$ and $\hat{\gamma}_{21}$ in Model 2.1 as the starting values for γ_{12} and γ_{21}.

Models fitted daily. We begin our investigation of the models considered by applying these to (n_{t-1}, x_t) for Allianz in Period 2 of the year 2004. The models are fitted on a daily basis in the month of May. For each of the trading days the two-state models are preferred by AIC. But no single type of two-state model are preferred for all days.

In order to assess the behaviour of the daily models fitted we thus inspect the most general type of two-state model we consider, i.e. models with N_{t-1} incorporated into the probability of price change in both states of the Markov chain. In these models $\pi_{t,i}(n_{t-1})$ is the probability of price change occurring in intraday interval t, as a function of the predicted volume, if the state is i. Figure 3.7 displays $\pi_{t,i}(n_{t-1})$ in state $i = 1$ and 2, plotted on the same set of axes, in the models fitted daily. In both states of the Markov chain the majority of the estimated functions are positively sloped. The curves in both states are roughly stable, although the curves in state 1 lie closer to each other than those in state 2.

In the t.p.m. of the underlying Markov chain, γ_{ij} denotes $\Pr(C_t = j | C_{t-1} = i)$, where γ_{12} and γ_{21} are the free parameters. Figure 3.8 depicts $\hat{\gamma}_{ij}$ for $i, j = 1, 2$ of the model fitted daily. Both free parameters are roughly stable; however, there is
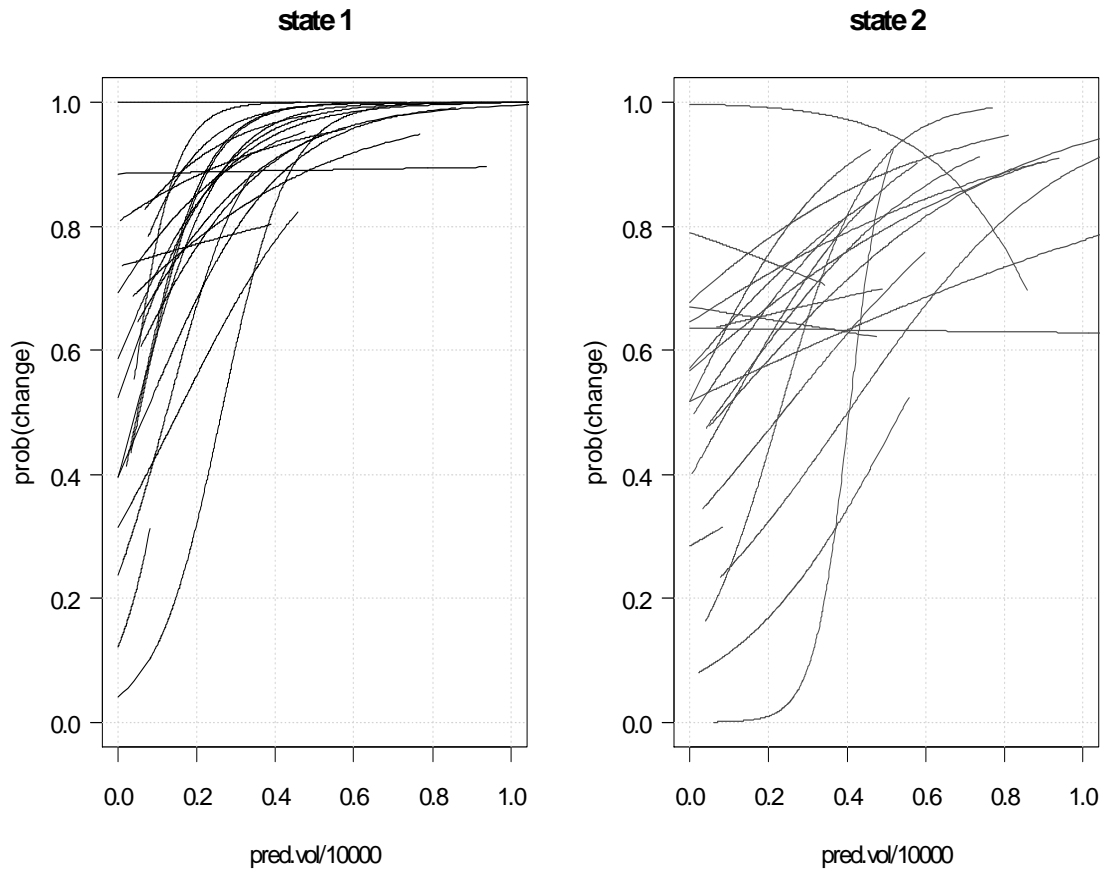
Figure 3.7: Two-state Bernoulli-HMMs: state-dependent probabilities of price change, as a function of N_{t-1}, in daily models for Allianz in May, 2004.
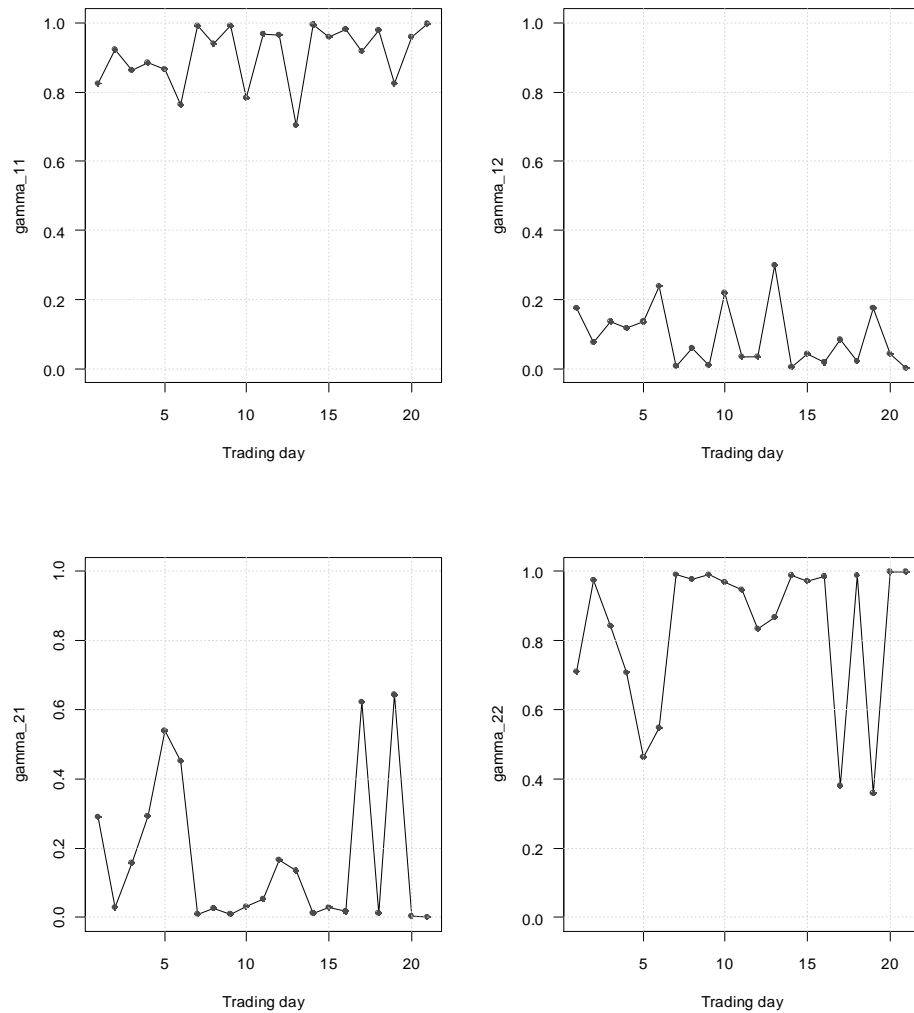
Figure 3.8: Two-state Bernoulli-HMMs: transition probabilities of Markov chain in daily models for Allianz in May, 2004.

considerable fluctuation in $\hat{\gamma}_{21}$.

Models fitted less frequently in a month. We use the two-state Bernoulli-HMM with $logit \; \pi_{t,i}(N_{t-1}) = \alpha_{0,i} + \alpha_{1,i} N_{t-1}$ in state $i = 1$ and 2. We apply this model to the calibration sample in Period 2 of the year 2004. We now fit the model less frequently in the month than daily as follows. We first divide the month into sub-periods then we calibrate the model for each of the sub-periods. The three cases that we investigate are the model recalibrated four times and twice in the month, and the model fitted monthly. In each of these cases we fit the model twice: firstly, by assuming in each sub-period that the bivariate observations in successive days are independent; and secondly, by assuming in each sub-period some serial dependence across successive days within each component of the bivariate observations (we achieve this by concatenating intraday volumes and returns in adjacent days by a series of missing observations, coded as `NA` in **R**, of length 10^{30}. We then apply the model to this consolidated bivariate series.) In the first two cases we sum the values of AIC for the sub-periods. In the case of models fitted daily the values of AIC for each of the days are summed. Table 3.2 compares the four cases, assuming serial dependence within the sub-periods in each case, on the basis of their total AIC. It indicates that the total AIC selects the case of fitting the model daily.

We consider again the application of the three types of two-state Bernoulli-

---

[30]Our choice of ten missing observations between adjacent days is arbitrary, but it does make the adjacent series close to independent. We use this choice to concatenate adjacent series from hereafter.

Frequency	Total AIC
daily	**23455**.**1**
four times	23502.8
twice	23542.4
monthly	23666.4

Table 3.2: Two-state Bernoulli-HMMs: comparison of models for Allianz recalibrated less frequently in May, 2004, by total AIC. The consolidated series is used in each sub-period.

HMM, as presented in Table 3.1, to the calibration sample for Allianz in Period 2 of the year 2004. On computational grounds we recalibrate all of these models monthly. Within each component series we postulate a weak serial dependence across the days by using a series of missing observations of length 10[31] to concatenate adjacent intraday series. The monthly model is an approximation to each daily model, since the values of parameters for the monthly model apply to all the daily models. For each type of model we summed the value of AIC for all the daily models. We compare the three types of two-state Bernoulli-HMM, recalibrated monthly, on the basis of their AIC[32]. We found that Model 2.1 is preferred by AIC. Parameter estimates in this model are presented in row 3 of Table 3.3. Figure 3.9 plots $\pi_{t,i}$ as a function of N_{t-1} in each of the models for Period 2.

In the same way (i.e. monthly recalibration with serial dependence across days

---

[31] As before, this choice arbitrary.

[32] In the bracket we give the value of total AIC for the corresponding daily models.
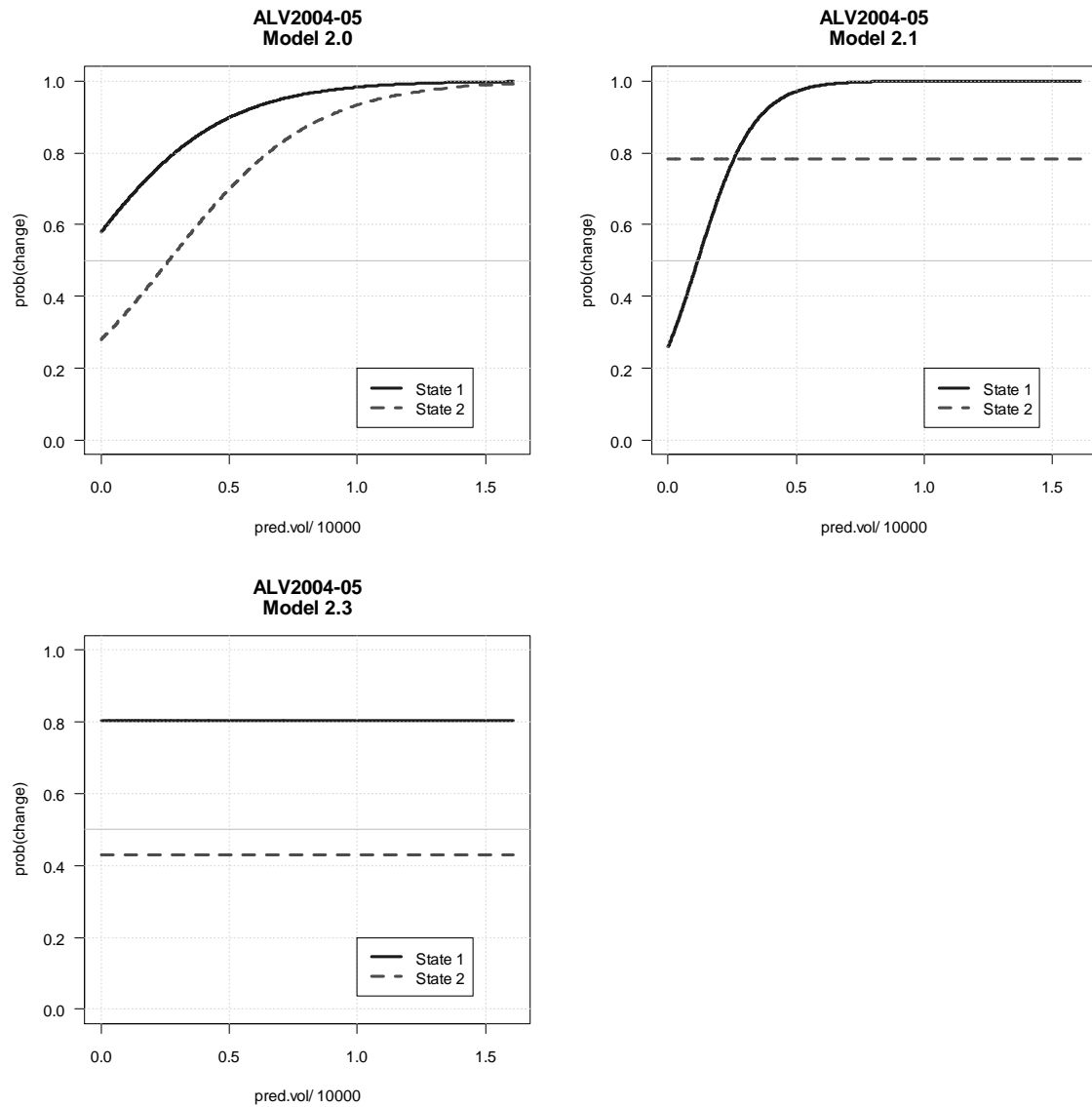
Figure 3.9: Monthly models for Allianz in May, 2004: Model 2.1 was selected.

ALV	$\hat{\alpha}_{0,1}$	$\hat{\alpha}_{0,2}$	$\hat{\alpha}_{1,1}$	$\hat{\alpha}_{1,2}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{22}$	$\hat{\delta}_1$	$\hat{\delta}_2$	#(obs.)
03.2004	0.746	0.0075	3.136	1.890	0.991	0.980	0.690	0.310	23483
05.2004	-1.050	1.284	9.079	—	0.972	0.968	0.533	0.467	21441
08.2004	0.548	-0.356	2.686	3.313	0.995	0.993	0.583	0.417	22462
11.2004	0.423	-0.275	2.852	1.792	0.990	0.971	0.744	0.256	22462

Table 3.3: Estimates of parameters for the selected two-state Bernoulli-HMM for Allianz in March, May, August and November, 2004.

in both component series) we apply the two-state Bernoulli-HMMs considered to the calibration sample for Allianz in Periods 1, 3 and 4 of the year 2004. For the calibration sample in all three periods the AIC preferred Model 2.0. Details of the selected model for each period are presented in rows 2, 4 and 5 of Table 3.3. Figures 3.10, 3.11 and 3.12 plot $\pi_{t,i}$, as a function of N_{t-1}, in the models for Periods 1, 3 and 4 respectively. In all these periods Model 2.0 was selected. Two notable features of these models are as follows. Firstly, the state-dependent probabilities $\pi_{t,i}$ are a monotonically increasing function of N_{t-1}. Secondly, the functions $\pi_{t,i}$ do not cross: the probability of price change in one of the states is higher than that in the other state for all N_{t-1}. The estimated Markov chain underlying the selected models for all four periods are strongly persistent. Their unconditional probabilities are given in Table 3.3, under $\hat{\delta}_i$ which denotes $\Pr(C_t = i)$. Note in Table 3.3 that the estimates of $\alpha_{0,1}$, $\alpha_{0,2}$, $\alpha_{1,1}$ and $\alpha_{1,2}$ can look quite different from period to period but still lead to similar probabilities (cf. Figures 3.9 to
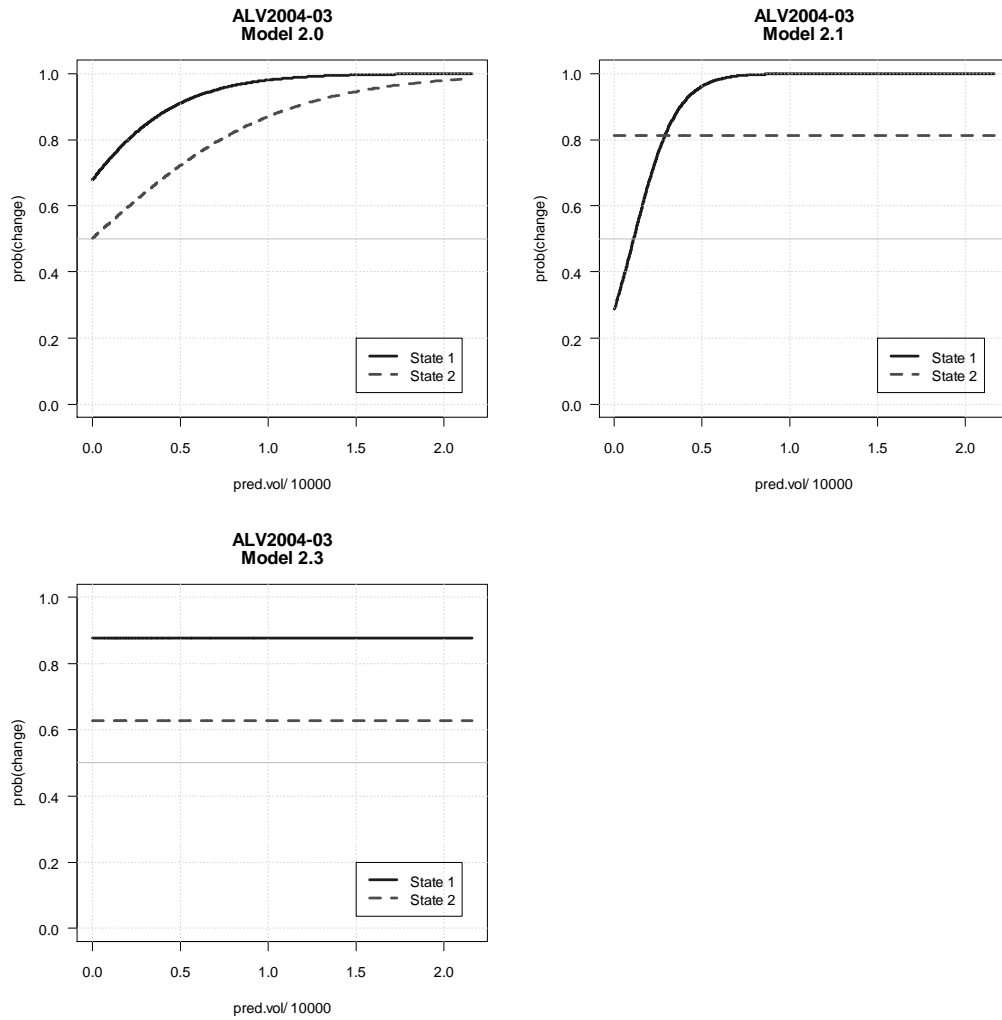
Figure 3.10: Two-state Bernoulli-HMMs: state-dependent probabilities of price change, as a function of N_{t-1}, in monthly models for Allianz in March, 2004. Model 2.0 was selected. Solid horizontal line indicates probability of 0.5.
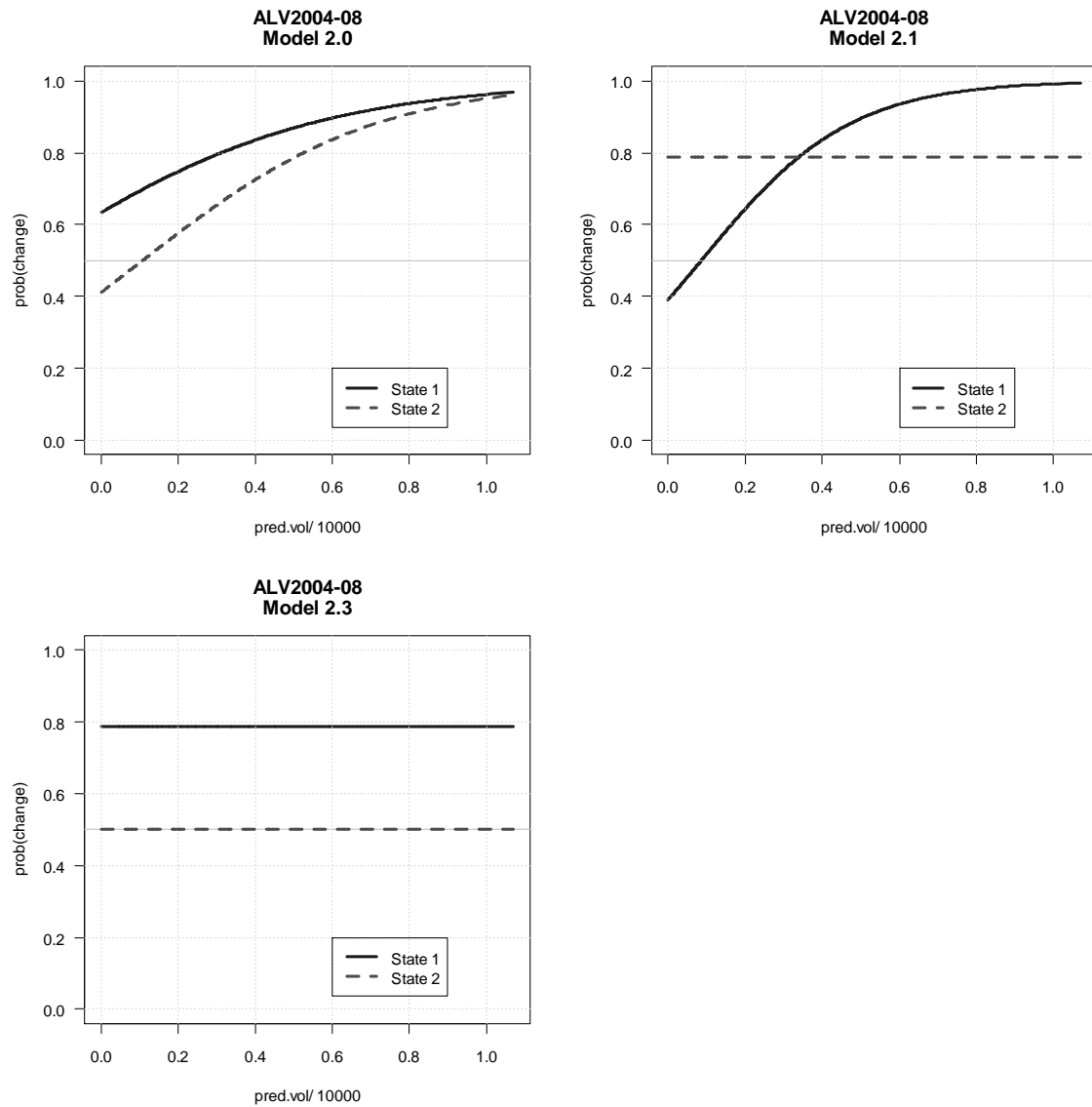
Figure 3.11: Monthly models for Allianz in August, 2004: Model 2.0 was selected.
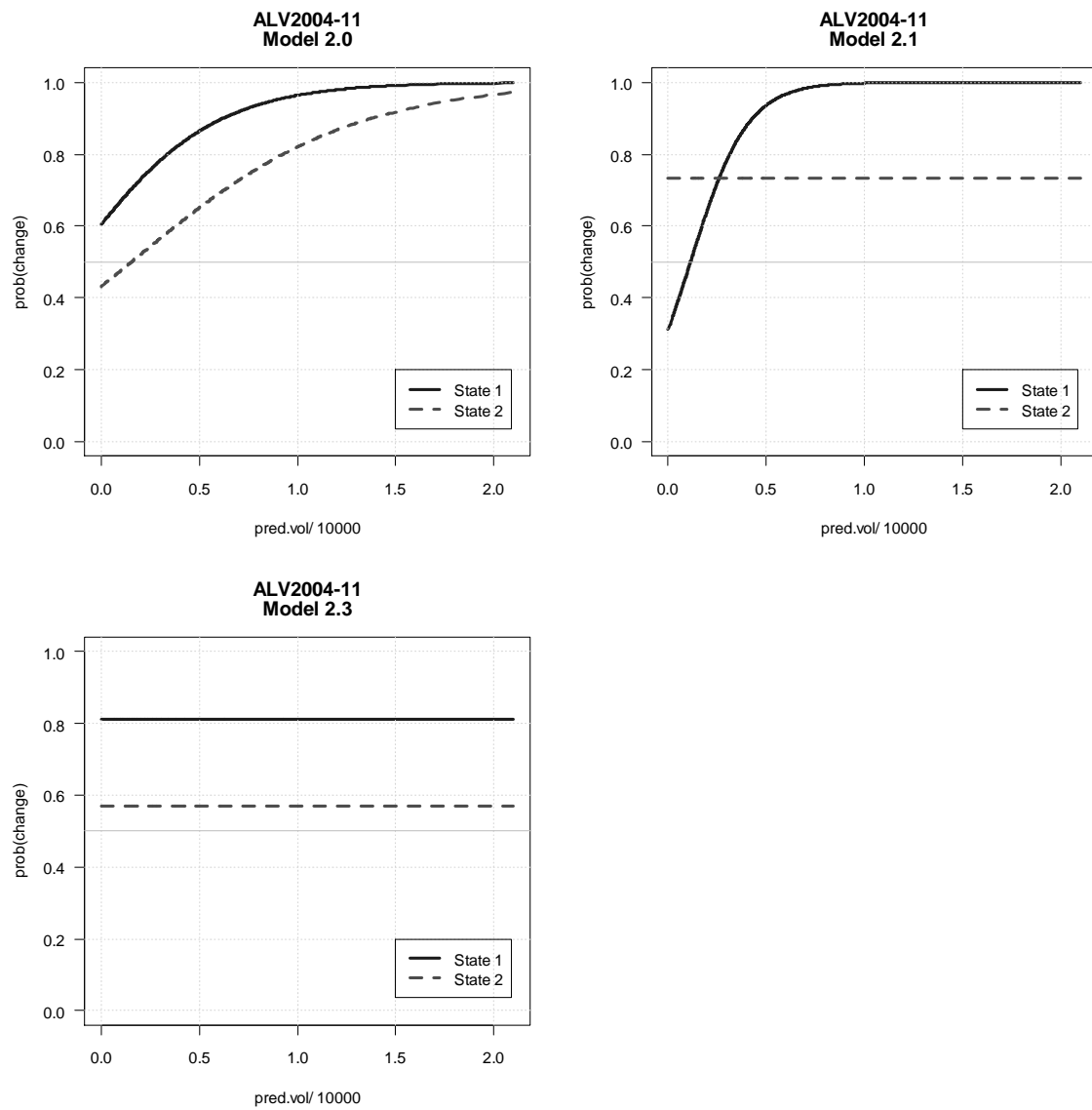
Figure 3.12: Monthly models for Allianz in November, 2004: Model 2.0 was selected.

