Technische Universität München

Zentrum Mathematik

Statistical Analysis of Intraday Option Price Changes using extended Count Regression Models

Diplomarbeit

von

Andreas Kolbe

Themensteller:	Prof. Dr. C. Czado
Betreuer:	Prof. Dr. C. Czado, Dipl.-Math G. Müller
Abgabetermin:	31. März 2004

Hiermit erkläre ich, dass ich die Diplomarbeit selbständig angefertigt und nur die angegebenen Quellen verwendet habe.

München, 31. März 2004

Acknowledgement

I would like to thank Professor Claudia Czado for giving me the possibility of preparing a diploma thesis in the interesting field of statistical market microstructure research. I am also grateful for her continuous support and valuable advice throughout the development of this thesis.

Furthermore I would like to thank Gernot Müller, who contributed to this thesis by giving me valuable technical advice and who also helped me with questions concerning LaTeX and formal issues.

Finally, I would like to thank Stefan Dirnstorfer, who helped to collect the data used and analyzed in this thesis.

Contents

Chapter 1

Introduction

In the last decades options and other financial derivatives have become more and more important. At the EUREX, the world's largest futures and options exchange, jointly operated by Deutsche Börse AG and SWX Swiss Exchange, a record volume of 801 million of derivative contracts were traded in 2002. Naturally, options and option markets have also become a major topic in various fields of academic research.
In the field of financial econometrics and market microstructure theory the recent widespread availability of intraday quotation and transaction data for exchange-traded securities, sometimes also called ultra-high frequency financial data, has had an important impact on research. The investigation and statistical modelling of these data has become a new focus of academic interest since quotation and transaction data provide detailed information about the trading and asset pricing process and therefore contribute to the verification of various theoretical implications in these academic fields. One focus of current research is the further development of autoregressive conditional duration (ACD) models based on the work of Engle (2000), while other publications (e.g. Liesenfeld and Pohlmeier (2003)) are concerned with the adequate modelling of the price process at transaction level within a count data framework. Since transaction price changes of exchange-traded securities are measured in multiples of a smallest possible incremental price change, count data models which incorporate other marks of the trading process as regressors are used. Applications of these models have had their focus on stock and bond markets.
In this thesis we study non-zero transaction price changes of equity index options with a representative example of an option on the XETRA DAX index, traded at the EUREX exchange. We consider absolute non-zero transaction price changes of the option and develop and fit statistical regression models to investigate the relationship between these price movements and other marks of the trading process such as the time between two consecutive price changes, the quoted Bid-Ask spreads and the number of new quotations between the price changes. We also consider the influence of option-specific quantities such

as the intrinsic value of the option at the time of the trade and the price development of the underlying on the absolute transaction price changes.
A natural starting point for a count data regression is the Poisson-Generalized Linear Model (GLM) which we extend in order to account for the time series structure and possibly for a non-standard variance structure of the data. Regression models for time series of counts have been discussed for example by Zeger (1988) or Davis et al. (1999) with econometric and biometric applications. A detailed introduction to GLMs can be found for example in McCullagh and Nelder (1989).
Since parameter estimation in the more complex models is not feasible with standard techniques such as maximum-likelihood estimation we use Markov Chain Monte Carlo (MCMC) methods for parameter estimation. MCMC techniques are discussed for example in Gamerman (1997). It is, of course, not straightforward to compare the adequacy of different modelling approaches, especially in a Bayesian context. We therefore introduce an information criterion proposed by Spiegelhalter et al. (2002) and finally compare and assess the adequacy of the different models by using this criterion. For the computation the S-Plus and WinBUGS software packages are used throughout this thesis.

The thesis is organized as follows: Chapter 2 gives a short introduction to the basic economic background and terminology related to options and option markets in general and to the EUREX exchange in particular. Chapter 3 provides an overview of the most important mathematical concepts which we need for the statistical modelling of our option data. In Chapter 4 we discuss regression models for time series of counts in general and explicitly introduce two examples for models of this type. The available data is presented and exploratively analyzed in Chapter 5. In Chapter 6 we fit several regression models to our data and give first interpretations of the estimation results. In Chapter 7 the adequacy of the previously considered models is assessed and finally, Chapter 8 relates the results to some implications of market microstructure theory. WinBUGS programme codes for the models are collected in the appendix.

Chapter 2

Economic background

In this chapter we give a short introduction to the economic background of options and to the trading of these options at derivative exchanges or over-the-counter. A detailed discussion of all economic aspects of options and other financial derivatives can be found in Hull (2003). Information about trading procedures at EUREX exchange and about EUREX in general can be obtained from the official EUREX website www.eurexchange.com.

2.1 Basic terminology and characteristics of options

An option is a privilege sold by one party to another that offers the buyer the right, but not the obligation, to buy (*Call option*) or sell (*Put option*) an underlying asset by a certain date for an agreed-upon price. The price in the contract is called *exercise price* or *strike price*. The date in the contract is known as *expiration date* or *maturity*. Options that can be exercised at any time up to the expiration date are referred to as *American options*, while those options that can only be exercised on the expiration date itself are called *European options*.

Of course, there are two sides to every option contract. Firstly, the investor who has bought the option or taken a *long position* and secondly the investor who has sold the option or taken a *short position*. The seller of the option is also called the *writer* of the option. The writer of an option receives a certain amount of cash, the option price, up-front from the buyer. After paying the up-front option price, the holder of a long position in an option does not have any potential liabilities later. The potential losses from a short position, however, are theoretically unlimited. The following table shows the payoffs of the four possible option positions at maturity of the option. S_T denotes the price of the underlying asset at maturity, K the strike price of the option.