3.12).

Finally, we repeated the application of the two-state Bernoulli-HMMs considered to the remaining seven shares in all four periods of the year 2004. As for Allianz these models were recalibrated monthly with serial dependence within the volume and return series across the days. For all seven shares in all four periods the AIC preferred Model 2.0. Features of these models are the same as those for Allianz.

3.3.2 Model checking

For each of the eight shares we used, a model is selected for each of the four periods in the year 2004. We used the model-checking techniques in Section 3.2.2 to assess whether these monthly models are adequate. Of the 32 models we examined this way the selected two-state models provide an acceptable fit in 26 cases[33]. These are the fitted models for ALV, EOA, BAS, DBK and SCH (all four periods), LHA (periods 1, 3 and 4) and TUI (periods 1,2 and 4).

In all six of the remaining cases (i.e. LHA for period 2, TUI for period 3 and DCX for all periods) we conclude that the selected models are unsatisfactory. In each of these cases a plot of fitted spline against the identity function reveals that the selected model overestimates the price change probability. In order to correct for these deficiencies we consider alternative models in Section 3.5.

---

[33]H_0 cannot be rejected (technique 1) and the general correspondence between the observed and the mean forecasted probability of price change is acceptable (technique 2).

To illustrate model checking for the out-of-sample period we present two cases: a monthly model that is adequate and the other that is not. For each t in the validation period we compute the conditional distribution (under the selected model) of X_t given the history $\mathbf{x}^{(t-1)}$.

For the first case we present the model selected for ALV for Period 3 in the year 2004. The parameter estimates in this model are:

$$\hat{\mathbf{\Gamma}} = \begin{pmatrix} 0.995 & 0.005 \\ 0.007 & 0.993 \end{pmatrix},$$

$$logit\,(\pi_{t,1}) = 0.548 + 2.686 \cdot N_{t-1},$$

$$logit\,(\pi_{t,2}) = -0.356 + 3.313 \cdot N_{t-1}.$$

We apply the first model checking technique of Section 3.2.2 to the logit of $\hat{\pi}_t$. Figure 3.13 displays the plot of fitted smoothing spline against the identity function. In this Figure we note the following. Firstly, the fitted smoothing spline lies closely around the identity function, particularly over the part of $x-$axis ranging from 0.5 to 0.8. The selected model is satisfactory for the the validation sample in the sense that the null hypothesis cannot be rejected. Secondly, there is an apparent departure from identity function at both tails of the fitted smoothing spline function. The application of the second model-checking technique of Section 3.2.2 reveals that fewer forecast price change probabilities take on either relatively small (between 0.4 and 0.5) or relatively large (between 0.8 and 1) values.

For the second case of our illustrations we present the model selected for LHA
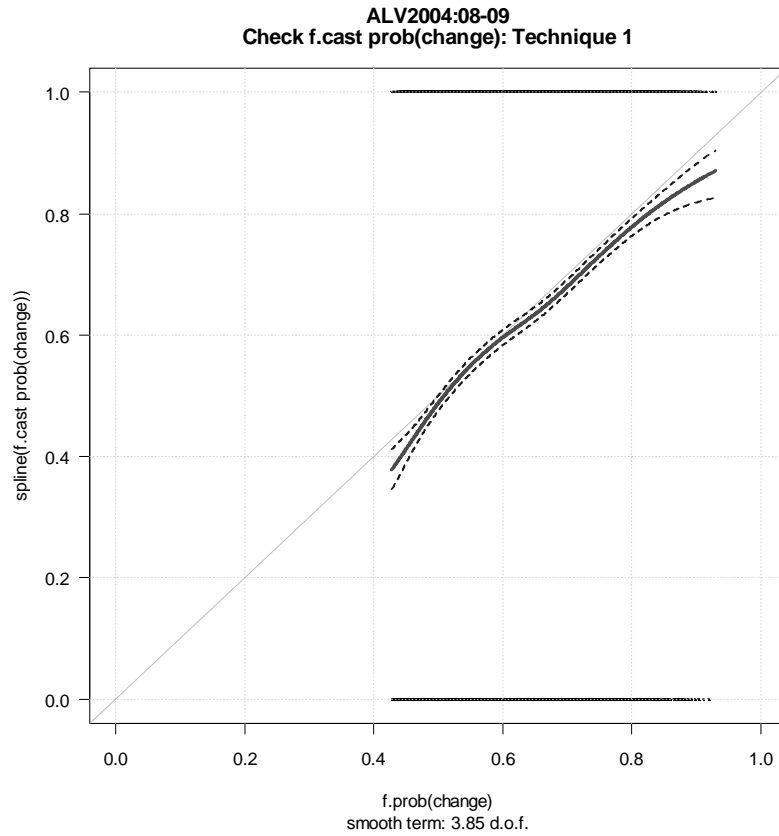
Figure 3.13: Model-checking technique 1 of Section 3.2.2 applied to model selected for Allianz for period 3, 2004. Plot of smoothed values, computed by function `gam` in **R** library `mgcv`, against forecast probabilities (solid line) with confidence bands (dashed). The data points (binary observations against forecast probabilities) appear as solid horizontal lines.

for Period 2 in the year 2004. We applied all three model-checking techniques of Section 3.2.2 to this model. Figure 3.14 depicts the plot for each of the techniques we applied. We note in this figure that the selected model systematically over-estimates the observed price change probability in the out-of-sample period. The plot for technique 2 confirms our conclusion[34]. In the plot for technique 3 a smooth of the standardised residual deviance remains negative over the entire out-of-sample period, which again confirms our conclusion.

3.4   Modelling the conditional return given an occurrence

In this section we model the conditional returns given price change occurrence in the bivariate series of intraday volumes and returns. We consider HMMs as models for the size of return when price change does occur. The state-dependent distributions are normal distributions with zero mean and a variance that depends on N_{t-1}. In order to assess whether such models can be used successfully to forecast the conditional volatility patterns, we used the 32 series of (v_t, r_t) computed in Section 3.3.

---

[34] A closer examination of the forecast probabilities that we computed reveals that $(21585/22462) \times 100 = 96.1\%$ of them achieve values in the range of $0.2 - 0.55$. In each of the bins that cover this range the average of forecast probabilities is still larger than the corresponding relative frequency of price changes that occurred. This indicates that our conclusion is still valid after we ignore in our assessment forecasts that we obtain rarely.
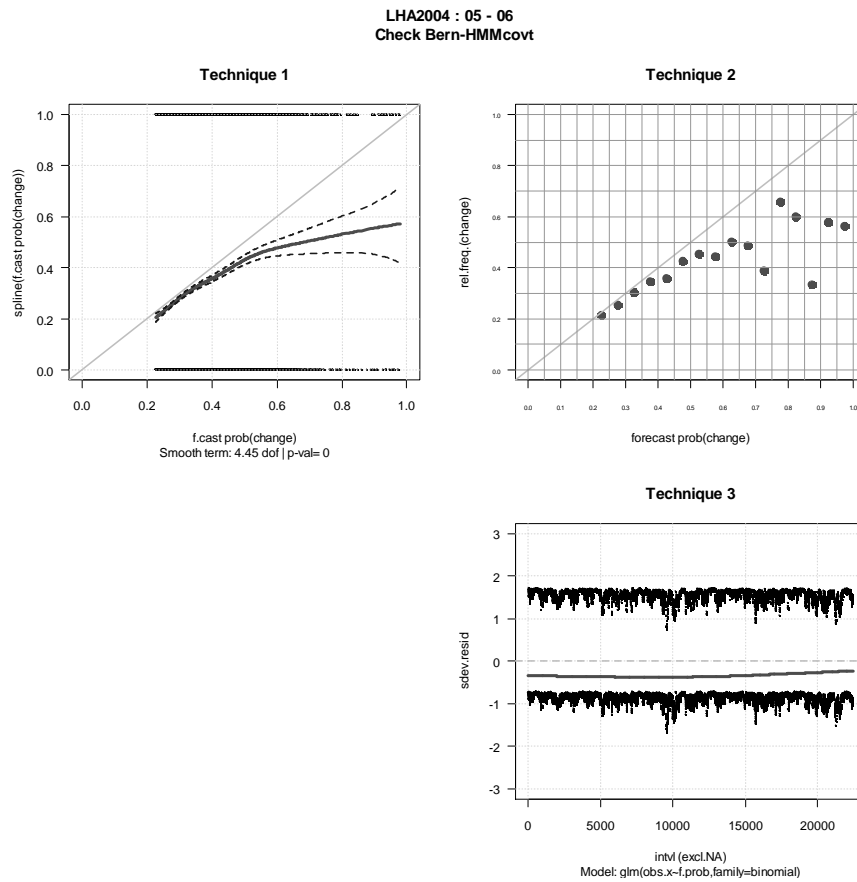
Figure 3.14: Price change indicators of LHA in May-Jun, 2004: applying the model-selection techniques of Section 3.2.2. Smoothed values against forecast probabilities of price change (top left). Relative frequencies (bin width of 0.05) of occurrences of price change against average forecast probabilities (top right). Standardised deviance residuals (under fitted model) and its smooth computed by **R** function `lowess` (solid line).

Model	$\log\left(\sigma_{t,1}^2\right)$	$\log\left(\sigma_{t,2}^2\right)$
2.0	$\alpha_{0,1} + \alpha_{1,1} N_{t-1}$	$\alpha_{0,2} + \alpha_{1,2} N_{t-1}$
2.1	$\alpha_{0,1} + \alpha_{1,1} N_{t-1}$	$\alpha_{0,2}$
2.3	$\alpha_{0,1}$	$\alpha_{0,2}$

Table 3.4: Two-state normal-HMMs: models for the logarithm of state-dependent volatilities of conditional return, given a price change has occurred.

3.4.1   Models fitted

We considered various models for the bivariate time series (n_{t-1}, m_t), with n_{t-1} being the prediction[35] of v_t, and m_t being the returns series in which a missing observation represents a zero return. We use the one-state normal-HMMs[36] as the baseline model.

We then considered three types of two-state normal-HMMs based on a stationary Markov chain. In these models $\sigma_{t,i}^2(n_{t-1})$ is the volatility of nonzero return in intraday period t, which is a function of the predicted volume, if the Markov chain is in state i. Table 3.4 gives the expressions for $\sigma_{t,i}^2(n_{t-1})$ in each type of model.

The likelihood of a normal-HMM incorporating volume is given by Equation (3.11) in Section 3.2.2, where the state-dependent distributions are zero mean

---

[35]The exponential smoother (as given in Section 3.1.3) with $\alpha = 0.5$ was used to compute the series of predicted volumes $\{n_t\}$ on a daily basis. As we have defined in Equation (3.11), we set $N_0 = 0$.

[36]This is the model which assumes that the nonzero returns are realisations of independent normal random variables with zero mean and the logarithm of variance in intraday interval t equals to $a + bn_{t-1}$. The constant term and the coefficient are the parameters.

normal and the covariate is N_{t-1}. The normal-HMMs are fitted via the **R** function `nlm` by direct numerical maximisation of the log-likelihood.

Like in Section 3.3.1, the calibration sample is used to fit the models and the validation sample is used to check the selected model. As before the determination of starting values and choosing the recalibration frequency are the main problems for model fitting. In this subsection we first describe the method we used to determine the starting values and then we discuss models calibrated on a daily basis, under the heading "Models fitted daily." We end this subsection by discussing "Models calibrated less frequently in a month."

Starting values for the iterations. We first fit the one-state model using zero as the starting value for both the constant term and the coefficient in the model. We then fit the two-state models in the order: Model 2.0, 2.1 then 2.3.

In Model 2.0 we use the parameter estimates of the one-state model as the starting values for the parameters of the model for $\log\left(\sigma_{t,1}^2\right)$. In the model for $\log\left(\sigma_{t,2}^2\right)$ we set $\alpha_{0,2} = \alpha_{1,2} = 0$. In Model 2.1 we use $(\hat{\alpha}_{0,1}, \hat{\alpha}_{1,1})$ and $\hat{\alpha}_{0,2}$ in Model 2.0 as the starting values for $(\alpha_{0,1}, \alpha_{1,1})$ and $\alpha_{0,2}$. In Model 2.3 we use $\hat{\alpha}_{0,1}$ and $\hat{\alpha}_{0,2}$ in Model 2.1 as the starting value for $\alpha_{0,1}$ and $\alpha_{0,2}$.

The starting values for the parameters of the Markov chain are chosen as in Section 3.3.1.

Models fitted daily. We start by applying the models considered to (n_{t-1}, m_t) for Allianz for Period 2 in the year 2004. We fit the models on a daily basis for the

month of May. For all trading days the two-state models are preferred by AIC. But no single type of two-state model from Models 2.0, 2.1 and 2.3 is preferred for all days.

We assess the behaviour of the daily models fitted by inspecting the two-state models with N_{t-1} incorporated into the volatility in both states of the Markov chain. In these models $\sigma^2_{t,i}(n_{t-1})$ is the volatility of nonzero return in interval t, as a function of the predicted volume, if the state is i. Figure 3.15 displays $\sigma^2_{t,i}(n_{t-1})$ in state $i = 1$ and 2, plotted on the same set of axes, of the model fitted in each day. In both states of the Markov chain the majority of the functions are positively sloped. The curves in both states are roughly stable, although the curves in state 2 lie closer to each other than those in state 1.

In the t.p.m. of the underlying Markov chain γ_{12} and γ_{21} are the free parameters. Figure 3.16 depicts $\hat{\gamma}_{ij}$ for $i, j = 1, 2$ of the model fitted in each day. Both free parameters are roughly stable, but there is more fluctuation in $\hat{\gamma}_{12}$.

Models fitted less frequently in a month. We use the two-state normal-HMM with log $\sigma^2_{t,i}(N_{t-1}) = \alpha_{0,i} + \alpha_{1,i}N_{t-1}$ in state $i = 1$ and 2. We apply this model to the calibration sample in Period 2 of the year 2004. We now fit the model less frequently in the month than daily as follows. We first divide the month into sub-periods then we calibrate the model for each of the sub-periods. The three cases we investigate, as in Section 3.3.1, are the model recalibrated four times and twice in the month, and the model fitted monthly. In each of these
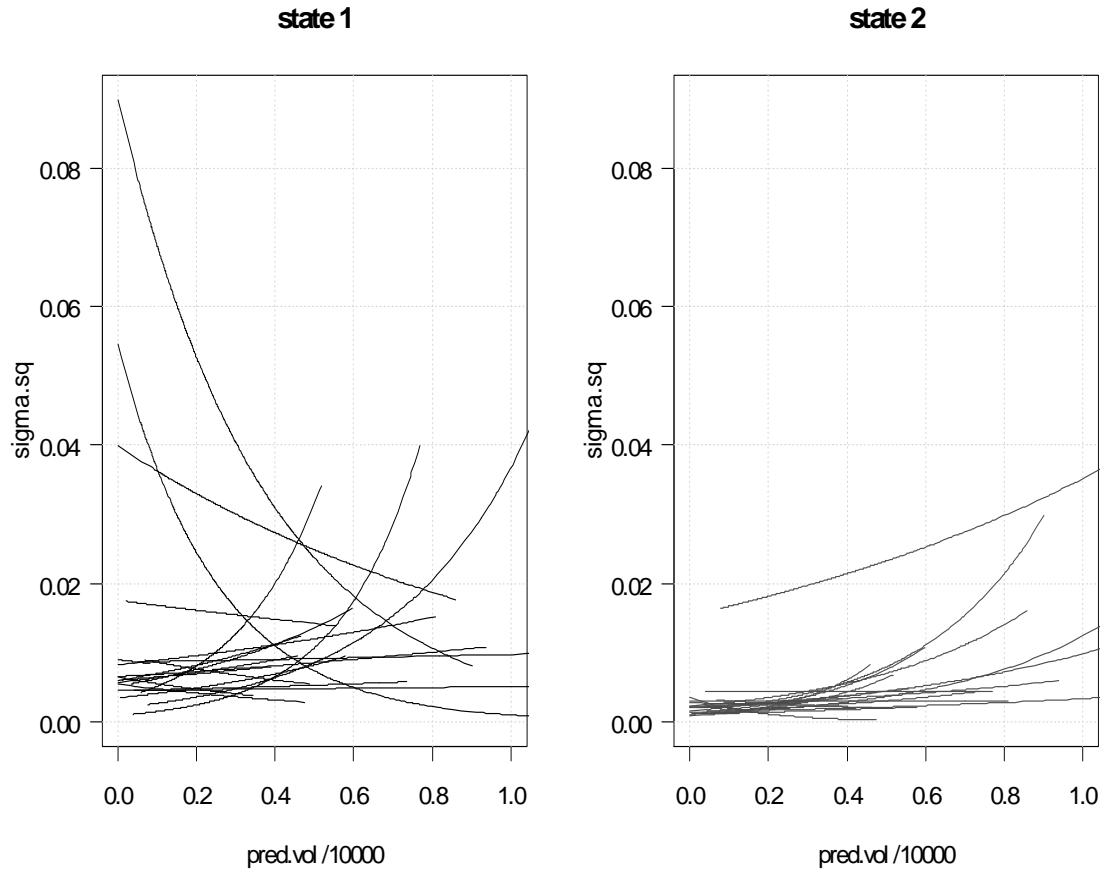
Figure 3.15: Two-state normal-HMMs: state-dependent volatilities, as a function of N_{t-1}, in daily models for Allianz in May, 2004.

Figure 3.16: Two-state normal-HMMs: estimated transition probabilities of Markov chain in daily models for Allianz in May, 2004.

Frequency	Total AIC
daily	**−39488.7**
four times	−39253.8
twice	−39232.4
monthly	−39230.3

Table 3.5: Two-state normal-HMMs: comparison of models for Allianz recalibrated less frequently in May, 2004, by total AIC. The consolidated series is used in each sub-period.

cases we fit the model twice: firstly, by assuming the independence of observations in successive days within each subperiod; and secondly, by assuming some serial dependence within each component of the observations[37] across successive days in each subperiod. For the cases of recalibrating the model four times and twice in the month we sum the values of AIC for the sub-periods. For the case of models fitted daily the we sum the values of AIC for each of the days. Table 3.5 compares the monthly model with the models fitted at each of the three different frequencies in the month, where serial dependence within the sub-periods are assumed in each of these cases, on the basis of their total AIC. It indicates that the total AIC selects the case of fitting the model daily.

We consider again the application of the three types of two-state normal-HMM, as presented in Table 3.4, to the calibration sample for Allianz in Period

---

[37] This is done as in Section 3.3.1.

Model	AIC
2.0	$-\mathbf{39230.3}\,(-39488.7)$
2.1	$-39177.1\,(-39395.1)$
2.3	$-39101.2\,(-39292.6)$

Table 3.6: Two-state normal-HMMs: comparison of monthly models for Allianz in May, 2004, by AIC. The consolidated series is used. In the brackets total AICs for models fitted daily are given.

2 of the year 2004. On computational grounds we recalibrate all these models monthly. Within each component series we postulate serial dependence across the days by using a series of missing observations of length 10^{38} to concatenate adjacent intraday series. The monthly model is an approximation to each daily models, since each parameter of its fit represents the average for the corresponding parameter in each of the daily models fitted. For each type of model we sum the values of AIC for all the daily models. Table 3.6 compares the three types of two-state normal-HMM, recalibrated monthly, on the basis of their AIC[39]. It indicates that Model 2.0 is preferred by AIC. Parameter estimates for this model are presented in row 3 of Table 3.7. Figure 3.17 plots $\sigma_{t,i}^2$ as a function of N_{t-1} in each of the models for Period 2.

In the same way (i.e. monthly recalibration with serial dependence across days in both component series) we apply the two-state normal-HMMs considered

---

[38] As before, our choice is arbitrary.

[39] In the bracket we give the value of total AIC for the corresponding daily models.

ALV	$\hat{\alpha}_{0,1}$	$\hat{\alpha}_{0,2}$	$\hat{\alpha}_{1,1}$	$\hat{\alpha}_{1,2}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{22}$	$\hat{\delta}_1$	$\hat{\delta}_2$	#(obs.)
03.2004	-6.208	-4.912	1.600	1.571	0.979	0.939	0.744	0.256	18012
05.2004	-4.941	-6.186	1.002	1.165	0.965	0.986	0.286	0.714	15044
08.2004	-6.357	-4.863	1.500	0.693	0.984	0.949	0.761	0.239	14905
11.2004	-6.867	-5.641	1.034	0.753	0.976	0.949	0.680	0.320	15666

Table 3.7: Estimates of parameters for the selected two-state normal-HMM for Allianz in March, May, August and November, 2004.
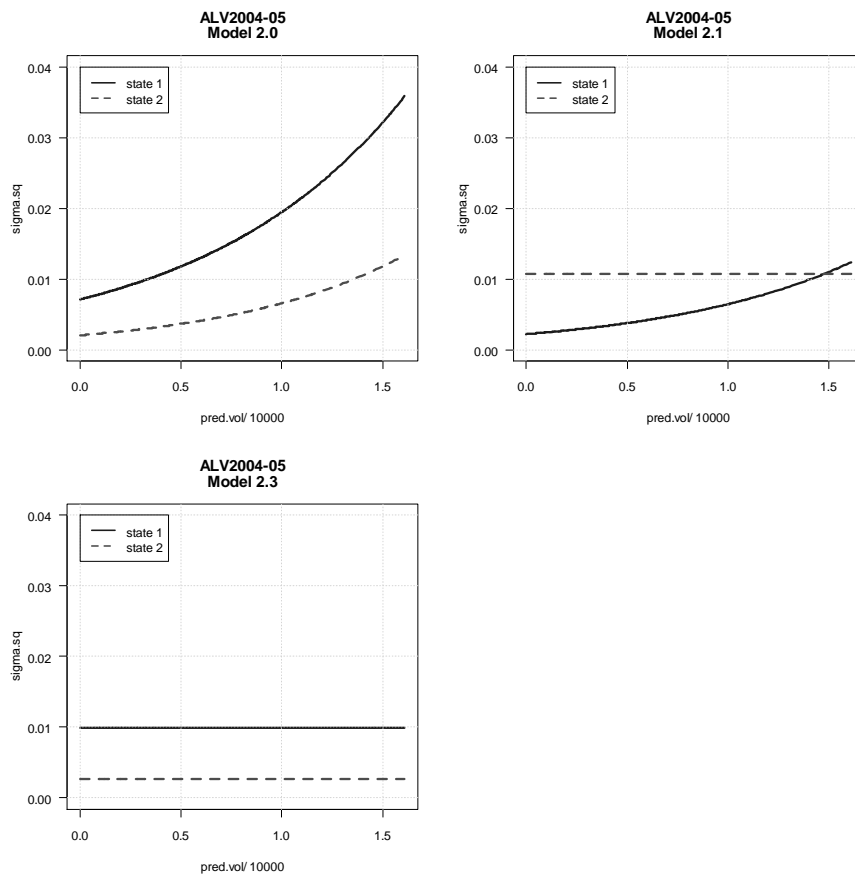
Figure 3.17: Monthly models for Allianz in May, 2004: Model 2.0 was selected.

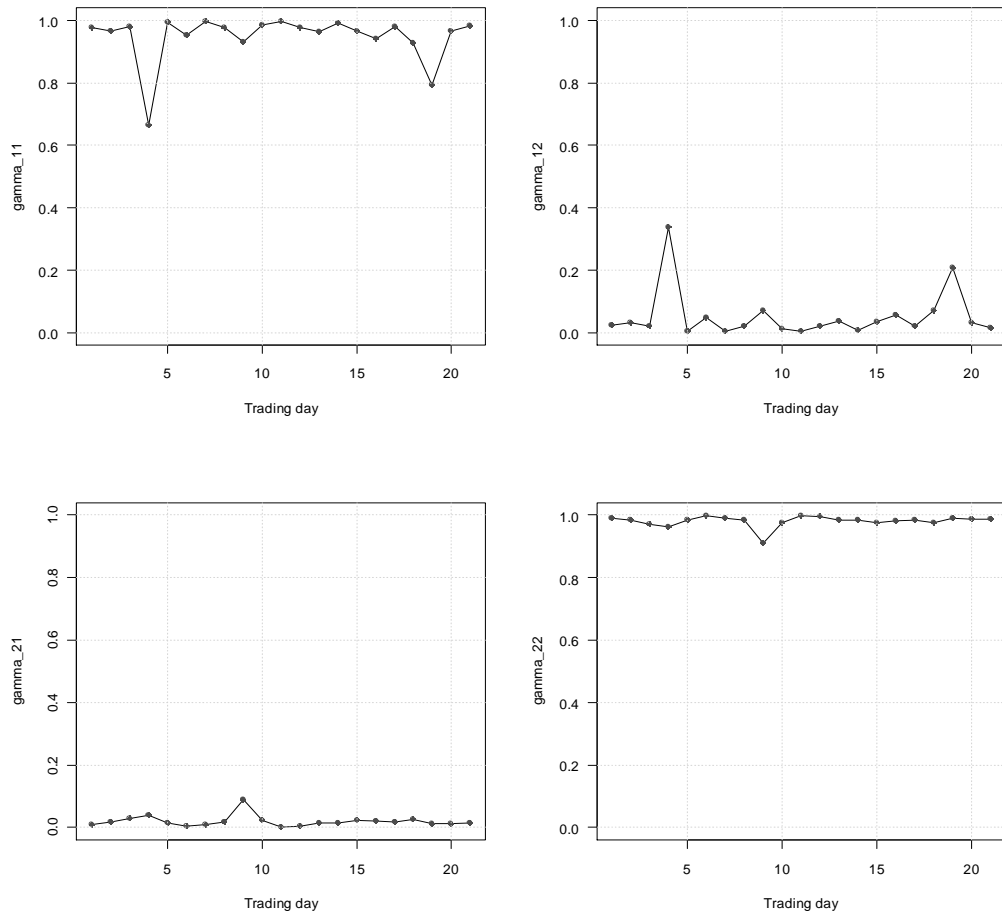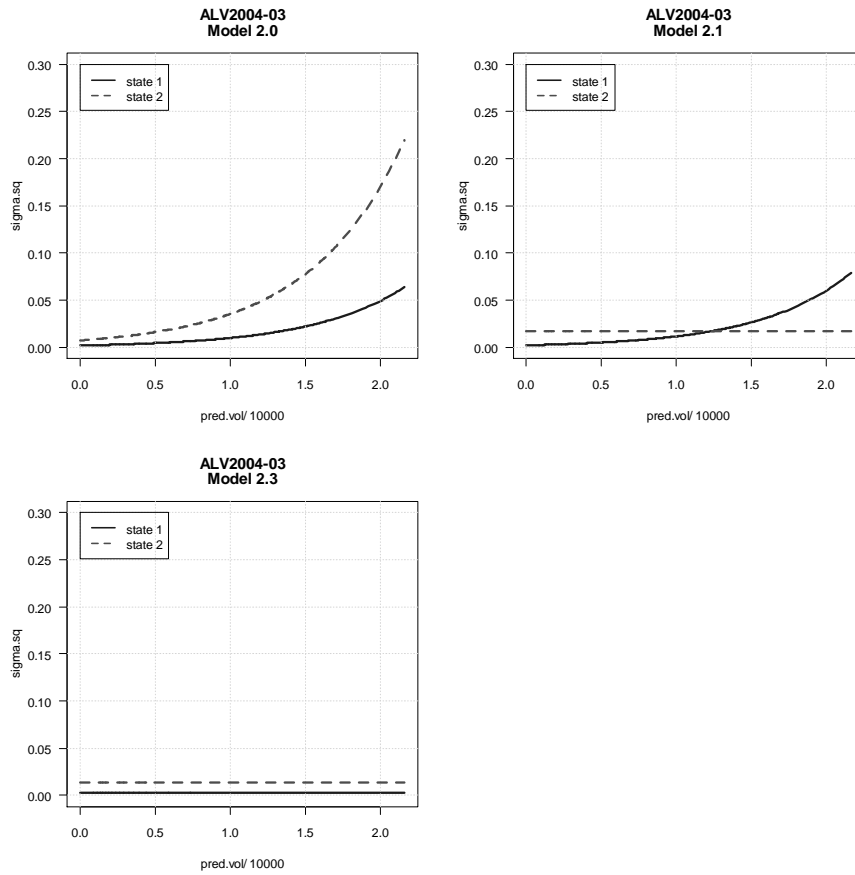Figure 3.18: Two-state normal-HMMs: state-dependent volatilities, as a function of N_{t-1}, in monthly models for Allianz in March, 2004. Model 2.0 was selected.

to the calibration sample for Allianz in Periods 1, 3 and 4 of the year 2004. For the calibration sample in all three periods the AIC prefers Model 2.0. Parameter estimates for the selected model for each of the periods are presented in rows 2, 4 and 5 of Table 3.7. Figures 3.18, 3.19 and 3.20 plot $\sigma_{t,i}^2$, as a function of N_{t-1}, in the models for Periods 1, 3 and 4 respectively. In all these periods Model 2.0 was selected. Two notable features of these models are as follows. Firstly,
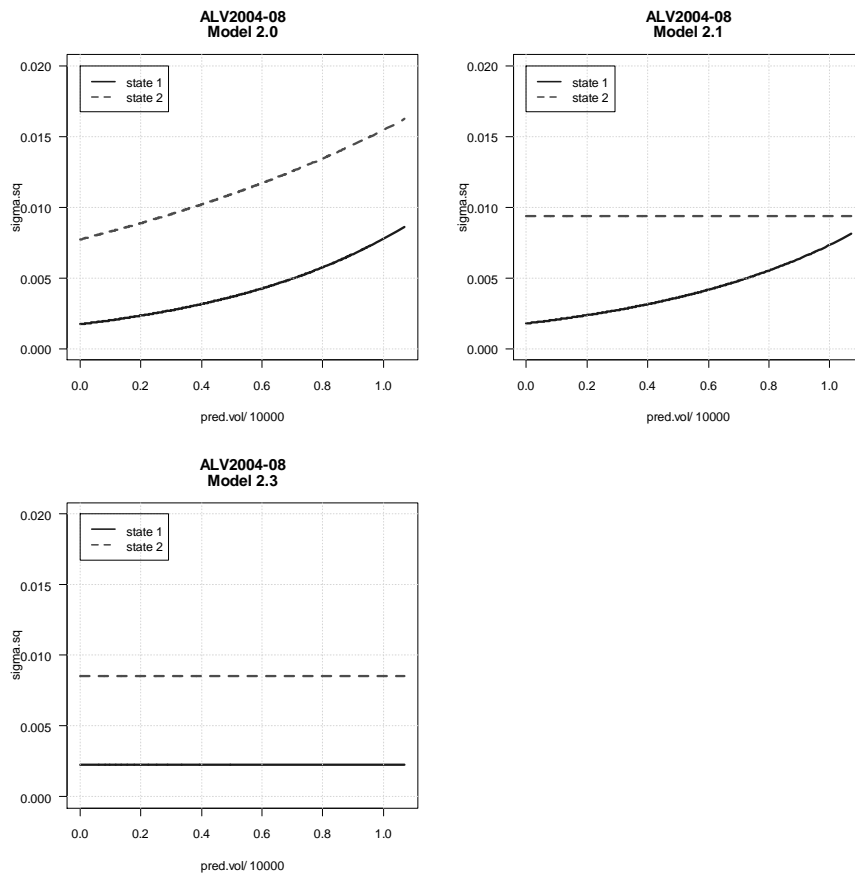
Figure 3.19: Monthly models for Allianz in August, 2004: Model 2.0 was selected.
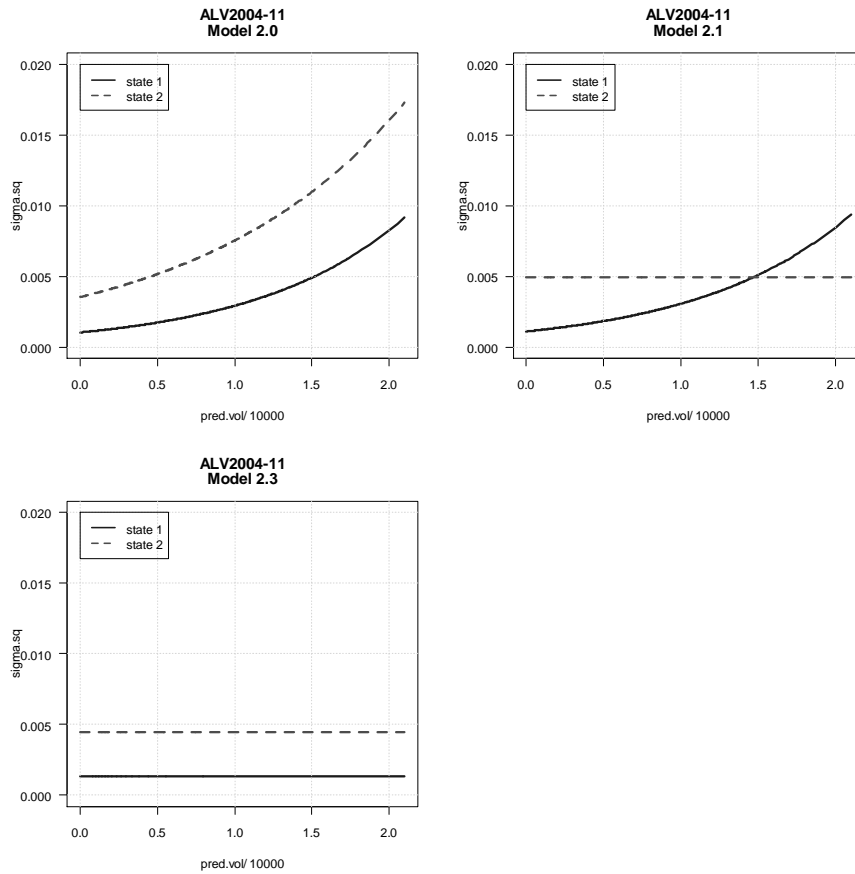
Figure 3.20: Monthly models for Allianz in November, 2004: Model 2.0 was selected.

the state-dependent volatilities $\sigma^2_{t,i}$ are monotonically increasing as a function of N_{t-1}. Secondly, the functions $\sigma^2_{t,i}$ do not cross; the volatility of nonzero return in one of the states is higher than that in the other state for all N_{t-1}. For each of the four periods we investigated the estimated t.p.m. of the Markov chain has diagonal elements close to 1. This indicates persistence in each of the two states. The corresponding stationary distribution $\left(\hat{\delta}_1, \hat{\delta}_2\right)$ is also given in Table 3.7. In March 2004, for example, state 1 is more likely than state 2, because the estimated stationary distribution is $(0.744, 0.256)$.

Finally, we repeat the application of the two-state normal-HMMs considered to the remaining seven shares for all four periods in the year 2004. As for Allianz we recalibrate these models monthly and for each series used we assumed serial dependence within the volumes and returns series across the days. For all seven shares and in all four periods the AIC prefers Model 2.0. Features of these models are the same as those for Allianz.

3.4.2   Model checking

For each of the eight shares we used, a model is selected for each of the four periods in the year 2004. We analyse the (forecast) NPRs, as defined in Section 3.2.2, to assess whether these monthly models are adequate. All 32 models we examined in this way provided an acceptable fit.

To illustrate model checking for the out-of-sample period we present the model

selected for ALV for Period 3 in the year 2004. The parameter estimates in this

model are:

$$\hat{\mathbf{\Gamma}} = \begin{pmatrix} 0.984 & 0.016 \\ 0.051 & 0.949 \end{pmatrix},$$

$$\log\left(\sigma_{t,1}^2\right) = -6.357 + 1.500 \cdot N_{t-1},$$

$$\log\left(\sigma_{t,2}^2\right) = -4.863 + 0.693 \cdot N_{t-1}.$$

For each t in the validation sample we computed the forecast distribution (under

the selected model) of the conditional return (given price change has occurred)

M_t. Figure 3.21 displays the histogram of the (forecast) UPRs and the plots of

the (forecast) NPRs.

In Figure 3.21 the histogram of the UPRs does not appear to be $U(0, 1)$. The

UPRs here reflect two features of the conditional returns. Firstly, the conditional

returns are nonzero. The values of the UPR that lie in the range $0.45 - 0.55$ occur

less frequently than expected. That's because the value of UPR cannot be 0.5

combined with the fact that the price change are discrete-valued[40]. Secondly, the

conditional returns take on a finite number of values, if the price level varies slowly

over the day. We have computed the UPRs by applying continuous distributions

to the conditional returns which are discrete in nature. The discreteness of the

UPRs is reflected in its histogram as peaks at bins 5, $7 - 9$, $12 - 14$ and 16.

In Figure 3.21 the histogram of the NPRs appears to be well approximated by

the standard normal, because the bin width used here is quite large. In the qq-plot

---

[40] The minimum amount of change allowed by XETRA is EUR0.01. Price changes involving
volume-weighted average prices we computed, however, are continuous-valued.
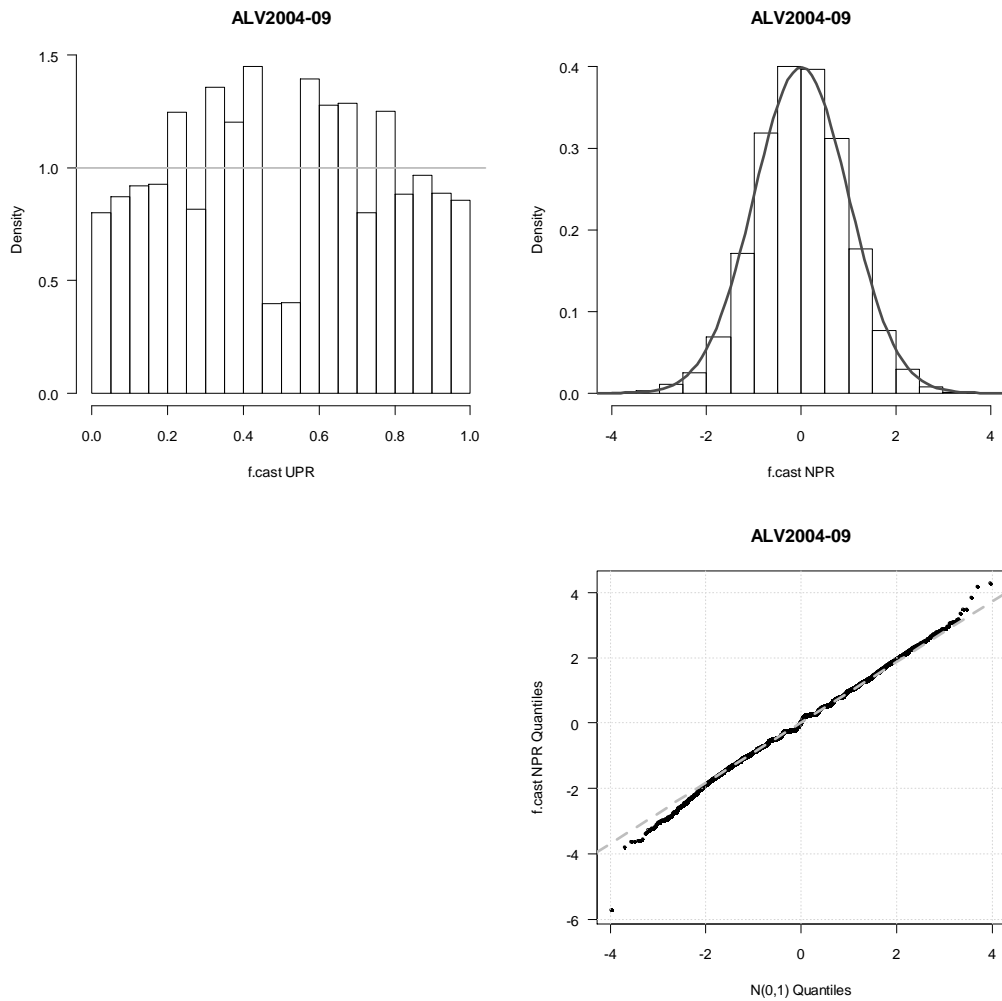
Figure 3.21: Validation sample of Allianz in Period 3, 2004: plots of forecast pseudo residuals (under selected two-state normal HMM incorporating volume). Histogram of forecast UPR (top left) and NPR (top right). Quantile-quantile plot of forecast NPR with standard normal quantiles on horizontal axis.

of the NPRs the points lie closely around the identity function, but deviate from linearity at both tails of the plot. This indicates that the selected model fits well to the observations that are not extreme (relative to the model and all preceding observations). A notable feature in this qq-plot is the discreteness of NPRs[41]. This confirms the discrete nature of the conditional returns, and that the normal distribution is only an approximation of the forecast distribution of M_t.

3.5 Models for the state-dependent probabilities of change

In Section 3.3.2 we examined the Bernoulli-HMM incorporating volume selected for each share in each of the four periods in the year 2004. We concluded there that the selected models were inadequate in the following six cases: LHA (Period 2), DCX (Periods 1-4), TUI (Period 3). In all these cases the selected model over-estimated the probabilities of price change. Forecast probabilities with values greater than 0.5 correspond particularly poorly to the occurrences of price change in the validation sample.

One way to correct this over-estimation is by considering transformations of the predicted volume in the model for the price change probabilities in state i, $\pi_{t,i}$. For this purpose we considered two types of transformations: the square root (model 1) and the (natural) logarithm[42] (model 2). These models are presented in Table 3.8, where $\alpha_{0,i}, \alpha_{1,i}$ are the state-dependent coefficients. In order to allow

---

[41] This is, at values at which the NPRs cluster the plot appears horizontal.

[42] Here, $\log(0)$ is set equal to $\log\left(10^{-323}\right)$: this is the smallest finite natural logarithm value allowed by **R**.

model	$logit\left(\pi_{t,i}\right)$
1	$\alpha_{0,i} + \alpha_{1,i} \cdot \sqrt{N_{t-1}}$
2	$\alpha_{0,i} + \alpha_{1,i} \cdot \log\left(N_{t-1}\right)$

Table 3.8: Models for the logit of the probability of price change in state i (for i=1,2).

flexibility in the form of its dependence on predicted volumes, we also model the means of the binary series by applying smoothing splines.

In the case of LHA for period 2 we fitted the following four models to the calibration sample. The first two models we fitted are the two-state Bernoulli-HMM in which we use model 1 and then model 2 in Table 3.8. The third model we fitted is a smoothing spline function for the logit of price change probabilities π_t[43] and the last model we used is the logistic regression model for π_t. We then applied the model-checking techniques of Section 3.2.2 to check the fit of these models in the out-of-sample period. We found that all four models still over-estimate the probability of price change. In order to check the fit in other sample periods we re-examine two of the fitted models: the smoothing spline function and the logistic regression model.

We first check the fit in the in-sample period by applying the model-checking techniques of Section 3.2.2 to these two models. Figure 3.22 displays the plot for technique 1 applied to the fitted logistic regression model. We note in this figure

---

[43]The function is fitted using the **R** function `gam` provided in the `mgcv` library. The degrees of freedom is selected via cross validation.
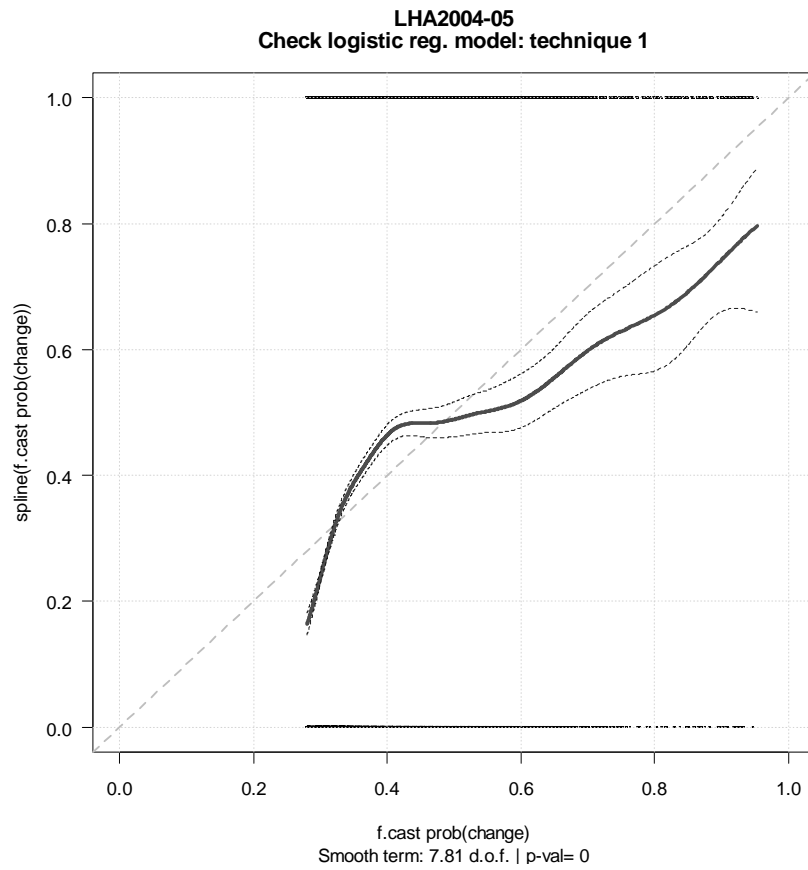
Figure 3.22: Model-checking technique 1 of Section 3.2.2 applied to calibration sample of Lufthansa in Period 2, 2004. Forecast probabilities of price change under fitted logistic regression model.

that the plot departs strikingly from the identity function. This indicates that the logistic regression model is inadequate also for the in-sample period. Figure 3.23 displays technique 1 applied to the fitted smoothing spline. We note in this figure that the plot lies closely around the identity function. This indicates that there is insufficient evidence to reject the null hypothesis. By comparing the plot in Figure 3.23 with that in Figure 3.22 we conclude that the smoothing spline function[44] is a more successful model than the logistic regression model. We then check the fit in the first out-of-sample day. From the plots for technique 1 we again conclude that the smoothing spline function is a more successful model than the logistic regression model.

For each of the remaining five cases for which the selected Bernoulli-HMMs is inadequate we fit the four alternative models we considered and then check the models as we have done for LHA for period 2. In each of these cases we arrived at the same conclusions as those for LHA for period 2.

3.6 Comparisons of improvement in forecast accuracy

We assess the forecast performance of our two-component model by applying the $RMSFE$ (page 152, Equation (3.20)). Table 3.9 presents this ratio computed for each of the 32 cases investigated in this chapter. In this table we note the following features. Firstly, our two-component model has in general performed better than the baseline model: in 25 of the 32 cases the $RMSFE$ has a value

---

[44]We use here 7.98 degrees-of-freedom that we obtained by using the cross-validation method.
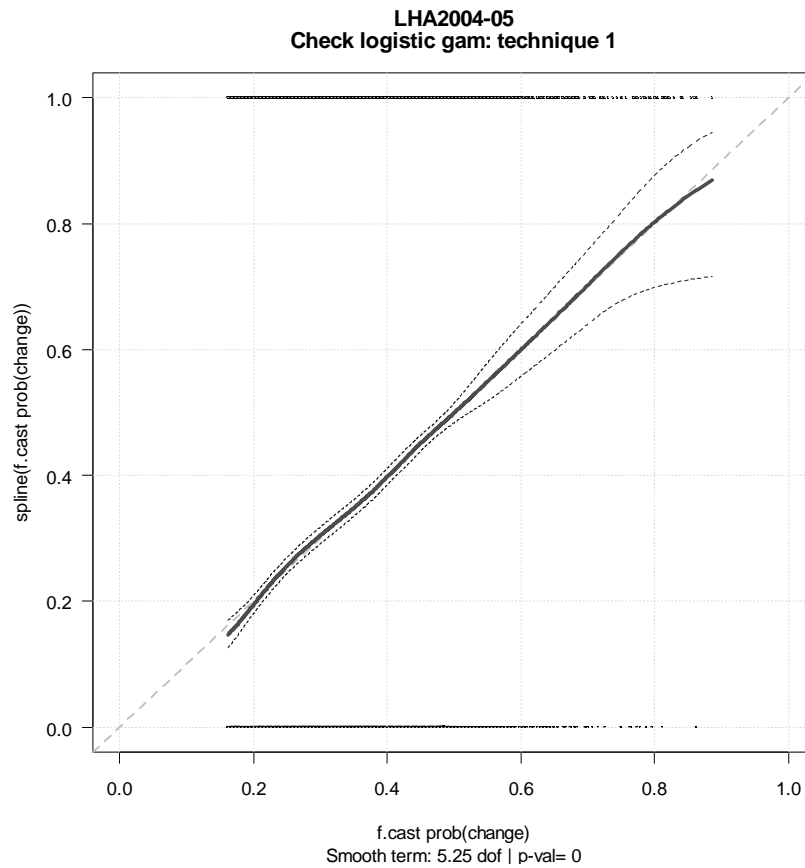
Figure 3.23: Model-checking technique 1 of Section 3.2.2 applied to calibration sample of Lufthansa in Period 2, 2004. Forecast probabilities of price change under fitted smoothing spline function.

Period	1	2	3	4
ALV	0.993	1.014	1.036	2.443
	(0.171)	(0.073)	(0.111)	(0.125)
BAS	1.006	1.017	1.065	1.208
	(0.096)	(0.080)	(0.106)	(0.300)
DBK	1.135	1.165	0.885	2.192
	(0.425)	(0.330)	(0.085)	(0.267)
DCX	≈ 0.000	1.035	1.052	1.171
	(0.094)	(0.080)	(0.106)	(0.300)
EOA	1.609	1.016	1.124	1.049
	(0.076)	(0.055)	(0.070)	(0.152)
LHA	1.029	1.028	1.019	1.091
	(0.099)	(0.190)	(0.078)	(0.234)
SCH	0.002	1.014	0.977	1.021
	(0.050)	(0.048)	(0.029)	(0.163)
TUI	0.990	1.947	1.065	0.994
	(0.040)	(0.053)	(0.216)	(0.050)

Table 3.9: Improvement in forecast accuracy achieved by the two-component HMM incorporating volume: RMSFE for eight shares in Periods 1-4, 2004. The alpha estimated from the calibration sample is given in the bracket.

exceeding 1. The success rate achieved by our model varies across periods (all eight shares in period 2; four shares only in period 1) and between shares (all periods for BAS, EOA and LHA; two periods only for SCH and TUI).

Secondly, our two-component model fared particularly poorly in two of the cases: DCX and SCH (period 1). A closer examination of the sample in each case revealed the following. In both cases the range for predicted volume N_{t-1} in the out-of-sample period is considerably wider than that in the calibration period. Extrapolations thus occurred when the variance of conditional return in state i, $\sigma_{t,i}^2$ was predicted beyond the range for N_{t-1} on which the model was fitted. An additional feature of the predicted volumes in both cases is that the variation in the out-of-sample period is considerably higher than that in the calibration period. The application of extrapolation is thus unreliable, since there is considerable doubt on the validity of the fitted model for the out-of-sample period.

3.7 Summary and discussion

In this chapter we forecast the volatilities of intraday return. We found that the squared returns depend on the volumes traded. We proposed to decompose the returns into two components: the occurrences of price change and the conditional returns given price change has occurred. We applied the Bernoulli-HMMs incorporating volume to model the first component, and the normal-HMMs incorporating volume to model the second.

For each component we examined the transaction data for eight shares covering four separate periods in the year 2004. In our model selection exercises we found that the models fitted daily are preferred (by AIC) to the model fitted monthly. We fitted our model monthly nevertheless, since the computational effort required is relatively modest. This made our analysis of eight shares over four periods computationally feasible.

In our application of the Bernoulli-HMMs incorporating volume we found that the selected models were inadequate in 6 of the 32 cases investigated. In all six cases the models over-estimated the probability of price change in the out-of-sample periods. In order to correct this over-estimation we fitted four alternative models to the binary series. We found that the smoothing spline function for $logit\,(\pi_t)$ provided adequate fit to the calibration sample.

In our application of the normal-HMMs incorporating volume we found that the discreteness of conditional returns (given price change has occurred) was unaccounted for in these models. The normal distribution can thus be regarded as an approximation only. We also found that these models were not adequate to describe the frequency of extreme observations.

In our assessment of the forecast performance we found that our two-component model are superior (according to $RMSFE$) to the baseline model in 25 of the 32 cases investigated. A closer examination of the remaining seven cases revealed that the forecasts derived under our model were unreliable when the state-dependent

variances were predicted at values that were outside the range of predicted volumes to which the models were fitted. In these cases the predictions of the state-dependent variances were computed by assuming that the regression lines fitted to the calibration sample holds true for values of N_{t-1} outside the range of the observations. These regression lines, however, can grossly misrepresent observations in the out-of-sample period. Thus they provided predictions that can differ substantially from the observed values.

CHAPTER 4

CONCLUDING DISCUSSION AND FUTURE

RESEARCH

The aim of this dissertation is to study empirically the volatility of intra-daily price changes for transaction data from electronic trading on the XETRA platform of Frankfurt Stock Exchange. In our study we investigated two variables that can be used to measure volatility. These are the time it takes for the price to change by a specified amount, and the amount of price change over a fixed interval of time. The two main questions that we addressed are as follows. Firstly, do the models that assume unobserved heterogeneity in the observations provide better forecasts than those that do not? Secondly, does the trading volume (possibly lagged) helps to improve the forecast of price change volatility?

In Chapter 2 we forecast the waiting time between successive sm-M price changes. We investigated the MSACD models proposed by Hujer et al. (2002) for this purpose. We applied these models to the price durations for three shares. In each of the three periods covered two stages of preprocessing were needed to standardise the series before these models could be fitted. In the first stage the price durations were deseasonalised. However, after deseasonalising, using

a spline function, much of the variability in the series remains; the day-to-day variability in the values of average deseasonalised price duration over a trading day is considerable. Thus in the second stage we standardised the deseasonalised price durations that occurred over each day, by dividing these values by their average over the day. Our findings suggest that the two sources of variability, namely the intraday seasonality[1] and the day-to-day variation, create the chief difficulties when attempting to forecast price durations.

A number of different specifications of the MSACD models were fitted to the standardised price durations. We found that a two-state model with Burr state-dependent distributions was preferred. In most of the selected models we found that, given the Markov chain is in state j, the conditional mean is a linear function of the lagged duration and the preceding conditional mean in state j. A difficulty encountered was that of maximising the likelihood, which has to be done using numerical methods. We found that the results were sensitive to the starting values, some of which lead to a local (rather than the global) maximum, while the others cause the method to diverge. This problem not only reduces the confidence in the estimates but also substantially increases the computational burden.

We assessed the fits of the selected models in the out-of-sample periods. One-month periods were first used for this purpose. We found that when the validation samples were preprocessed before forecasting the standardised durations, the se-

---

[1]There were some regularity in the seasonal pattern. For example, price durations that occurred in the commencement and the conclusion of trading phases are frequently short; price durations that occurred in the lunchtime period (around 13:00 in Germany) are frequently long.

lected models were adequate in the out-of-sample periods. This is what was done in, for example, Hujer et al. (2002) and Hautsch (2004)[2]. However, the price duration forecasts in this case were not truely out-of-sample, since the seasonal and the day-to-day components were determined from data that included observations that were not available at the time of forecast (i.e. at the start of each successive days in an out-of-sample period).

We then considered rolling samples with a fixed window to assess the out-of-sample fits. In each rolling sample we forecasted one day ahead. As our forecast we used the seasonal and the day-to-day component determined from the calibration sample. Out of the nine cases investigated the selected models were adequate in only one case. We found that the parameters of the fitted models vary substantially over the successive rolling samples. In the unsuccessful cases we found considerable variability in both seasonal and the day-to-day components. As a result, it is not clear whether the standardised durations still contain seasonal variation that were unaccounted for, or the MSACD model was indeed inadequate for the standardised durations.

In our applications of the MSACD model we encountered several problems. The performance of this model was sensitive to the instability of the intraday seasonality, since this is a model for the standardised durations. One approach is to introduce covariate information into the MSACD models. This can be done via the state-dependent conditional means or the transition probabilities of the

---

[2]To our knowledge truely out-of-sample forecasting were not investigated in previous papers.

Markov chain. One potential covariate is the total volume of shares traded in the preceding interval of successive significant price changes. This can be used to explain the seasonal variation in the state-dependent conditional mean. The parameters of the MSACD model were estimated using numerical maximization of the likelihood. They can also be estimated by the expectation-maximization (EM) algorithm. Neither method is guaranteed to converge to the global maximum, but the EM algorithm can have a larger circle of convergence (see Bulla and Berzel (2008) in the context of the HMMs). Another variable that could be modelled by the MSACD model is the volume duration[3], which could be used as a measure of liquidity[4].

In Chapter 3 we forecast volatilities of intraday returns. We proposed the two-component HMMs incorporating volume and applied these to model intraday returns for eight shares. In each of the four periods covered we found that a two-state model was preferred for both components of return. For each of these components we found a dependence on predicted volume in both state-dependent distributions of each of the selected models. In 25 out of the 32 out-of-sample periods we found that the volatility forecasts derived under our model are, on average, more accurate than those computed by the exponential smoother, which was used as a baseline model.

In both components of our model we circumvented the forecasting of season-

---

[3]This is defined as the time between consecutive transactions during which a specified number shares is traded.

[4]See Hautsch (2004, Section 3.4.2).

ality by allowing for the influence of predicted volume in the respective state-dependent distributions. However, there are weaknesses in our model. For the price change occurrences component we found that our model over-estimated the probability of occurrence in some cases. Among the alternative models considered for these cases, we modelled the logit of price change probabilities. By using the smoothing spline to estimate the regression of the logit quantity on the predicted volume we found that the resulting model was adequate for the calibration sample in each of the unsuccessful cases. This suggests that using a nonparametric model for $logit(\pi_{t,i})$ in the two-state Bernoulli HMM might be a promising alternative. For the other component of return we found that, conditional on the occurrence of price change, the discreteness of the return was unaccounted for by our model. We also found that the frequency of extreme observations was inadequately described by our model. In this respect a family of heavy-tailed distributions, such as the (student's) $t-$distributions, might be a useful alternative to the normal. In forecasting volatilities we found that, when the state-dependent variances were extrapolated, our model can be invalid in some cases. In our applications we forecast volatility of intraday return 30-second-ahead. This forecast period might be too short for decision-making.

The two-component HMM incorporating volume can also be applied to other measures of volatility. The model can be applied, for example, to the intraday absolute (or squared) return. In that case the exponential distribution could be

used as the state-dependent distribution in the model for the size of price change conditional on the occurrence of a price change.

The conclusion that may be drawn from our findings in Chapters 2 and 3 is that, when intraday observations are available, models based on regular time intervals are more successful in the forecasting of volatility than those that are based on price durations. The main difficulty in the first case is that of forecasting the mean level and seasonality, a problem that has been ignored in the literature.

BIBLIOGRAPHY

[1] Andersen, T.G. and Bollerslev, T. (1997), "Intraday periodicity and volatility persistence in financial markets," *Journal of Empirical Finance*, **4:2-3**, pp. 115-158.

[2] Andersen, T.G.; Bollerslev, T. and Diebold, F.X. (2007), "Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility," *The Review of Economics and Statistics*, **89:4**, pp. 701-720.

[3] Bauwens, L. and Giot, P. (2001), *Econometric modelling of stock market intraday activity*, Dordrecht: Kluwer Academic Publishers.

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

[5] Box, G.E.P.; Jenkins, G.M. and Reinsel, G.C. (1994), *Time Series Analysis: Forecasting and Control*, Third Edition, New Jersey: Prentice-Hall, Inc..

[6] Bulla, J. and Berzel, A. (2008), "Computational Issues in Parameter Estimation for Stationary Hidden Markov Models," *Computational Statistics*, **23:1**, p. 1-18.

[7] Collett, D. (1991), *Modelling Binary Data*, London: Chapman & Hall.

[8] Czado, C.; Müller, G. and Nguyen, T. (2010), "Ordinal- and continuous-

response stochastic volatility models for price changes: an empirical comparison," In: *Statistical Modelling and Regression Structures*, Kneib, T. and Tutz, G. (eds.), pp. 301-320, Berlin Heidelberg: Springer-Verlag.

[9] De Luca, G. and Zuccolotto, P. (2003), "Finite and infinite mixtures for financial durations," *Metron*, **LXI:3**, pp. 431-455.

[10] De Luca, G. and Gallo, G.M. (2004), "Mixture Processes for Financial Intradaily Durations," *Studies in Nonlinear Dynamics & Econometrics*, **8:2**, article 8.

[11] De Luca, G. and Gallo, G.M. (2009), "Time-Varying Weights in Mixture Autoregressive Conditional Duration Models," *Econometric Reviews*, **28:1**, pp. 102-120.

[12] Deutsche Börse AG (2004), "Xetra Release 7.1: Market Model Equities," Xetra Publication, 63 pages. Document dated 10 September 2004. Downloaded from `http://deutsche-boerse.com`.

[13] Diebold, F.X.; Gunther, T.A. and Tay, A.S. (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, **39:4**, pp. 863-883.

[14] Engle, R.F. and Russell, J.R. (1997), "Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model," *Journal of Empirical Finance*, **4:2-3**, pp. 187-212.

[15] Engle, R.F. and Russell, J.R. (1998), "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, **66: 5**, pp. 1127-1162.

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

[17] Gerhard, F. and Hautsch, N. (2002), "Volatility estimation on the basis of price intensities," *Journal of Empirical Finance*, **9:1**, pp. 57-89.

[18] Grammig, J. and Maurer, K. (2000), "Non-monotonic Hazard Functions and the Autoregressive Conditional Duration Model," *Econometrics Journal*, **3**, pp. 16-38.

[19] Hamilton, J.D. (1994), *Time Series Analysis*, New Jersey: Princeton University Press.

[20] Haug, S. and Czado, C. (2006), "Mixed effect models for absolute log returns of ultra high frequency data," *Applied Stochastic Models in Business and Industry*, **22**, pp. 243-267.

[21] Hautsch, N. (2004), *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models*, Lecture Notes in Economics and Mathematical Systems vol. 539, Berlin Heidelberg: Springer-Verlag.

[22] Hujer, R. and Vuletić, S. (2007), "Econometric analysis of financial trade processes to discrete mixture duration models," *Journal of Economic Dynamics & Control*, **31:2**, pp. 635-667.

[23] Hujer, R.; Vuletić, S. and Kokot, S. (2002), "The Markov Switching ACD Model," Finance & Accounting Working Paper Series, No.90, Johann Wolfgang Goethe-University, Frankfurt/Main.

[24] Kwok, S.S.M.; Li, W.K. and Yu, P.L.H. (2009), "The autoregressive conditional marked duration model: statistical inference to market microstructure," *Journal of Data Science*, **7**, pp. 189-201.

[25] Leroux, B.G. and Puterman, M.L. (1992), "Maximum-penalized-likelihood Estimation for Independent and Markov-dependent Mixture Models," *Biometrics*, **48:2**, pp. 545-558.

[26] Lunde, A. (1999), "A Generalized Gamma Autoregressive Conditional Duration Model," Discussion Paper, Aarlborg University.

[27] McAleer, M. and Medeiros, M.C. (2008), "Realized volatility: a review," *Econometric Reviews*, **27:1**, pp. 10-45.

[28] Müller, G. and Czado, C. (2009), "Stochastic volatility models for ordinal-valued time series with application to finance," *Statistical Modelling*, **9:1**, pp. 69-95.

[29] Ng, W.L. (2010), "Dynamic order submission and herding behavior in electronic trading," *The Journal of Financial Research*, **33:1**, pp. 27-43.

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

[31] Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, **23**, pp. 470-472.

[32] Silverman, B. W. (1986), *Density estimation for statistics and data analysis*, London: Chapman and Hall.

[33] Tay, A.C.; Tse, Y.K. and Warachaka, M. (2004), "Transaction-data analysis of marked durations and their implications for market microstructure," *Working Paper*, Singapore Management University.

[34] Russell, J.R. and Engle, R.F. (2005), "A discrete-state continuous-time model of financial transactions prices and times: The autoregressive conditional multinomial-autoregressive conditional duration model," *Journal of Business & Economic Statistics*, **23:2**, pp. 166-180.

[35] Shephard, N.G. (1996), "Statistical aspects of ARCH and stochastic volatility," In *Time Series Models: In econometrics, finance and other fields*, D.R.

Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (eds.), pp. 1-67, London: Chapman & Hall.

[36] Zhang, Q.; Cai, C.X. and Keasey, K. (2009), "Forecasting using high-frequency data: a comparison of asymmetric financial durations models," *Journal of Forecasting*, **28**, pp. 371-386.

[37] Zucchini, W. (2000), "An Introduction to Model Selection," *Journal of Mathematical Psychology*, **44**, p. 41-61.

[38] Zucchini, W. and MacDonald (2009), *Hidden Markov Models for Time Series: An Introduction Using R*, Monographs on Statistics & Applied Probability 110, Boca Raton: Chapman & Hall/CRC.

[39] Zucchini, W.; Raubenheimer, D. and MacDonald, I.L. (2008), "Modeling Time Series of Animal Behavior by Means of a Latent-State Model with Feedback," *Biometrics*, **64**, p. 807-815.

APPENDICES

A.   Transaction data

code	form (German)	form (English)
E	Eröffnung	Open
EA	Eröffnungsauktion	Opening auction
VA	Auktion nach Volatilitätsunterbrechung	Auction after trading halt due to excess volatility
IA		Intraday auction
V	Variabler Handel	Continuous trading
S	Schluß	Close
SA	Schlußauktion	Closing auction

Table 4.1: Code indicating the form of trading.

B.   Markov switching ACD models

Derivation of Hamilton's algorithm

Our aim here is to derive Hamilton's algorithm to compute the likelihood of N consecutive observations $x_1, x_2, \ldots, x_N$ assumed to be generated by a Markov switching model. This model has m states (for $m = 1, 2, \ldots$), a lag order of u (for $u = 1, 2, \ldots$) and state-dependent errors that are independent with common distribution $F\left(\epsilon; \boldsymbol{\theta}_{\epsilon, j}\right)$ (for $j = 1, 2, \ldots, m$).

The Markov switching model consists of two parts. The first is a state process (C_t), for $t = 2, 3, \ldots, N$, which is a (discrete index) Markov chain, with m states (for $m = 1, 2, \ldots$) and transition probability matrix defined by $\boldsymbol{\Gamma} = \left(\gamma_{ij}\right)$ (for $i, j = 1, 2, \ldots, m$). The second part is the state-dependent process (X_t), for $t = (u + 1), \ldots, N$ such that if $C_t = j$ (for $j = 1, 2, \ldots, m$) the model for X_t is $X_t = \mu\left(X_{t-1}, \ldots, X_{t-u}, \boldsymbol{\theta}_{\mu, j}\right) + \epsilon_{t, j}$ and $\left(\epsilon_{t, j}\right)$ are independent with common distribution $F\left(\epsilon; \boldsymbol{\theta}_{\epsilon, j}\right)$.

For $t = 1, 2, \ldots, N$, we define the following notation. Firstly, $\mathbf{C}^{(t)} = \left(C_1, \ldots, C_t\right)'$ is the history of (C_t) up to t; $\mathbf{X}^{(t)}$ and $\mathbf{x}^{(t)}$ are defined analogously. Secondly, $\mathbf{x}_u^t = \left(x_u, \ldots, x_t\right)'$ are the observations from u $(u > 1)$ to t $(t > u)$.

The parameters for the Markov chain are contained in the vector $\boldsymbol{\theta}_{\Gamma} = \left(\gamma_{ij}\right)_{i \neq j}$, where i and $j = 1, 2, \ldots, m$. The parameters for the conditional mean and the error distribution are contained in the vectors $\boldsymbol{\theta}_{\mu} = \left(\boldsymbol{\theta}_{\mu, 1}, \ldots, \boldsymbol{\theta}_{\mu, m}\right)$ and $\boldsymbol{\theta}_{\epsilon} = \left(\boldsymbol{\theta}_{\epsilon, 1}, \ldots, \boldsymbol{\theta}_{\epsilon, m}\right)$, respectively. Then the parameters for the model are contained in

the vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\epsilon, \boldsymbol{\theta}_\Gamma)'$.

Problem statement. Assume that $\mathbf{x}$ is a realisation of (X_t). For $t = 1, 2, \ldots, N$ what is the formula for $\Pr\left(C_t \middle| \mathbf{X}^{(t)} = \mathbf{x}^{(t)}\right)$?

Solution. Let $\Pr\left(C_t \middle| \mathbf{x}^{(t)}\right)$ represent $\Pr\left(C_t \middle| \mathbf{X}^{(t)} = \mathbf{x}^{(t)}\right)$. We make the following assumption.

(A0) $\mathbf{x}$ and $\boldsymbol{\theta}$ are known

(A1) $\Pr\left(C_t \middle| \mathbf{C}^{(t-1)}\right) = \Pr\left(C_t \middle| C_{t-1}\right)$ Markov property

(A2) $f\left(x_t \middle| \mathbf{C}^{(t)}, \mathbf{x}^{(t-1)}\right) = f\left(x_t \middle| C_t, \mathbf{x}^{(t-1)}\right)$ X_t, $\mathbf{C}^{(t-1)}$ independent

(A3) $\Pr\left(C_t \middle| \mathbf{C}^{(t-1)}, \mathbf{x}^{(t-1)}\right) = \Pr\left(C_t \middle| \mathbf{C}^{(t-1)}\right)$ C_t, $\mathbf{X}^{(t-1)}$ independent

We now derive a recursion to compute the likelihood.

Step $t = 1, 2, \ldots, u$ (Initialise)

Take $\mathbf{x}^{(0)}$ to be the empty set. At index points $t = 1, 2, \ldots, u$:

1. Set the distribution conditioned on $\mathbf{x}^{(t-1)}$ to its unconditional distribution:

$$\Pr\left(C_t \middle| \mathbf{x}^{(t-1)}\right) = \Pr\left(C_t\right) \quad \text{distribution of } C_t.$$

$$f\left(x_t \middle| C_t, \mathbf{x}^{(t-1)}\right) = f\left(x_t \middle| C_t\right) \quad \text{distribution of } X_t \text{ conditioned on } C_t.$$

2. Update:

$$f\left(x_t, C_t\right) = f\left(x_t \middle| C_t\right) \Pr\left(C_t\right) \tag{B.1}$$

$$f\left(x_t\right) = \sum_{j=1}^{m} f\left(x_t, C_t = j\right) \tag{B.2}$$

$$= \sum_{j=1}^{m} f\left(x_t \middle| C_t = j\right) \Pr\left(C_t = j\right) \tag{B.3}$$

$$\Pr\left(C_t \middle| x_t\right) = \frac{f\left(x_t, C_t\right)}{f\left(x_t\right)} \tag{B.4}$$

Equation (B.1) is the definition of conditional probability, Equation (B.2) represents the law of total probability, and equations (B.3) and (B.4) follows by definition of conditional probability.

3. Predict:

$$\Pr\left(C_{t+1}\middle|\mathbf{x}^{(t)}\right) = \sum_{j=1}^{m}\Pr\left(C_{t+1}, C_t = j\middle|\mathbf{x}^{(t)}\right) \tag{B.5}$$

$$= \sum_{j=1}^{m}\Pr\left(C_{t+1}\middle|C_t = j, \mathbf{x}^{(t)}\right)\Pr\left(C_t = j\middle|\mathbf{x}^{(t)}\right) \tag{B.6}$$

$$= \sum_{j=1}^{m}\Pr\left(C_{t+1}\middle|C_t = j\right)\Pr\left(C_t = j\middle|\mathbf{x}^{(t)}\right) \tag{B.7}$$

Equation (B.5) represents the law of total probability, Equation (B.6) follows by definition of conditional probability and Equation (B.7) follows by the independence of C_{t+1} and $\mathbf{x}^{(t)}$.

Step $t = (u+1), (u+2), \ldots, N$

(Updating equation)

Contribution to likelihood:

$f\left(x_t\middle|\mathbf{x}^{(t-1)}\right) = \sum_{j=1}^{m} f\left(x_t\middle|C_t = j, \mathbf{x}^{(t-1)}\right)\Pr\left(C_t = j\middle|\mathbf{x}^{(t-1)}\right).$

Update C_t:

$\Pr\left(C_t\middle|\mathbf{x}^{(t)}\right) = \frac{f\left(x_t\middle|C_t, \mathbf{x}^{(t-1)}\right)\Pr\left(C_t\middle|\mathbf{x}^{(t-1)}\right)}{f\left(x_t\middle|\mathbf{x}^{(t-1)}\right)}.$

(Prediction equation)

$$\Pr\left(C_{t+1}\middle|\mathbf{x}^{(t)}\right) = \sum_{j=1}^{m}\Pr\left(C_{t+1}\middle|C_t = j\right)\Pr\left(C_t\middle|\mathbf{x}^{(t)}\right).$$

Likelihood of $\mathbf{x}_{u+1}^N$ conditioned on $\mathbf{x}^{(u)}$

$$L_N = f\left(\mathbf{x}_{u+1}^N \middle| \mathbf{x}^{(u)}; \boldsymbol{\theta}\right) = \prod_{t=u+1}^{N} f\left(x_t \middle| \mathbf{x}^{(t-1)}\right).$$

log-likelihood:

$$l_N = \log L_N = \sum_{t=u+1}^{N} \log f\left(x_t \middle| \mathbf{x}^{(t-1)}\right).$$

Derivation of the explicit formula for the likelihood

Problem statement. Assume $\mathbf{x}$ is a realisation of $\{X_t\}$. Derive an explicit formula for the likelihood L_N conditioned on $x_1, \ldots, x_u$ of an m-state Markov-switching model.

Solution. The likelihood L_N of a Markov switching process $\{X_t\}$ conditioned on $x_1, \ldots, x_u$ can be written as

$$
\begin{aligned}
L_N &= f\left(x_{u+1}, \ldots, x_N \middle| x_1, \ldots, x_u\right) \\
&= \sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\cdots\sum_{j_N=1}^{m} f\left(\mathbf{x}_{u+1}^N, C_1 = j_1, \ldots, C_N = j_N \middle| \mathbf{x}^{(u)}\right) \qquad \text{(B.8)}
\end{aligned}
$$

The summands of Equation (B.8) can be represented by $f\left(\mathbf{x}_{u+1}^N, \mathbf{C}^{(N)} \middle| \mathbf{x}^{(u)}\right)$. Then applying the definition of conditional probability to $f\left(\mathbf{x}_{u+1}^N, \mathbf{C}^{(N)} \middle| \mathbf{x}^{(u)}\right)$, the summands of Equation (B.8) can be written as

$$f\left(\mathbf{x}_{u+1}^N, \mathbf{C}^{(N)} \middle| \mathbf{x}^{(u)}\right) = f\left(\mathbf{x}_{u+1}^N \middle| \mathbf{C}^{(N)}, \mathbf{x}^{(u)}\right) \cdot \Pr\left(\mathbf{C}^{(N)} \middle| \mathbf{x}^{(u)}\right) \qquad \text{(B.9)}$$

The first term on the right-hand side of Equation (B.9) can be written as:

$$f\left(\mathbf{x}_{u+1}^{N}\middle|\mathbf{C}^{(N)},\mathbf{x}^{(u)}\right)$$

$$= f\left(x_{N},\mathbf{x}_{u+1}^{N-1}\middle|\mathbf{C}^{(N)},\mathbf{x}^{(u)}\right) \tag{B.10}$$

$$= f\left(x_{N}\middle|\mathbf{x}_{u+1}^{N-1},\mathbf{C}^{(N)},\mathbf{x}^{(u)}\right)\cdot f\left(\mathbf{x}_{u+1}^{N-1}\middle|\mathbf{C}^{(N)},\mathbf{x}^{(u)}\right), \tag{B.11}$$

where Equation (B.11) follows from applying the definition of conditional probability to Equation (B.10). After we have applied the definition of conditional probability to $f\left(x_{t},\mathbf{x}_{u+1}^{t-1}\middle|\mathbf{C}^{(N)},\mathbf{x}^{(u)}\right)$, for $t = N, (N-1),\ldots,(u+2)$, the first term on the right-hand side of Equation (B.9) can be written as a product

$$f\left(\mathbf{x}_{u+1}^{N}\middle|\mathbf{C}^{(N)},\mathbf{x}^{(u)}\right) = f\left(x_{u+1}\middle|\mathbf{C}^{(N)},\mathbf{x}^{(u)}\right)\prod_{t=u+2}^{N} f\left(x_{t}\middle|\mathbf{x}_{u+1}^{t-1},\mathbf{C}^{(N)},\mathbf{x}^{(u)}\right) \tag{B.12}$$

The first factor in Equation (B.12) can be re-written as

$$f\left(x_{u+1}\middle|\mathbf{C}^{(N)},\mathbf{x}^{(u)}\right) = f\left(x_{u+1}\middle|C_{u+1},\mathbf{x}^{(u)}\right), \tag{B.13}$$

because X_{u+1} depends only on C_{u+1} and $\mathbf{x}^{(u)}$. For $t = (u+2),(u+3),\ldots,N$, the factors in Equation (B.12) can be re-written as

$$f\left(x_{t}\middle|\mathbf{x}_{u+1}^{t-1},\mathbf{C}^{(N)},\mathbf{x}^{(u)}\right) = f\left(x_{t}\middle|\mathbf{x}_{u+1}^{t-1},C_{t},\mathbf{x}^{(u)}\right) \tag{B.14}$$

$$= f\left(x_{t}\middle|\mathbf{x}_{t-u}^{t-1},C_{t}\right) \tag{B.15}$$

where Equation (B.14) follows from the independence of X_{t} and $\mathbf{C}^{(t-1)}$ and Equation (B.15) follows from the dependence of X_{t} on $\mathbf{x}_{t-u}^{t-1}$ only. After we replace the first factor of Equation (B.12) by equations (B.13) and the remaining factors of

Equation (B.12) by Equation (B.15), we can then re-write the first term on the right-hand side of Equation (B.9) as:

$$f\left(\mathbf{x}_{u+1}^{N}\middle|\mathbf{C}^{(N)}, \mathbf{x}^{(u)}\right) = f\left(x_{u+1}\middle|C_{u+1}, \mathbf{x}^{(u)}\right)\prod_{t=u+2}^{N} f\left(x_t\middle|\mathbf{x}_{t-u}^{t-1}, C_t\right). \qquad \text{(B.16)}$$

The second term on the right-hand side of Equation (B.9) can be written as:

$$\Pr\left(\mathbf{C}^{(N)}\middle|\mathbf{x}^{(u)}\right) = \Pr\left(C_N, \mathbf{C}^{(N-1)}\middle|\mathbf{x}^{(u)}\right) \qquad \text{(B.17)}$$

$$= \Pr\left(C_N\middle|\mathbf{C}^{(N-1)}, \mathbf{x}^{(u)}\right)\Pr\left(\mathbf{C}^{(N-1)}\middle|\mathbf{x}^{(u)}\right), \qquad \text{(B.18)}$$

where Equation (B.18) follows from applying the definition of conditional probability to Equation (B.17). After we have applied the definition of conditional probability to $\Pr\left(C_t, \mathbf{C}^{(t-1)}\middle|\mathbf{x}^{(u)}\right)$, for $t = N, (N-1), \ldots, 2$, the second term on the right-hand side of Equation (B.9) can be written as a product

$$\Pr\left(\mathbf{C}^{(N)}\middle|\mathbf{x}^{(u)}\right) = \Pr\left(C_1\right)\prod_{t=2}^{N}\Pr\left(C_t\middle|\mathbf{C}^{(t-1)}, \mathbf{x}^{(u)}\right). \qquad \text{(B.19)}$$

For $t = 2, 3, \ldots, N$, the factors in Equation (B.19) can be re-written as

$$\Pr\left(C_t\middle|\mathbf{C}^{(t-1)}, \mathbf{x}^{(u)}\right) = \Pr\left(C_t\middle|\mathbf{C}^{(t-1)}\right) \qquad \text{(B.20)}$$

$$= \Pr\left(C_t\middle|C_{t-1}\right), \qquad \text{(B.21)}$$

where Equation (B.20) follows from the independence of C_t and $\mathbf{x}^{(u)}$ and Equation (B.21) follows from the Markov property of $\{C_t\}$. After we replace the factors of Equation (B.19) by Equation (B.21), we can then re-write the second term on the right-hand side of Equation (B.9) as:

$$\Pr\left(\mathbf{C}^{(N)}\middle|\mathbf{x}^{(u)}\right) = \Pr\left(C_1\right)\prod_{t=2}^{N}\Pr\left(C_t\middle|C_{t-1}\right). \qquad \text{(B.22)}$$

Finally, by replacing the first term of Equation (B.9) by Equation (B.16) and the second term of Equation (B.9) by Equation (B.22), the summands can be re-written as

$$
f\left(\mathbf{x}_{u+1}^{N}, \mathbf{C}^{(N)}\middle|\mathbf{x}^{(u)}\right)
$$

$$
= \Pr\left(C_1\right)\prod_{t=2}^{N}\Pr\left(C_t\middle|C_{t-1}\right)
$$

$$
\times f\left(x_{u+1}\middle|C_{u+1}, \mathbf{x}^{(u)}\right)\prod_{t=u+2}^{N}f\left(x_t\middle|\mathbf{x}_{t-u}^{t-1}, C_t\right) \tag{B.23}
$$

and the likelihood of a Markov switching process conditioned on $x_1, \ldots, x_u$ can be written as

$$
L_N = \sum_{j_1,\ldots,j_N=1}^{m}\left(\delta_{j_1}\prod_{t=2}^{N}\gamma_{j_{t-1},j_t}\right)
$$

$$
\times\left(f_{j_{u+1}}\left(x_{u+1}\middle|\mathbf{x}^{(u)}\right)\prod_{t=u+2}^{N}f_{j_t}\left(x_t\middle|\mathbf{x}_{t-u}^{t-1}\right)\right) \tag{B.24}
$$

$$
= \sum_{j_1,\ldots,j_N=1}^{m}\left(\delta_{j_1}\prod_{t=2}^{u+1}\gamma_{j_{t-1},j_t}\right)f_{j_{u+1}}\left(x_{u+1}\middle|\mathbf{x}^{(u)}\right)
$$

$$
\times\prod_{t=u+2}^{N}\gamma_{j_{t-1},t}f_{j_t}\left(x_t\middle|\mathbf{x}_{t-u}^{t-1}\right) \tag{B.25}
$$

$$
= \boldsymbol{\delta}\boldsymbol{\Gamma}^u\mathbf{F}\left(x_{u+1}\middle|\mathbf{x}^{(u)}\right)
$$

$$
\times\boldsymbol{\Gamma}\mathbf{F}\left(x_{u+2}\middle|\mathbf{x}_2^{u+1}\right)\boldsymbol{\Gamma}\mathbf{F}\left(x_{u+3}\middle|\mathbf{x}_3^{u+2}\right)\cdot\cdots\cdot\boldsymbol{\Gamma}\mathbf{F}\left(x_N\middle|\mathbf{x}_{N-u}^{N-1}\right)\mathbf{1}', \tag{B.26}
$$

where $\boldsymbol{\delta}$ is a row vector with elements $\delta_{j_1} = \Pr\left(C_1 = j_1\right)$, $\boldsymbol{\Gamma}$ is the t.m.p. of $\{C_t\}$ with elements $\gamma_{j_{t-1},j_t} = \Pr\left(C_t = j_t\middle|C_{t-1} = j_{t-1}\right)$, $\mathbf{F}\left(x_t\middle|\mathbf{x}_{t-u}^{t-1}\right)$ is the diagonal matrix with j_t th element $f_{j_{u+1}}\left(x_{u+1}\middle|\mathbf{x}^{(u)}\right) = f\left(x_t\middle|\mathbf{x}_{t-u}^{t-1}, C_t = j_t\right)$ and $\mathbf{1}$ is a row vector of length m with one as its elements.

Software for estimation

Consider the two-state MSACD model with t.p.m.

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}$$

and conditional means in state 1 given by

$$\mu_{1,1} = \frac{\omega_1}{1-\alpha_1-\beta_1}$$

$$\mu_{1,1} = \frac{\omega_1}{1-\alpha_1-\beta_1} \mu_{n,1} = \omega_1 + \alpha_1 Z_{n-1} + \beta_1 \mu_{n-1,1} \quad \text{(for } n = 2, \ldots, N)$$

and in state 2 given by

$$\mu_{1,2} = \frac{\omega_2}{1-\alpha_2-\beta_2}$$

$$\mu_{1,2} = \frac{\omega_2}{1-\alpha_2-\beta_2} \mu_{n,2} = \omega_2 + \alpha_2 Z_{n-1} + \beta_2 \mu_{n-1,2} \quad \text{(for } n = 2, \ldots, N).$$

At Z_n the distribution of the error is a mixture distribution with components

$$\epsilon_{n,1} \sim Burr\left(\mu = 1, \kappa_1, \sigma_1^2\right) \quad \text{with} \quad E\left(\epsilon_n | C_t = 1\right) = 1$$

$$\epsilon_{n,2} \sim Burr\left(\mu = 1, \kappa_2, \sigma_2^2\right) \quad \text{with} \quad E\left(\epsilon_n | C_t = 2\right) = 1,$$

i.e. the component distribution selected by the Markov chain at Z_n is used to describe the error. We call a model of the kind we describe above as the Burr MSACD$(1,1)$. It has twelve parameters of which two are for the Markov chain, γ_{12}, γ_{21}, six are for the state-dependent conditional means, $\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2$, and the remaining four are for the state-dependent error distributions, $\kappa_1, \sigma_1^2, \kappa_2, \sigma_2^2$.

We checked our estimation software by using simulation to estimate the means and standard deviations of the maximum likelihood estimators in the two-state

model we describe above. We generated twenty series of length 10000^5 from this model by specifying its population parameters, and then we fitted the same type of model (using our software) to each of these series[6]. Table 4.2 gives the sample means, biases[7] and standard deviations for the estimators of the twelve parameters. We note in this table the following. Firstly, sample means are reasonably similar to the population parameters. Secondly, although the standard deviations are not large, it has to be kept in mind that very long series were used in the simulation. The results from this simulation exercise provide some support for the accuracy of our estimation software.

---

[5] This round number is in the range of the lengths of the series that we fit our models to. The series we generated has a length of 11000 of which the first 1000 observations are regarded as the burn-in sample.

[6] The computational time for the simulation is roughly eight hours.

[7] Defined here as the difference between the values of sample mean and population parameter.

$\boldsymbol{\theta}$	true	mean$\left(\hat{\boldsymbol{\theta}}\right)$	bias$\left(\hat{\boldsymbol{\theta}}\right)$	s.d.$\left(\hat{\boldsymbol{\theta}}\right)$
ω_1	2.00	1.9934	-0.0066	0.1328
α_1	0.15	0.1489	-0.0011	0.0074
β_1	0.80	0.7982	-0.0018	0.0101
κ_1	3.50	3.4590	-0.0410	0.0860
σ_1^2	2.00	1.9479	-0.0521	0.0700
ω_2	0.50	0.5023	0.0023	0.0377
α_2	0.10	0.0995	-0.0005	0.0113
β_2	0.50	0.4992	-0.0008	0.0289
κ_2	1.50	1.5119	0.0119	0.0273
σ_2^2	0.50	0.5068	0.0068	0.0588
γ_{12}	0.10	0.0997	-0.0003	0.0044
γ_{21}	0.10	0.1000	0.0000	0.0046

Table 4.2: Simulation of a two-state Burr MSACD(1,1) model: bias and standard error of the parameters.

Figures
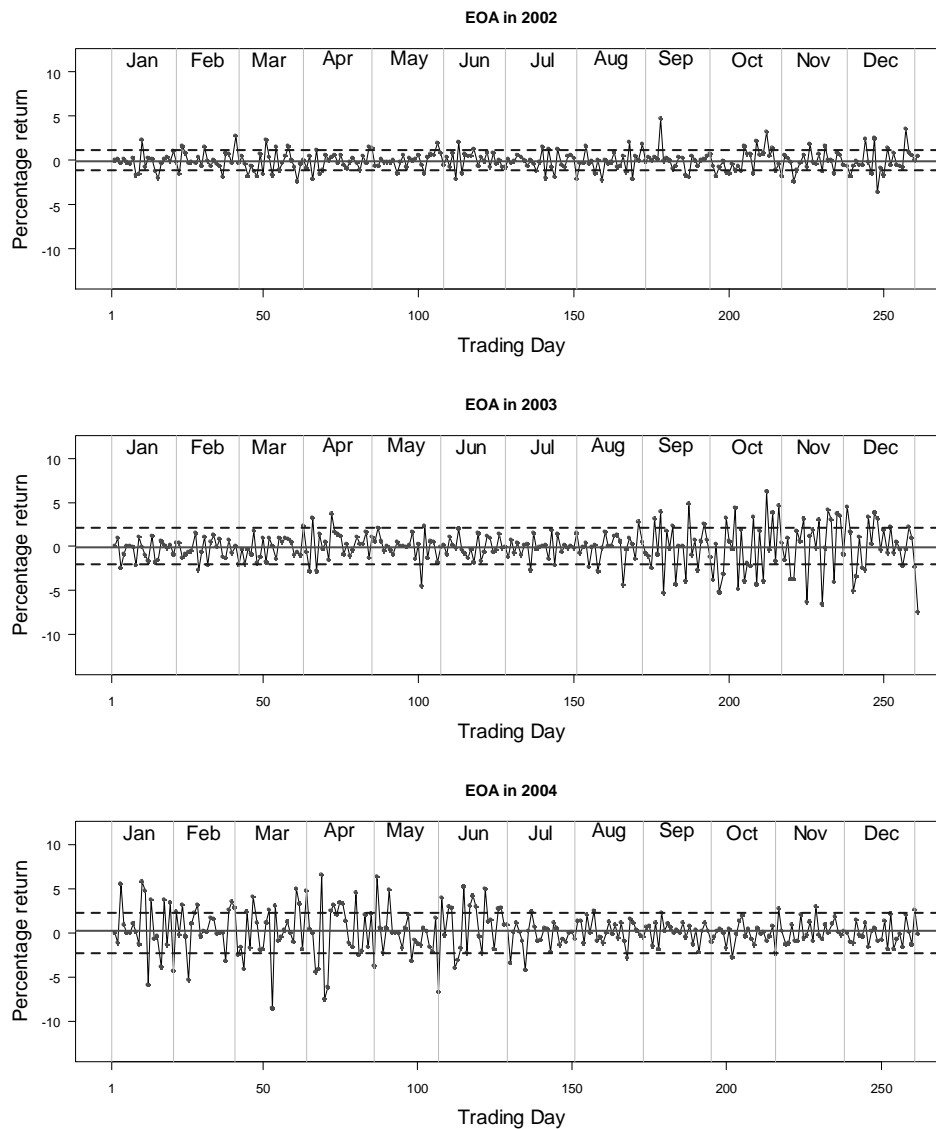
Figure 4.1: Index plot of daily return series of E.ON. The return is calculated on the opening price of the share on each trading day. Solid horizontal line marks the average return over the trading year. Dashed line marks three standard deviations in excess (above mean) and in deficit (below mean) of the mean.
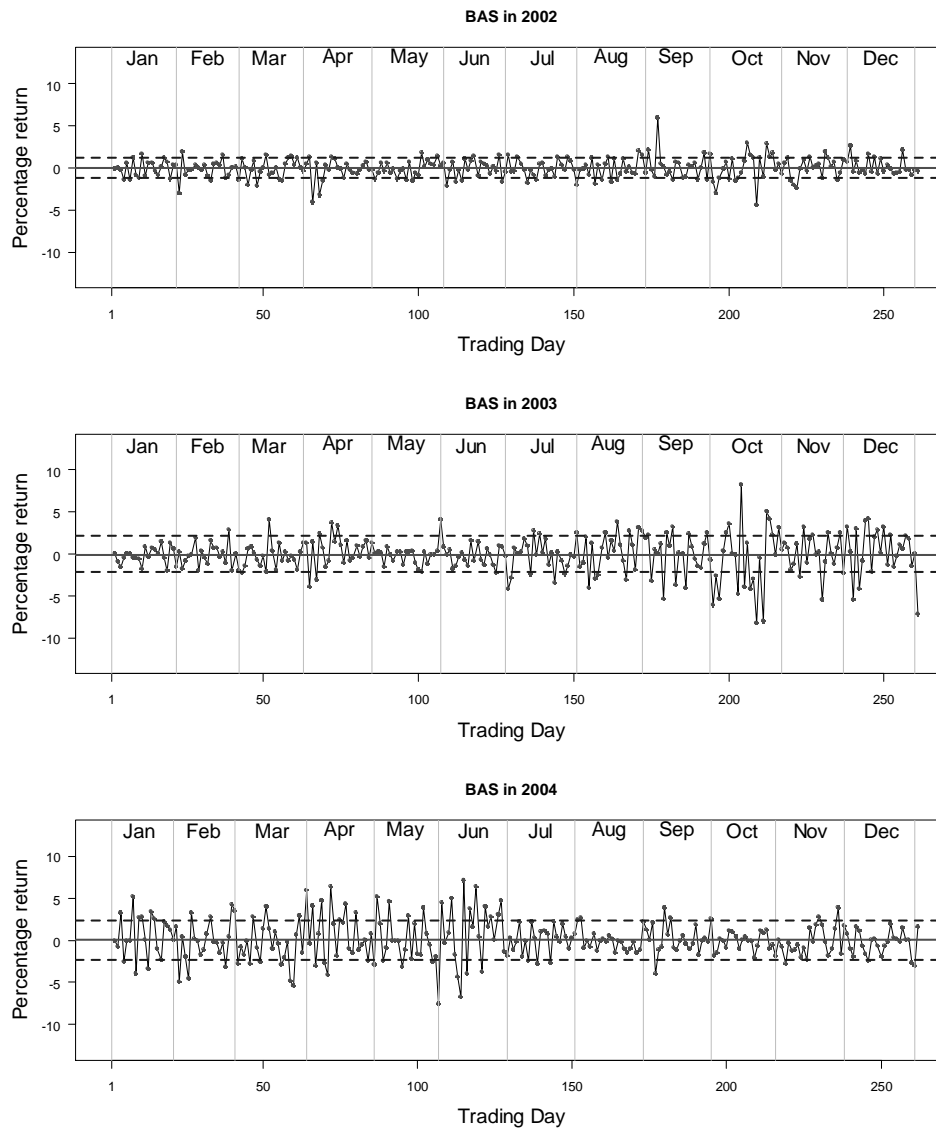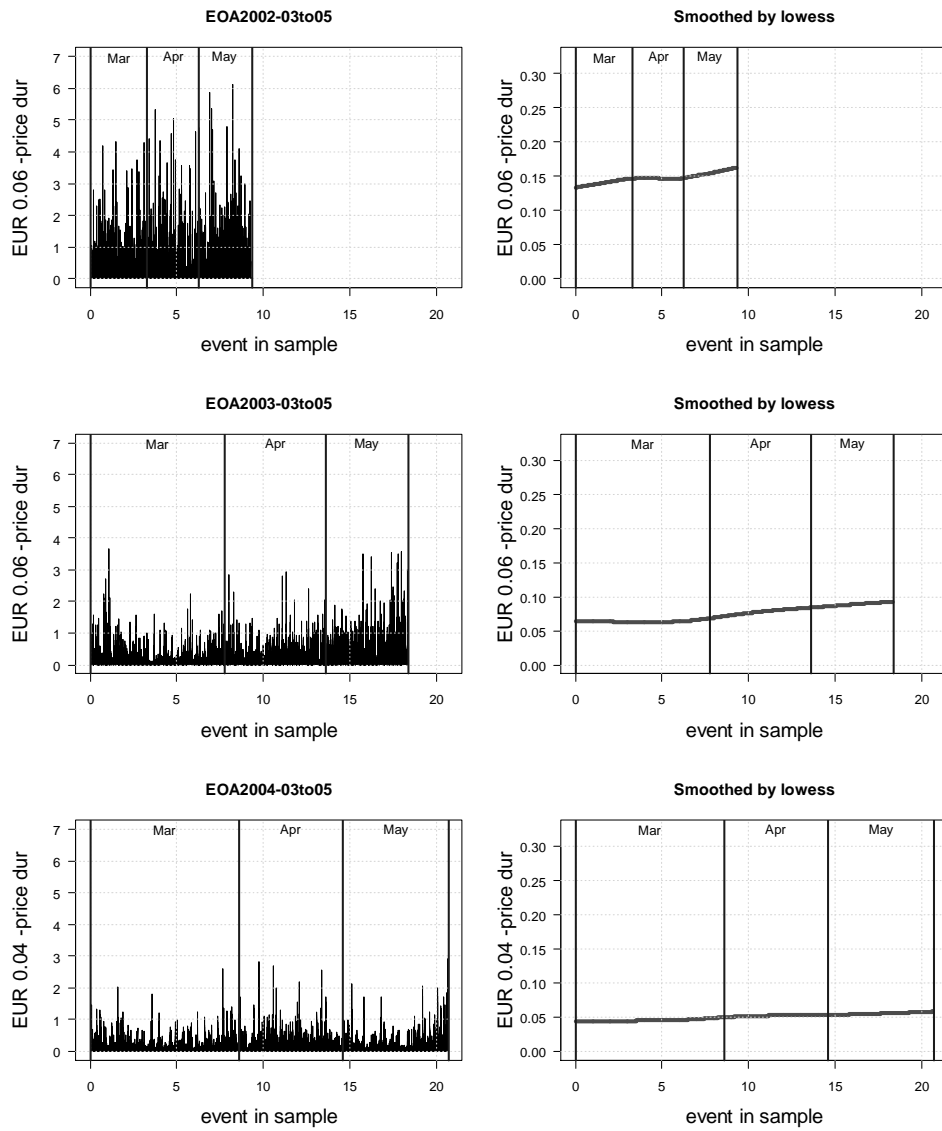
Figure 4.2: Index plot of daily return series of BASF. The return is calculated on the opening price of the share on each trading day. Solid horizontal line marks the average return over the trading year. Dashed line marks three standard deviations in excess (above mean) and in deficit (below mean) of the mean.

Figure 4.3: Plots of sm-M price durations (in units of 1000 seconds) for EOA against index of price event (in units of 1000) for the months March to May in the years 2002-2004 (left column). Nonparametric regression lines computed by the **R** function `lowess` that is a smooth of the points in each of the years (right column).
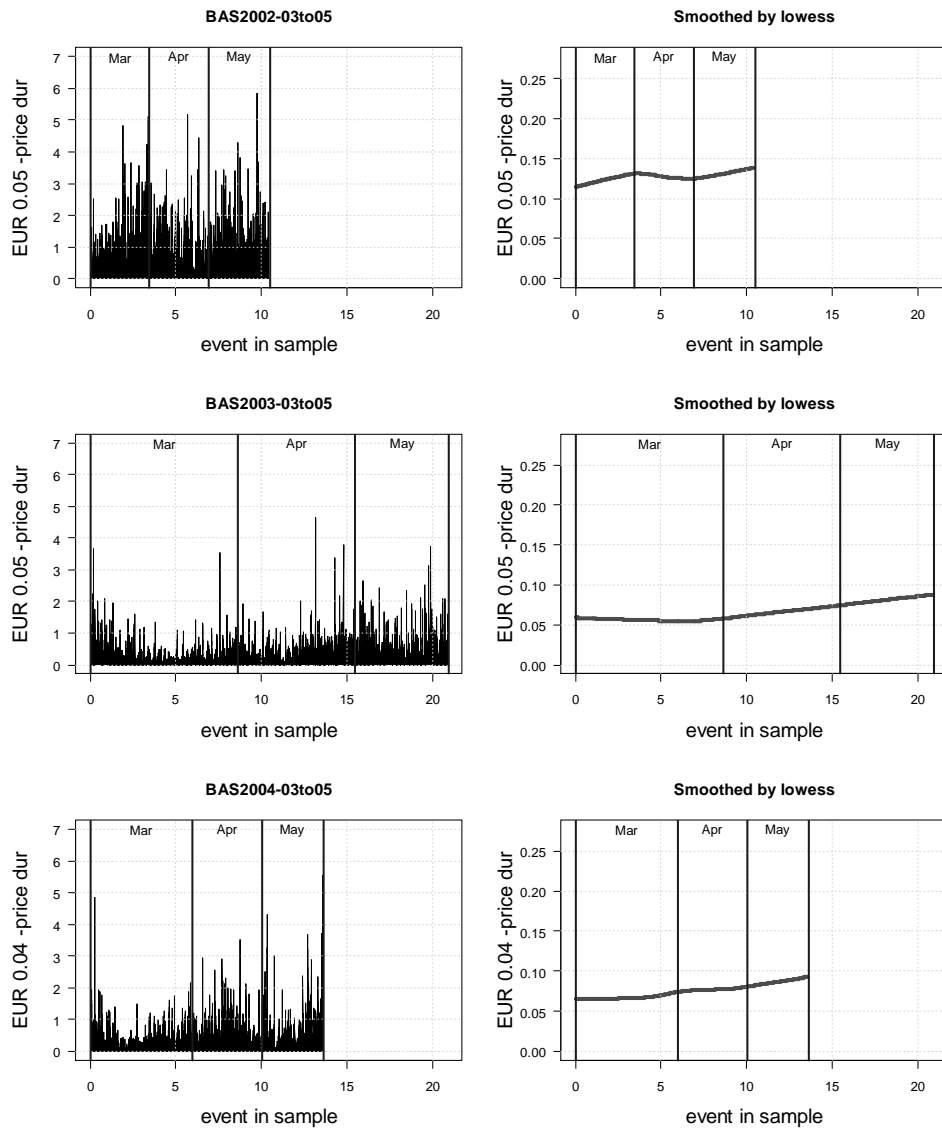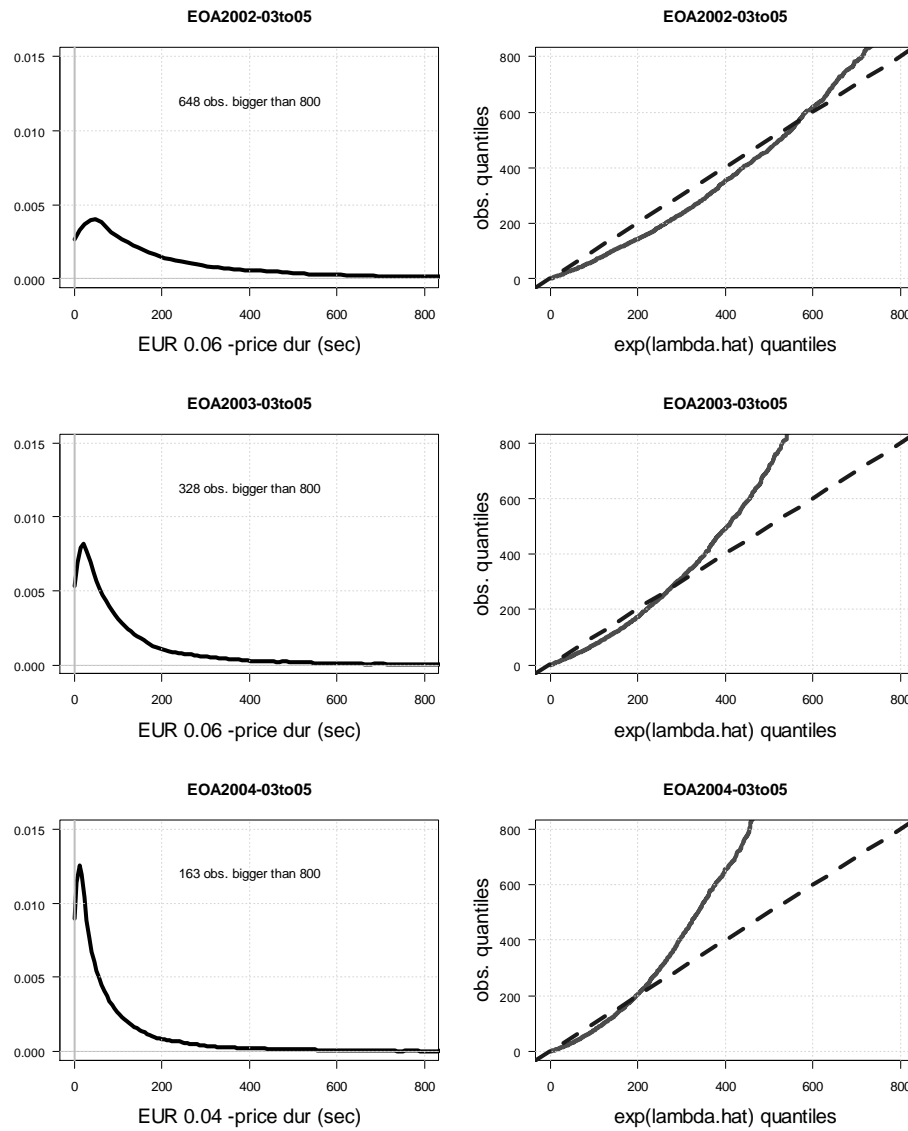
Figure 4.4: Plots of sm-M price durations (in units of 1000 seconds) for BAS against index of price event (in units of 1000) for the months March to May in the years 2002-2004 (left column). Nonparametric regression lines computed by the **R** function `lowess` that is a smooth of the points in each of the years (right column).
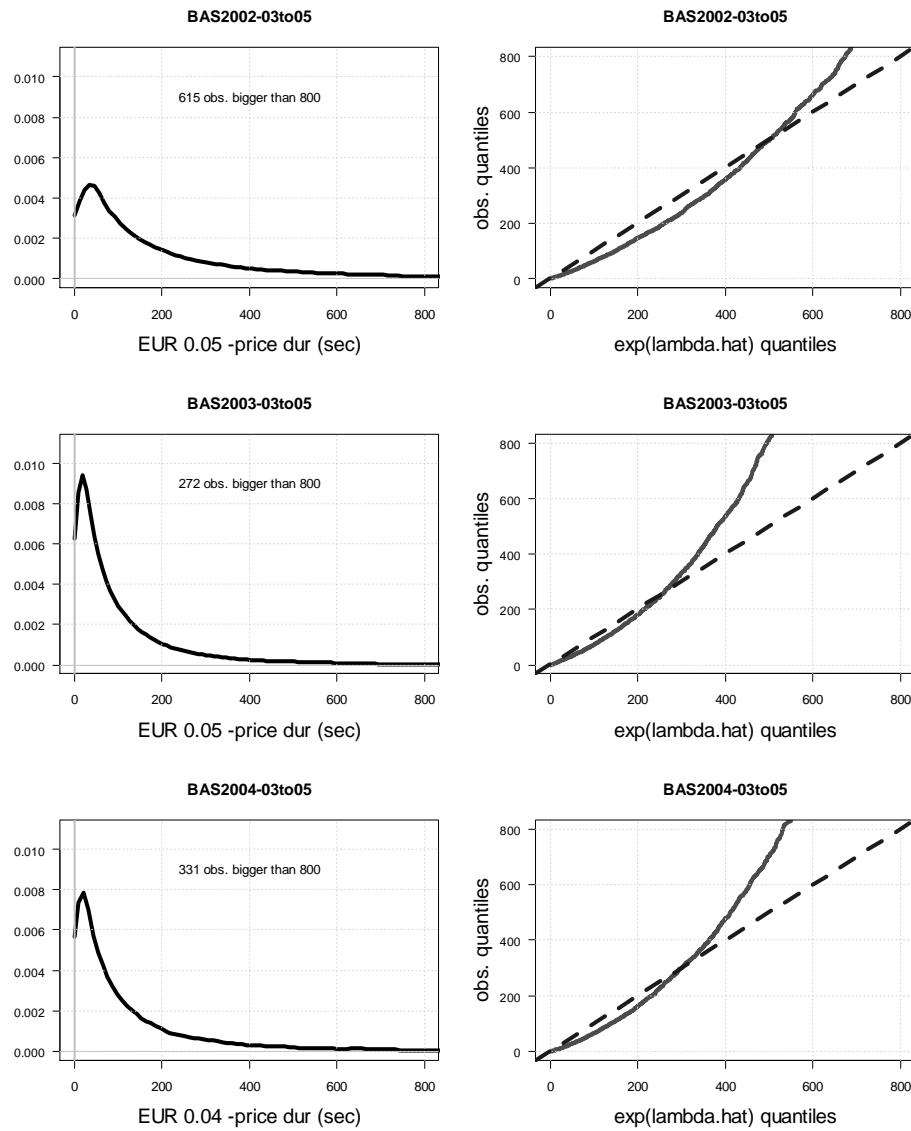
Figure 4.5: Kernel density estimates of the observed sm-M price durations of E.ON (left column). qq-plot of the observed price durations based on an exponential distribution (right column).

Figure 4.6: Kernel density estimates of the observed sm-M price durations of BASF (left column). qq-plot of the observed price durations based on an exponential distribution (right column).
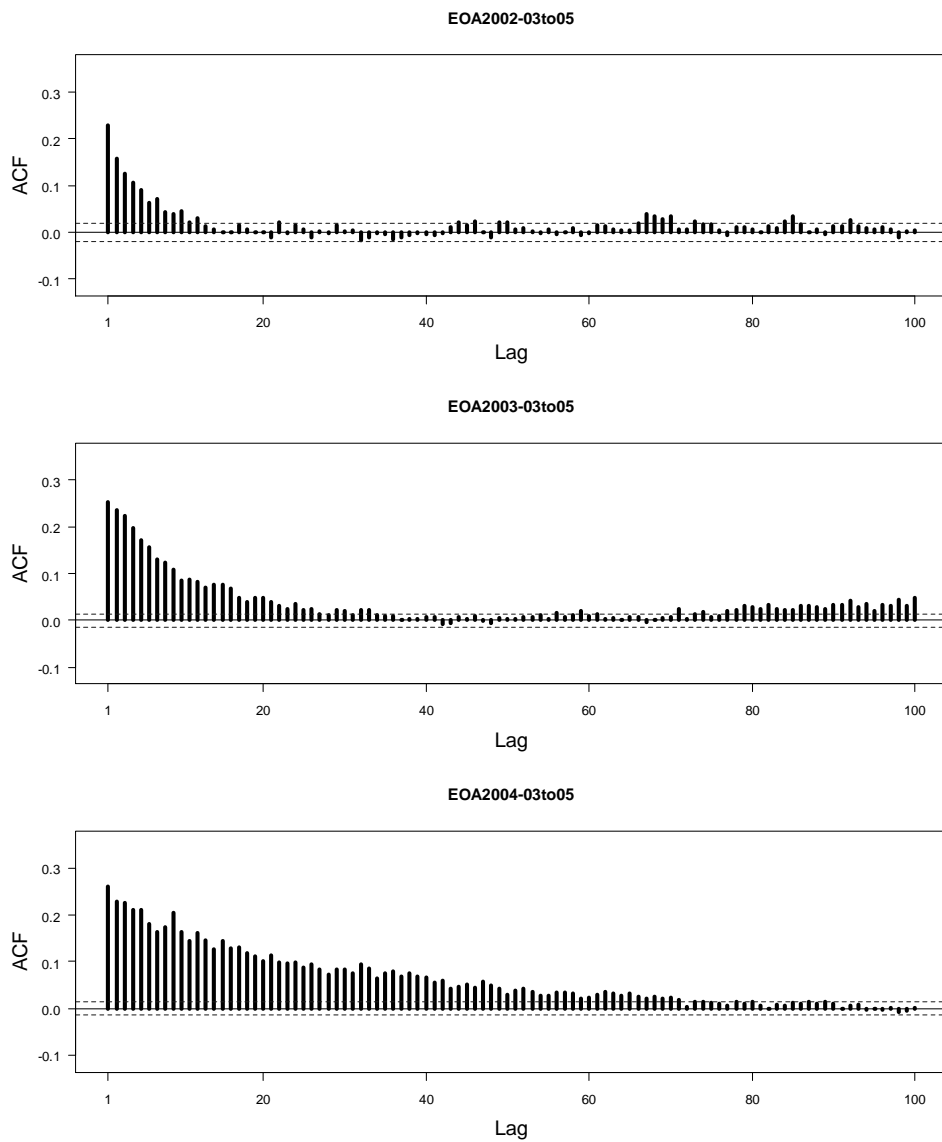
Figure 4.7: ACF of observed price durations of E.ON corresponding to small-M price changes for the period March to May in years 2002, 2003 and 2004.
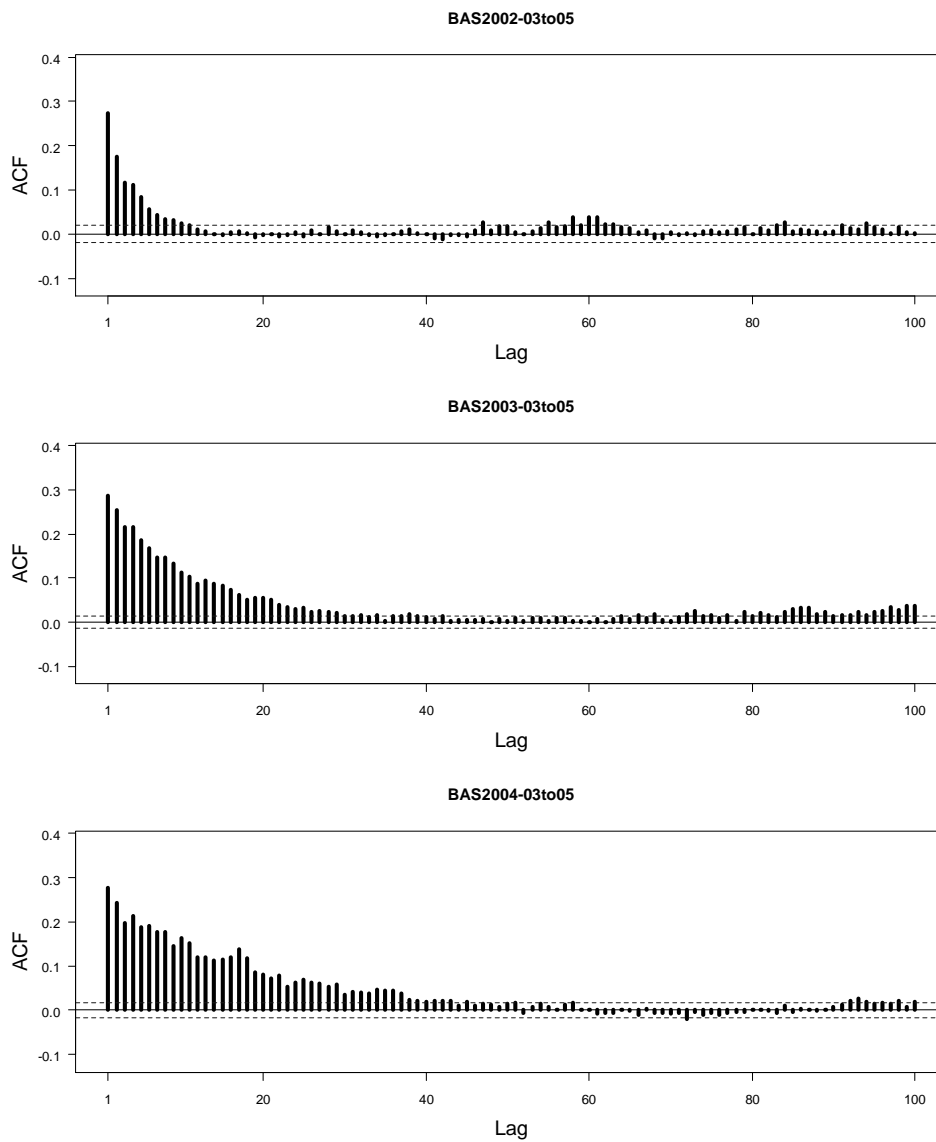
Figure 4.8: ACF of observed price durations of BASF corresponding to small-M price changes for the period March to May in years 2002, 2003 and 2004.
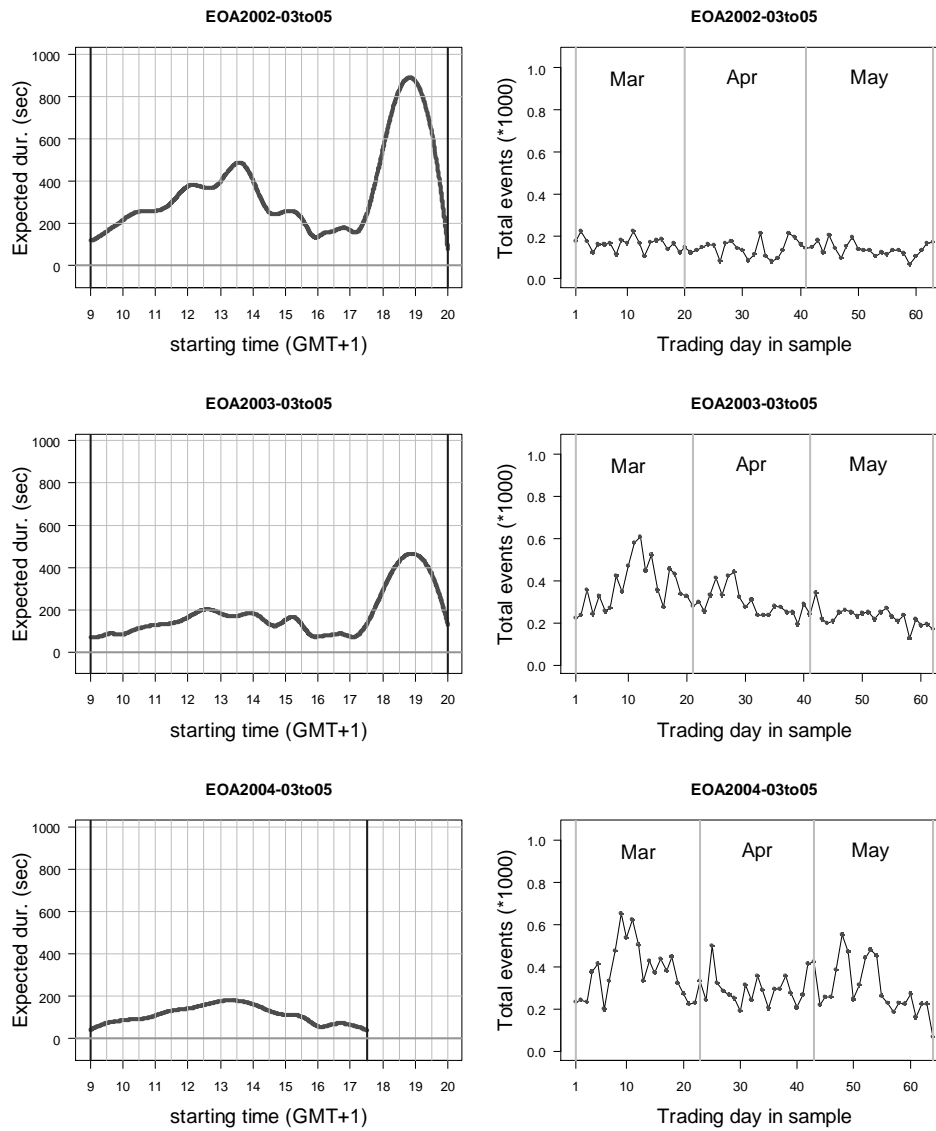
Figure 4.9: Observed price durations of E.ON: nonparametric estimate of intraday cyclical pattern (left column) and daily average (right column). The durations correspond to small-M price changes for the period March to May in years 2002, 2003 and 2004.
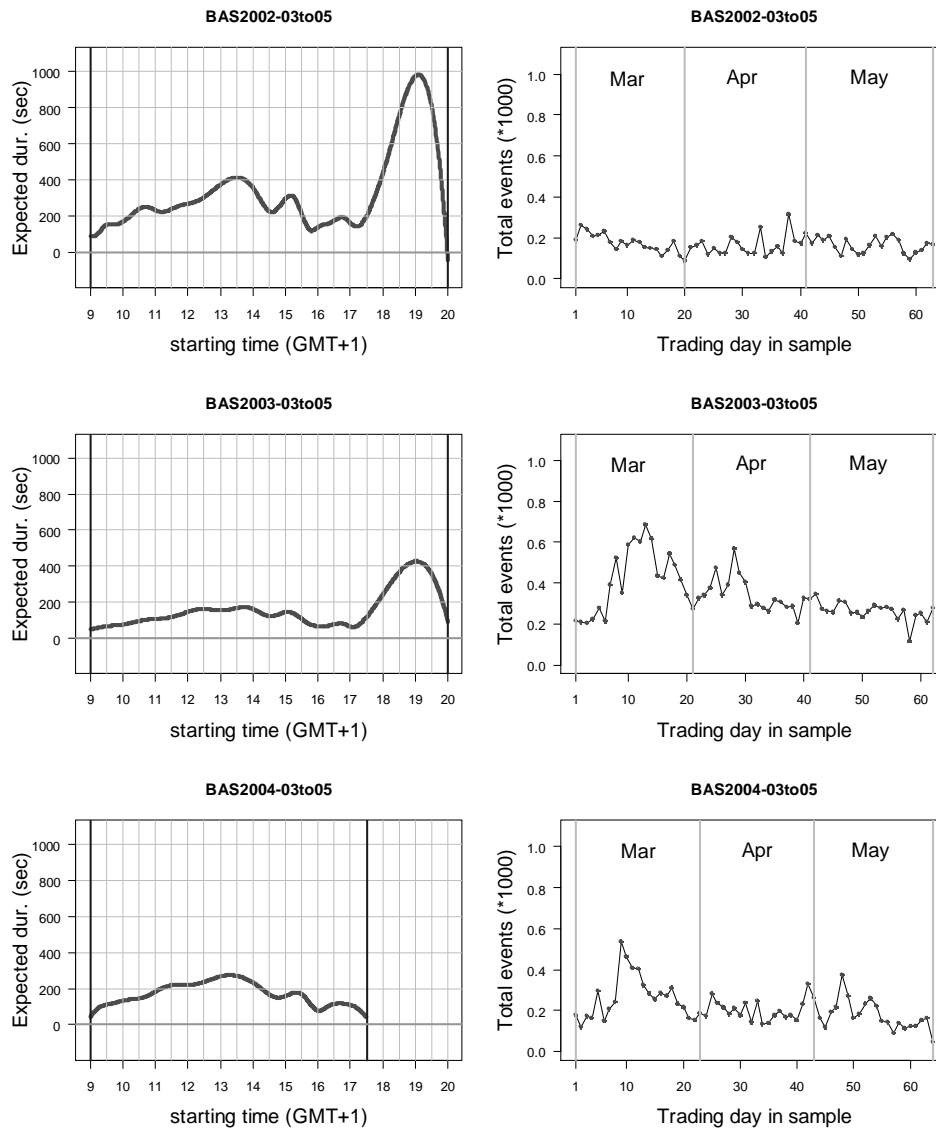
Figure 4.10: Observed price durations of BASF: nonparametric estimate of intra-day cyclical pattern (left column) and daily average (right column). The durations correspond to small-M price changes for the period March to May in years 2002, 2003 and 2004.
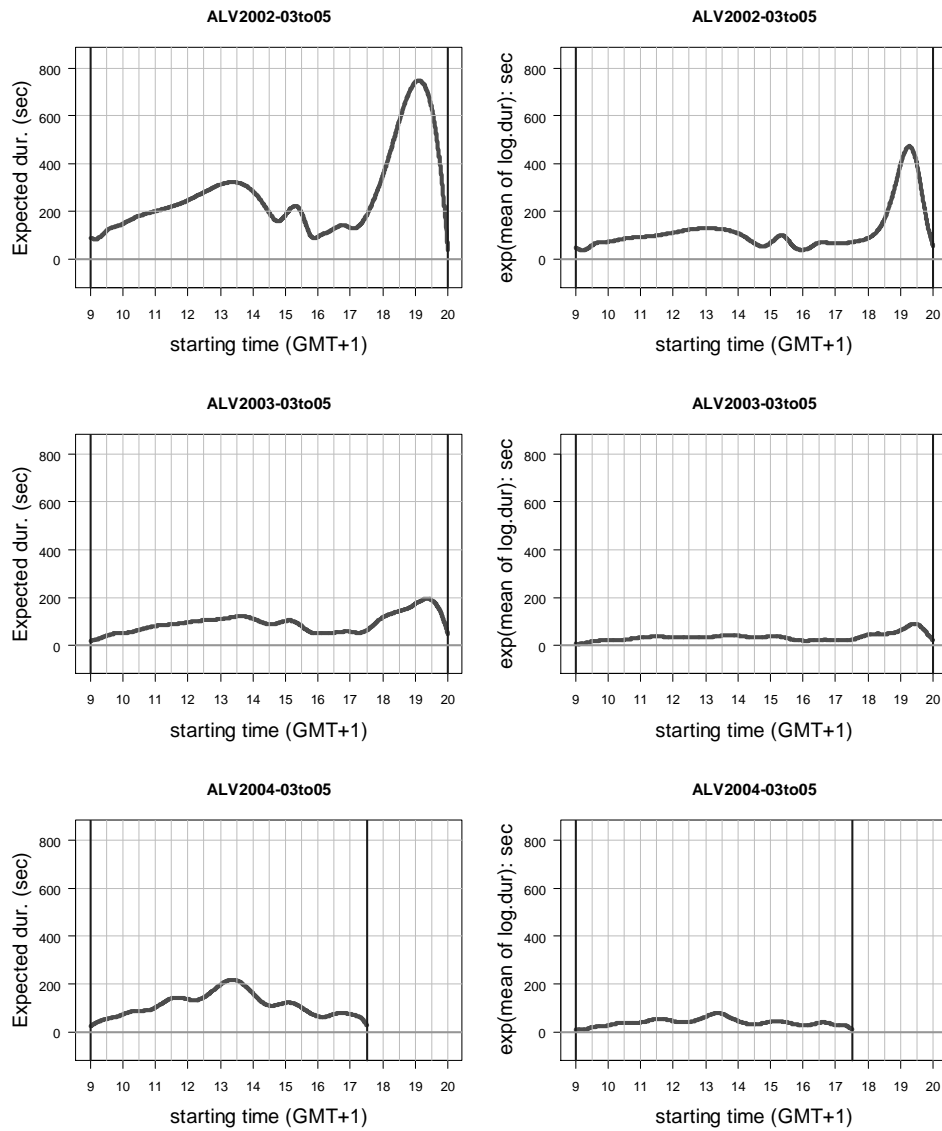
Figure 4.11: Observed price durations of Allianz: nonparametric estimate of mean duration, before transformations (left column) and after transformations (right column), as a function of time of last significant price change. The durations correspond to small-M price changes for the period March to May in years 2002, 2003 and 2004.
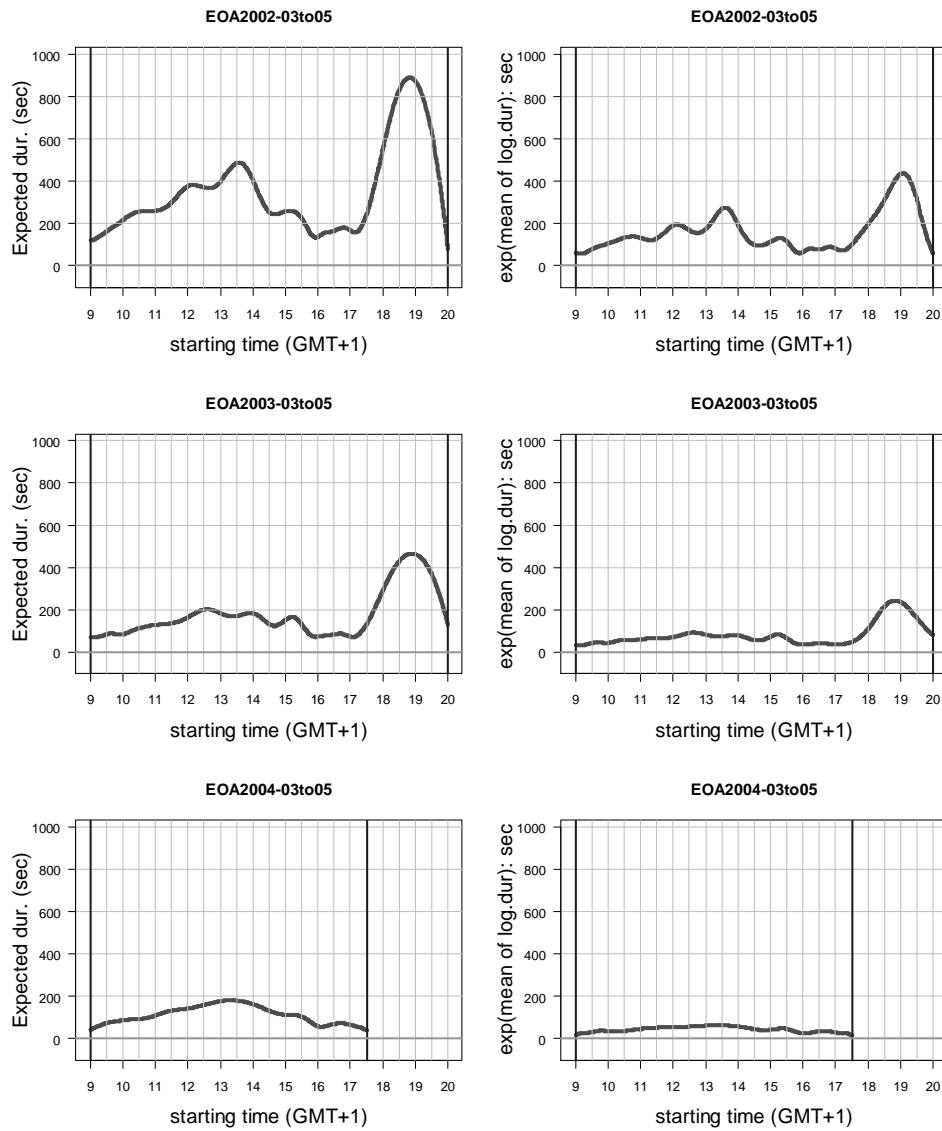
Figure 4.12: Observed price durations of E.ON: nonparametric estimate of mean duration, before transformations (left column) and after transformations (right column), as a function of time of last significant price change. The durations correspond to small-M price changes for the period March to May in years 2002, 2003 and 2004.
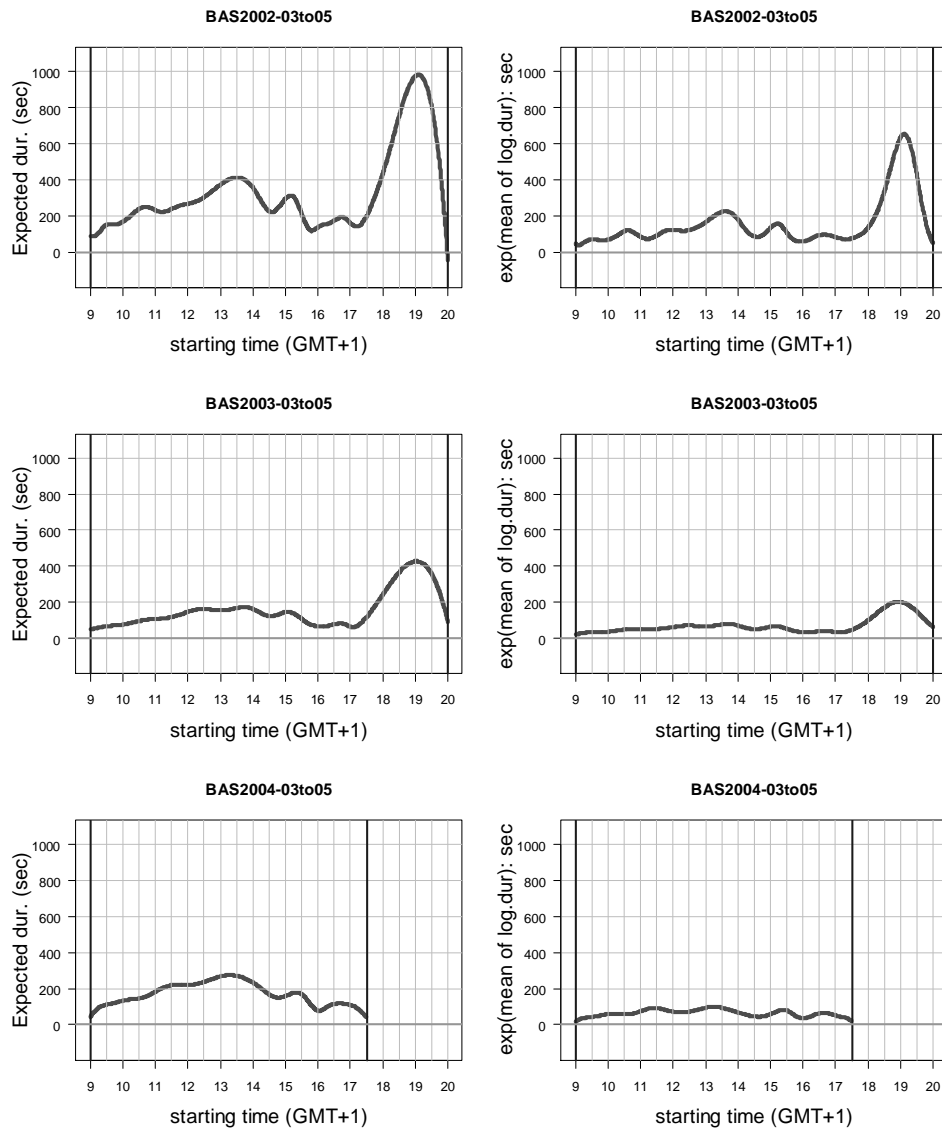
Figure 4.13: Observed price durations of BASF: nonparametric estimate of mean duration, before transformations (left column) and after transformations (right column), as a function of time of last significant price change. The durations correspond to small-M price changes for the period March to May in years 2002, 2003 and 2004.
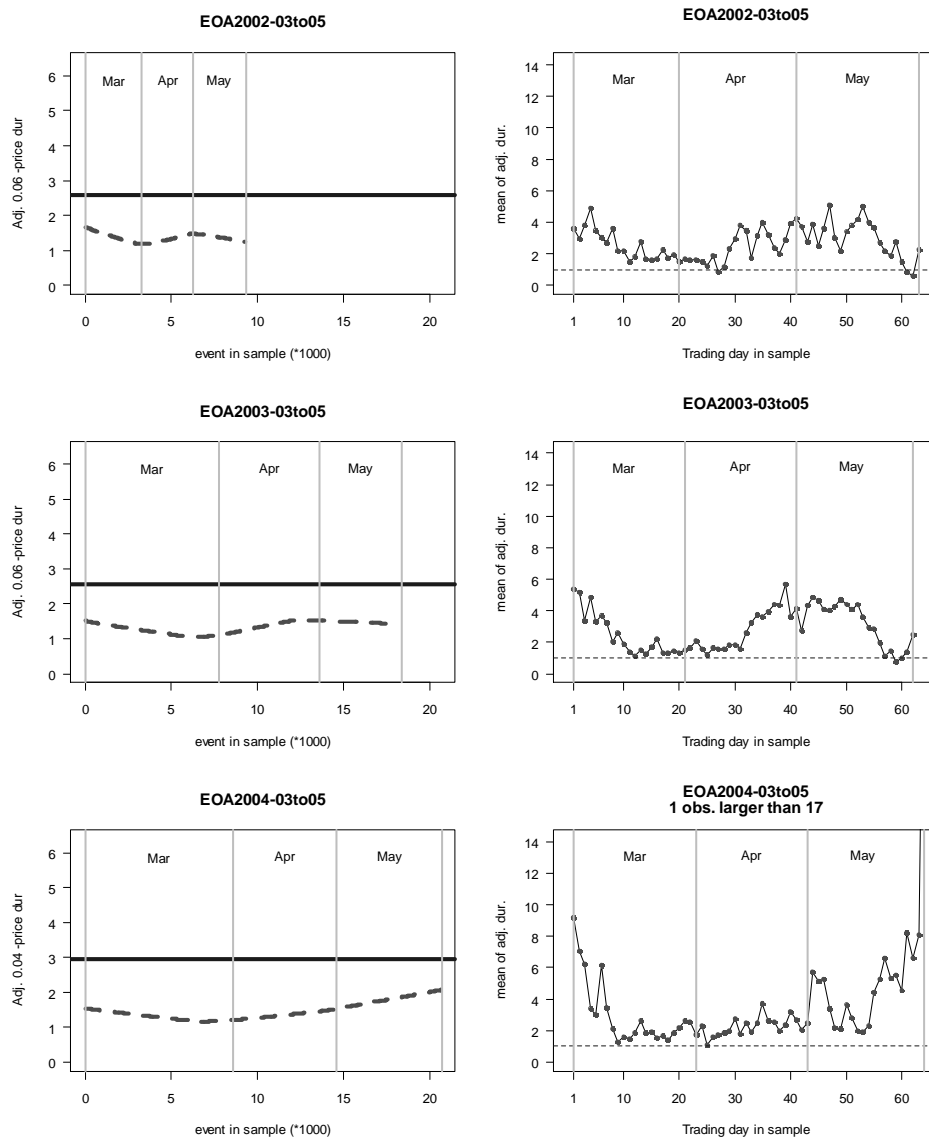
Figure 4.14: (Seasonally-) adjusted price durations of E.ON: mean durations, estimated nonparametrically, as a function of number of significant price changes (dotted line, left column) and daily average (right column). The durations correspond to small-M price changes for the period March to May in years 2002, 2003 and 2004. In each plot of left column, sample mean of durations in each year is indicated by the solid horizontal line.

Figure 4.15: (Seasonally-) adjusted price durations of BASF: mean durations, estimated nonparametrically, as a function of number of significant price changes (dotted line, left column) and daily average (right column). The durations correspond to small-M price changes for the period March to May in years 2002, 2003 and 2004. In each plot of left column, sample mean of durations in each year is indicated by the solid horizontal line.
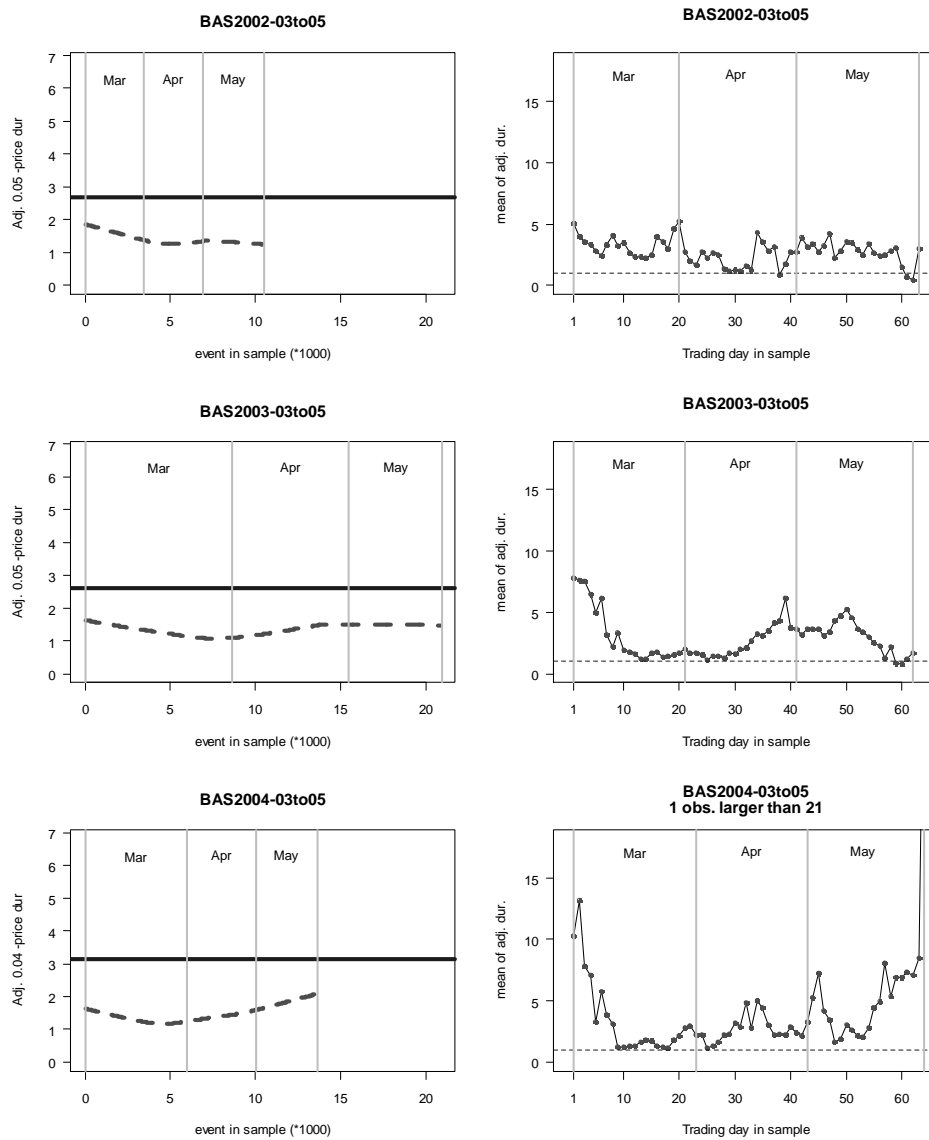
Figure 4.16: ACF of (seasonally-) adjusted price durations of E.ON corresponding to small-M price changes for the period March to May in years 2002, 2003 and 2004.
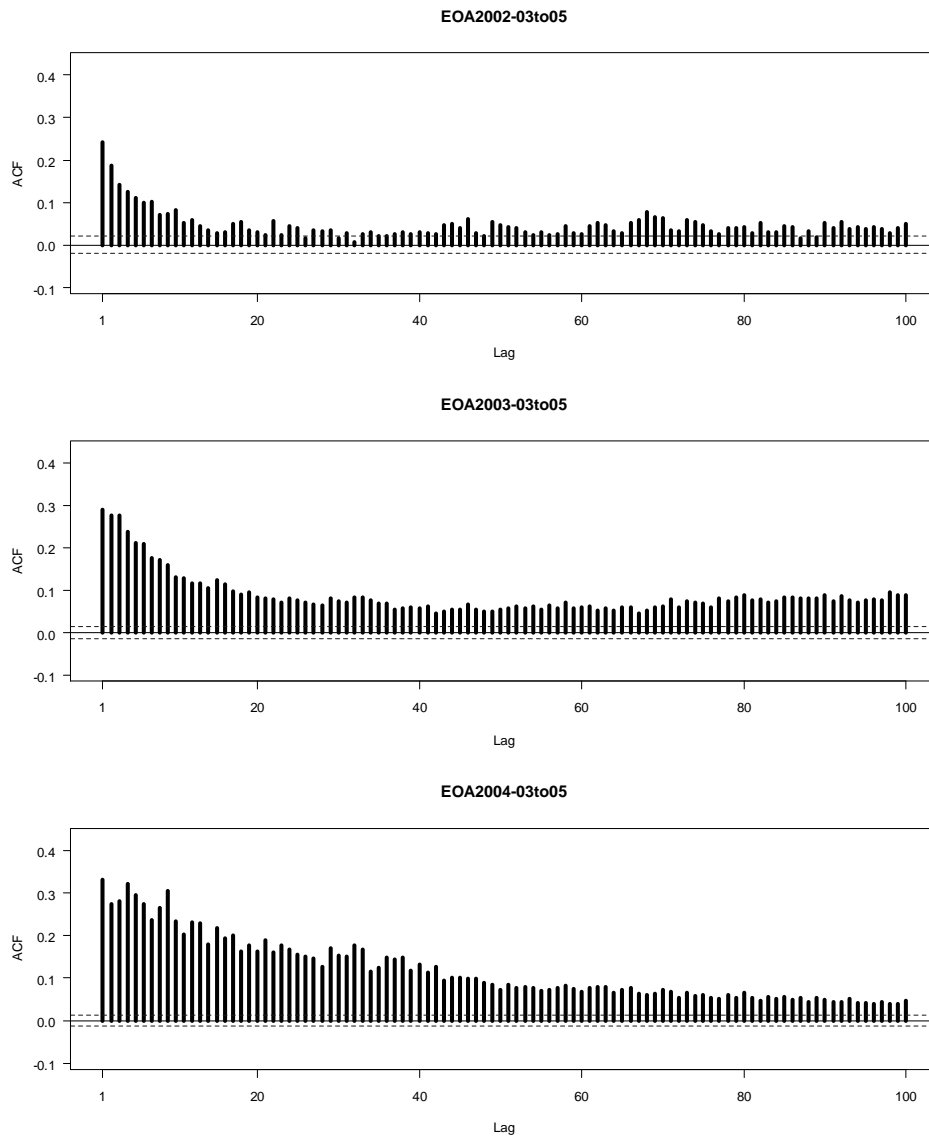
Figure 4.17: ACF of (seasonally-) adjusted price durations of BASF corresponding to small-M price changes for the period March to May in years 2002, 2003 and 2004.

EOA2002-03to05

EOA2003-03to05

EOA2004-03to05

Figure 4.18: ACF of standardised price durations of E.ON corresponding to small-M price changes for the period March to May in years 2002, 2003 and 2004.
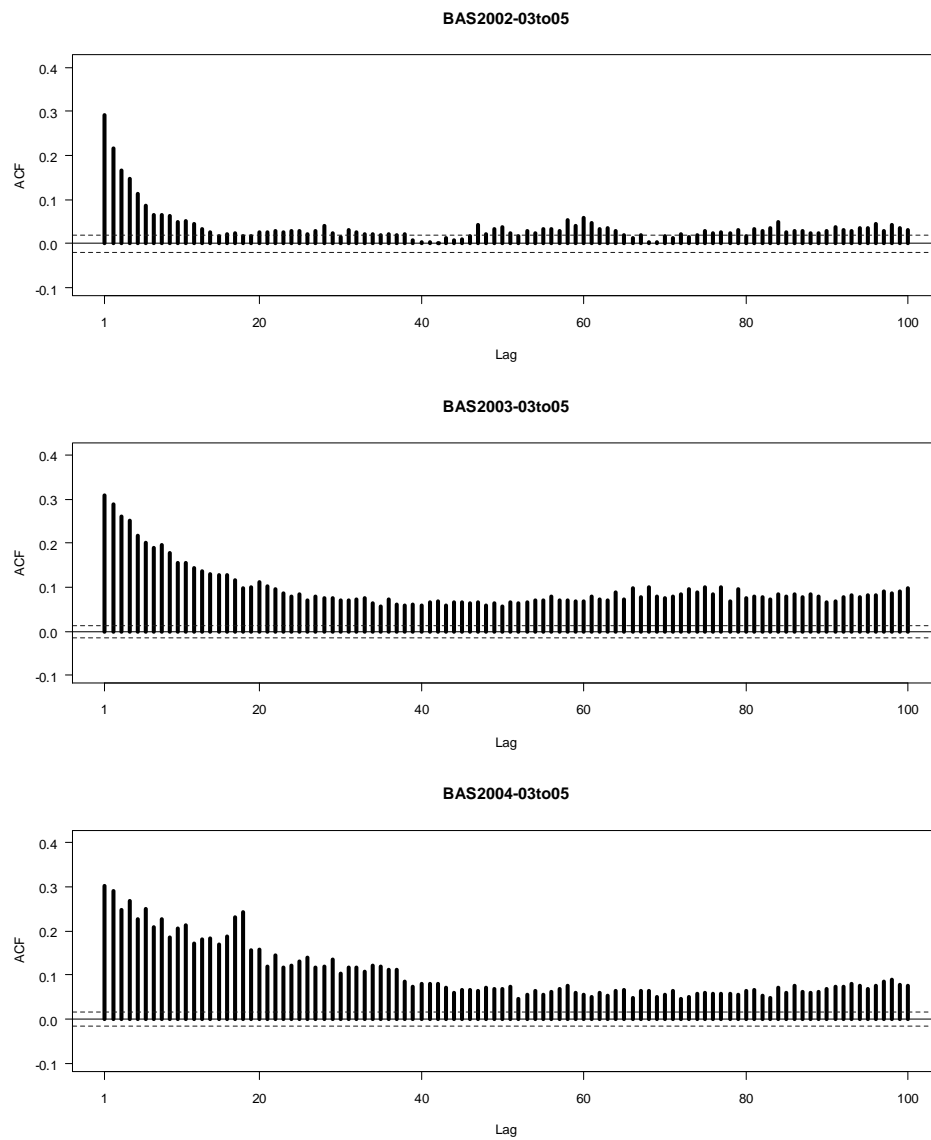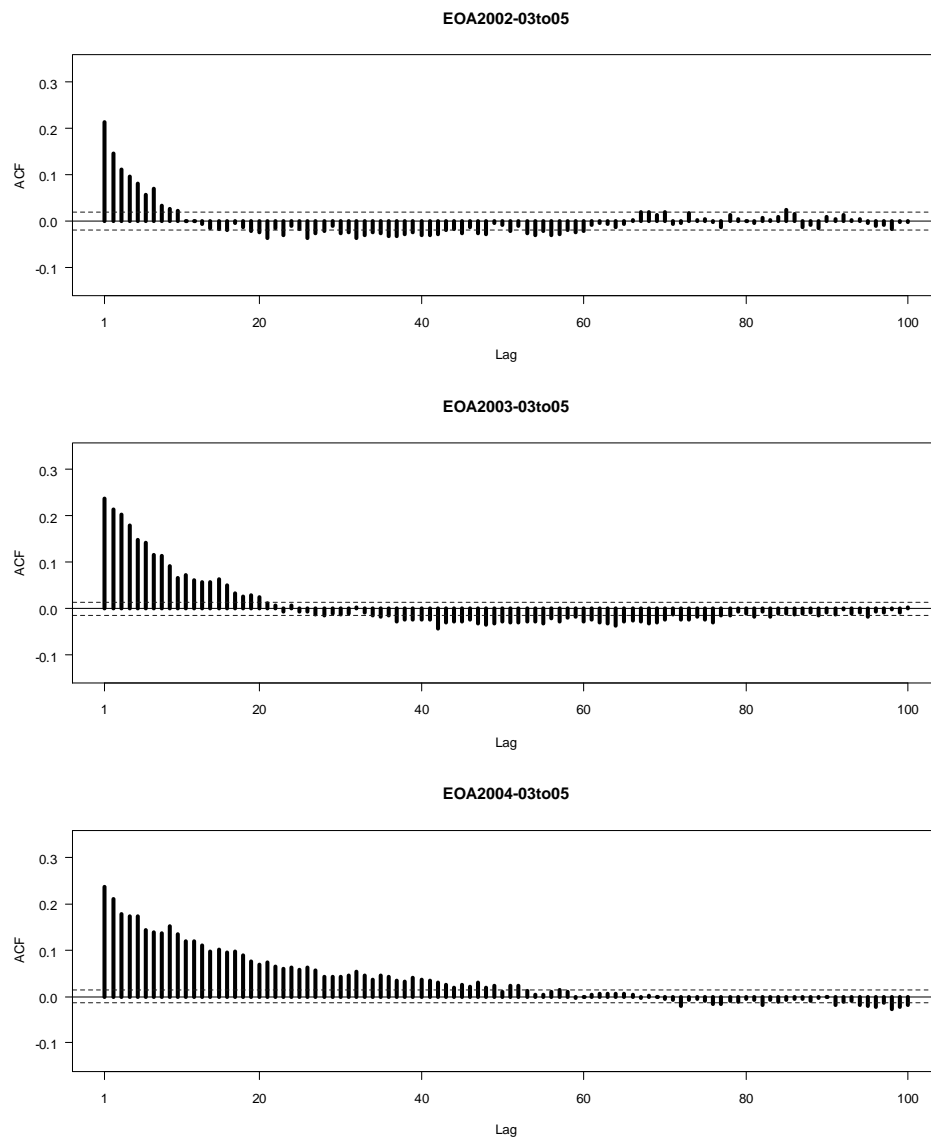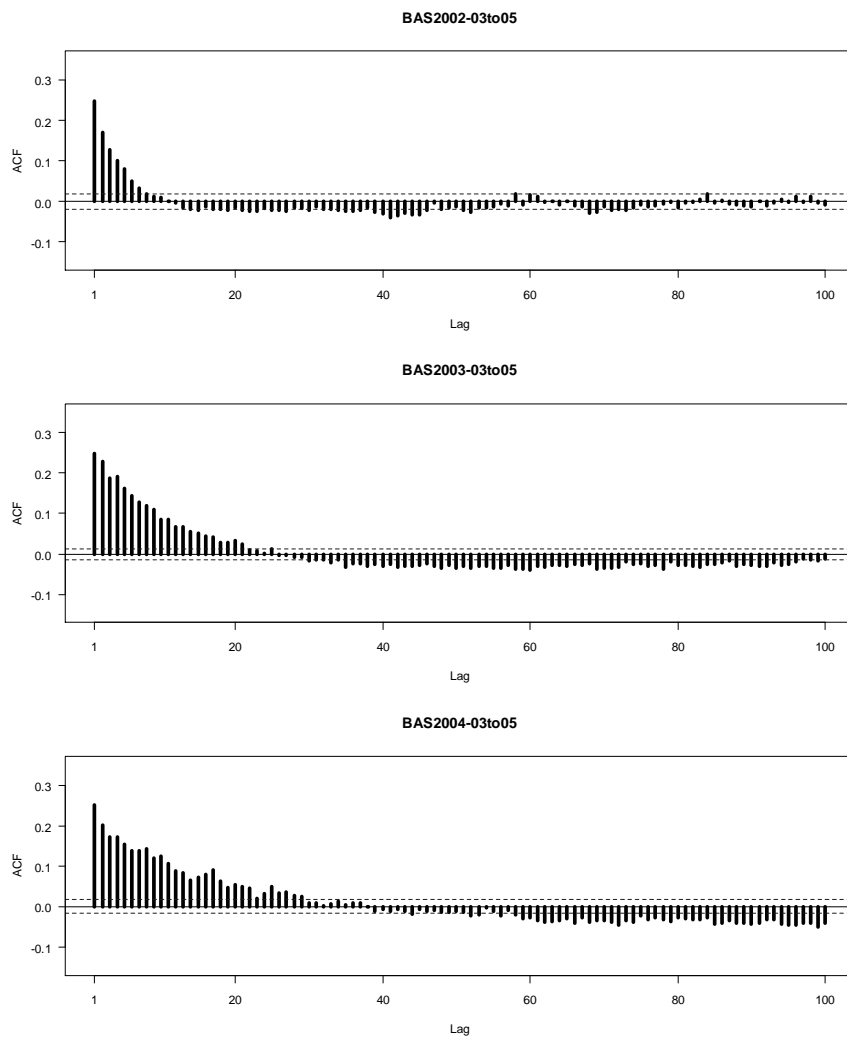
Figure 4.19: ACF of standardised price durations of BASF corresponding to small-M price changes for the period March to May in years 2002, 2003 and 2004.

C.   Hidden Markov Models

two-state HMM

Distribution in state i	*Probability functions*
Poisson	$p_1(s_t) = \frac{\lambda_1^{s_t} \exp(-\lambda_1 s_t)}{(s_t)!}$ $p_2(s_t) = \frac{\lambda_2^{s_t} \exp(-\lambda_2 s_t)}{(s_t)!}$
Bernoulli	$p_1(s_t) = \pi_1^{s_t} (1 - \pi_1)^{1-s_t}$ $p_2(s_t) = \pi_2^{s_t} (1 - \pi_2)^{1-s_t}$

two-state HMM incorporating volume

Distribution in state i	*Probability functions*	*Models for parameters*
Bernoulli	$p_{t,1}(s_t) = \pi_{t,1}^{s_t} \left(1 - \pi_{t,1}^{s_t}\right)^{1-s_t}$ $p_{t,2}(s_t) = \pi_{t,2}^{s_t} \left(1 - \pi_{t,2}^{s_t}\right)^{1-s_t}$	$\pi_{t,1}^{s_t} = \frac{\exp(\alpha_{0,1}+\alpha_{1,1}N_{t-1})}{1+\exp(\alpha_{0,1}+\alpha_{1,1}N_{t-1})}$ $\pi_{t,2}^{s_t} = \frac{\exp(\alpha_{0,2}+\alpha_{1,2}N_{t-1})}{1+\exp(\alpha_{0,2}+\alpha_{1,2}N_{t-1})}$
Exponential	$p_{t,1}(s_t) = \rho_{t,1} \exp\left(-\rho_{t,1} s_t\right)$ $p_{t,2}(s_t) = \rho_{t,2} \exp\left(-\rho_{t,2} s_t\right)$	$\rho_{t,1} = \exp\left(\alpha_{0,1} + \alpha_{1,1}N_{t-1}\right)$ $\rho_{t,2} = \exp\left(\alpha_{0,2} + \alpha_{1,2}N_{t-1}\right)$
Normal	$p_{t,1}(s_t) = \frac{1}{\sqrt{2\pi\sigma_{t,1}^2}} \exp\left[-\frac{1}{2}\left(\frac{s_t^2}{\sigma_{t,1}^2}\right)\right]$ $p_{t,2}(s_t) = \frac{1}{\sqrt{2\pi\sigma_{t,2}^2}} \exp\left[-\frac{1}{2}\left(\frac{s_t^2}{\sigma_{t,2}^2}\right)\right]$	$\sigma_{t,1}^2 = \exp\left(\alpha_{0,1} + \alpha_{1,1}N_{t-1}\right)$ $\sigma_{t,2}^2 = \exp\left(\alpha_{0,2} + \alpha_{1,2}N_{t-1}\right)$

Table 4.3:   Definitions of state-dependent distributions of the HMMs for

t=1,2,...,T.

two-state HMM

Distribution in state i	Working parameters	Natural parameters
Poisson	$\eta_i = \log(\lambda_i)$	$\lambda_i = \exp(\eta_i)$
Bernoulli	$\eta_i = \log\left(\frac{p_i}{1-p_i}\right)$	$p_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}$

two-state HMM incorporating volume

Distribution in state i	Working parameters	Natural parameters
Bernoulli	$\boldsymbol{\eta}_t = \begin{pmatrix} \log\left(\frac{p_{1,t}}{1-p_{1,t}}\right) \\ \log\left(\frac{p_{2,t}}{1-p_{2,t}}\right) \end{pmatrix}$ $= \begin{pmatrix} \alpha_{0,1} & \alpha_{1,1} \\ \alpha_{0,2} & \alpha_{1,2} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ N_t \end{pmatrix}.$	$\mathbf{P}_t = \begin{pmatrix} p_{1,t} \\ p_{2,t} \end{pmatrix}$ $= \begin{pmatrix} \frac{\exp(\eta_{1,t})}{1+\exp(\eta_{1,t})} \\ \frac{\exp(\eta_{2,t})}{1+\exp(\eta_{2,t})} \end{pmatrix}.$
exponential	$\boldsymbol{\eta}_t = \begin{pmatrix} \log(\rho_{1,t}) \\ \log(\rho_{2,t}) \end{pmatrix}$ $= \begin{pmatrix} \alpha_{0,1} & \alpha_{1,1} \\ \alpha_{0,2} & \alpha_{1,2} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ N_t \end{pmatrix}.$	$\mathbf{R}_t = \begin{pmatrix} \rho_{1,t} \\ \rho_{2,t} \end{pmatrix}$ $= \begin{pmatrix} \exp(\eta_{1,t}) \\ \exp(\eta_{2,t}) \end{pmatrix}.$
normal	$\boldsymbol{\eta}_t = \begin{pmatrix} \log(\sigma^2_{1,t}) \\ \log(\sigma^2_{2,t}) \end{pmatrix}$ $= \begin{pmatrix} \alpha_{0,1} & \alpha_{1,1} \\ \alpha_{0,2} & \alpha_{1,2} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ N_t \end{pmatrix}.$	$\boldsymbol{\Sigma}_t = \begin{pmatrix} \sigma^2_{1,t} \\ \sigma^2_{2,t} \end{pmatrix}$ $= \begin{pmatrix} \exp(\eta_{1,t}) \\ \exp(\eta_{2,t}) \end{pmatrix}.$

Table 4.4: Definitions of transformations of parameters of state-dependent distributions for t=1,2,...,T. The working parameters are unconstrained and the natural parameters are constrained.

Implementation of the HMM incorporating volume

In this section we give an account of the methods we used to overcome each of the main problems that arises when we implement the HMM incorporating volume.

1. (Independent gamma random variables) In the simulation to test the two-state Bernoulli-HMM incorporating volume we found that a sequence of volumes $\{N_t\}$ generated by some arbitrarily chosen distribution leads to a substantial bias. To overcome this problem we first choose a trading day in which the price change probability is strongly dependent on N_t (in this case the data for Allianz on 20 May 2004). Then we equate the average and the sample variance of $\{n_t\}$ to the expectation and the variance of a gamma random variable respectively. We solve for the shape and the scale parameters, which we denote by $\hat{a}$ and $\hat{s}$ respectively, of the gamma distribution simultaneously. Finally, we generate an N_t series of length n from a gamma distribution with parameters $\hat{a}$ and $\hat{s}$ by using the **R** function `rgamma`.

2. (Bias relative to population parameter) In order to verify our software we use simulation to determine the difference between the true value of a parameter and the expected value of its estimator. To assess the extent of the bias,

however, we use the ratio

$$\%bias = \frac{\left|mean\left(\hat{\theta}\right) - \theta\right|}{|\theta|} \cdot 100,$$

where θ denotes the true value of a parameter and $mean\left(\hat{\theta}\right)$ the average of the estimates we obtain for θ. That is, $\%bias$ represents the bias as a percentage of θ. In applying this measure we specify θ that is considerably different from 0.

3. (Independence of the consecutive intra-daily bivariate series) The series to which we fit a model consist of intraday series in consecutive trading days. However, no data are available in periods outside of the trading hours, since no trading took place. One way to account for this fact in our model is to assume that the intraday series are independent of each other. Let D denote the number of trading days in the period. For day d (for $d = 1, 2, \ldots, D$) we define $T^{(d)}$ as the length of the series, $\mathbf{N}^{(d)} = \left(N_1^{(d)}, \ldots, N_{T^{(d)}}^{(d)}\right)'$ as the volumes series and $\mathbf{S}^{(d)} = \left(s_1^{(d)}, \ldots, s_{T^{(d)}}^{(d)}\right)$ as the returns series. Then applying Equation (3.11) to the series in day d, we get the likelihood

$$L_{T^{(d)}}^{(d)} = \boldsymbol{\delta} \cdot \mathbf{P}_1\left(s_1^{(d)}, N_0^{(d)}\right) \cdot \boldsymbol{\Gamma}\mathbf{P}_2\left(s_2^{(d)}, N_1^{(d)}\right) \cdot \cdots \cdot \boldsymbol{\Gamma}\mathbf{P}_{T^{(d)}}\left(s_{T^{(d)}}^{(d)}, N_{T^{(d)}-1}^{(d)}\right) \cdot \mathbf{1}'.$$

By assuming independence of the intraday series the log-likelihood of the HMM incorporating volume is given by

$$l_D = l_{T^{(1)}}^{(1)} + \cdots + l_{T^{(D)}}^{(D)}.$$

4. (Missing values in bivariate series) An alternative way to account for a bivariate series that consists of observations in consecutive days in our model is to insert a sequence of missing values between the series in adjacent days. We can then fit our models to the consolidated series thus obtained. For example, we have an observation series that covers two trading days, which we display here

<table>
<tr><td align="center">Day 1</td><td></td><td></td><td align="center">Day 2</td><td></td><td></td></tr>
<tr><td>$x_1^{(1)}$</td><td>$x_2^{(1)}$</td><td>$x_3^{(1)}$</td><td>$x_1^{(2)}$</td><td>$x_2^{(2)}$</td><td>$x_3^{(2)}$</td></tr>
<tr><td>$N_1^{(1)}$</td><td>$N_2^{(1)}$</td><td>$N_3^{(1)}$</td><td>$N_1^{(2)}$</td><td>$N_2^{(2)}$</td><td>$N_3^{(2)}$</td></tr>
</table>

By inserting a missing value between days 1 and 2 we obtain the following bivariate series:

<table>
<tr><td>$x_2^{(1)}$</td><td>$x_3^{(1)}$</td><td>NA</td><td>$x_2^{(2)}$</td><td>$x_3^{(2)}$</td></tr>
<tr><td>$N_1^{(1)}$</td><td>$N_2^{(1)}$</td><td>NA</td><td>$N_1^{(2)}$</td><td>$N_2^{(2)}$</td></tr>
</table>

For the series that we use, ten missing values (coded, for example, as `NA` in **R**) are inserted between the trading days whenever we want to fit our models to the consolidated series. A test for the software in this case is as follows. In the likelihood Equation (3.4), the state-dependent distribution matrix for an observation that is missing is replaced by the identity matrix.