	Call	Put
Long position	$max(S_T - K, 0)$	$max(K - S_T, 0)$
Short position	$-max(S_T - K, 0)$	$-max(K - S_T, 0)$

The *intrinsic value of a Call option* at time $t \leq T$ is defined as (see e.g. Hull (2003)):

$$max(S_t - K, 0).$$

An option with intrinsic value > 0 is called *in-the-money*, an option with intrinsic value $= 0$ is called *at-the money* if the current price of the underlying is equal to the strike price and *out-of-the-money* if the current price of the underlying is below the strike price of the option. For Put options the intrinsic value is defined as $max(K - S_T, 0)$ and the in-the-money, at-the-money, out-of-the-money terminology is used analogically.
In general the present value of a European option equals the intrinsic value plus the *time value* of the option. The time value reflects the possibility of favourable movements in the price of the underlying in the future. It is of course equal to zero when the option has reached maturity.
Currently, options on stocks, stock indices, foreign currencies, futures contracts and other underlyings are actively traded at derivative exchanges.
We now explain who may be interested in having an option position and in what kind of markets options can be traded.

2.2 Option market mechanics

In the last decades options and other financial derivatives have become more and more important. At the EUREX only, the world's largest futures and options exchange, which we will introduce later in this chapter, a record volume of 801 million of derivative contracts were traded in 2002. The main reason for the success of the option markets is that they have attracted different types of traders and that they have a great deal of liquidity. In the following paragraphs we will introduce these different types of traders and their financial intermediaries that execute orders and provide liquidity to the market.

2.2.1 Market participants

Types of traders

There are three broad categories of traders that actively trade options: hedgers, speculators and arbitrageurs. *Hedgers* use options and other financial derivatives to reduce their risk exposure. As an example consider an investor (e.g. a life insurance company)

who owns 1,000 units of a particular underlying asset, e.g. a specific stock or a portfolio replicating a stock index like the German DAX index. The current price of one unit of the underlying is EUR 1,100.- which yields a total value of EUR 1.1 million for the entire holding. The investor is concerned that the economic development may cause the underlying asset to decline sharply in the next two months and wants protection. The investor could buy Put options with a strike price of EUR 1,000.- which would give him the right to sell 1000 units of the underlying asset for EUR 1,000.- each in two months time. If one Put option contract gives the buyer the right to sell one unit of the underlying and the quoted option price is EUR 25.-, the investor could buy 1,000 Put options and the total cost of the hedging strategy will be EUR $1000 \cdot 25 = 25,000.-$. If the market price of the underlying asset is below EUR 1000.- at maturity the options can be exercised so that EUR 975,000.- is realized for the entire holding with the initial cost for the options taken into account. If the market price of the underlying asset stays above EUR 1,000.-, the options expire worthless. In both cases, the value of the holding is always at least EUR 975,000.- with the cost for the options taken into account. The purchase of the Put options can therefore be regarded as a hedging strategy to reduce the risk of potential losses.

While hedgers seek to reduce their risk exposure, *speculators* explicitly wish to take a position in the market and bet on market movements. Suppose a speculator is convinced that the value of a certain underlying will increase over the next two months. The price of the underlying, e.g. a certain stock, is currently EUR 50.- and a two month Call option with a strike price of EUR 60.- is currently selling for EUR 1.-. Suppose the speculator is willing to invest EUR 5000.-. With this amount he can either buy 100 stocks or 5000 Call options. Transaction costs are assumed to be zero in this example. Let us consider two scenarios. In the first scenario, we assume the speculator's hunch to be correct and the price of the underlying rises to EUR 70.-. In the second scenario, the price of the underlying decreases to EUR 45.-. The following table shows the profits and losses of the speculator with the two investment strategies under the two scenarios.

Profits and losses of the speculator		
	Price of underlying after 2 months	
Investor's strategy:	EUR 45.-	EUR 70.-
Buy underlying	$-500.-$	$2000.-$
Buy Call options	$-5000.-$	$45,000.-$

If the speculator decides to buy the Call options and his hunch is correct, the value of one option will be EUR 10.- at maturity (which is equal to its intrinsic value). The total

value of his position is therefore EUR 45,000.- if one takes into account the initial cost of EUR 5,000.- for the purchase of the options. Under this scenario the option strategy is much more profitable than the direct purchase of the underlying. However, options also give rise to a greater potential loss if the price of the underlying is below the strike price at maturity and the options expire worthless. Generally speaking, for a given investment the use of options magnifies the financial consequences. This effect is called *leverage effect*.

The third important group of traders are *arbitrageurs*. Arbitrage involves locking in risk-less profits by simultaneously entering into transactions in two or more markets. As an example suppose that the price of a certain underlying is EUR 30.- and that a European Call option with maturity in two months and strike price EUR 30.- sells for EUR 2.-. A European Put option on the underlying with the same strike price and maturity sells for EUR 1.-. Consider the two portfolios:

Portfolio A:	1 Call option plus EUR 30.- of cash
Portfolio B:	1 Put option plus 1 unit of the underlying

Let us assume that there are no transaction costs and that the risk-free interest rate is equal to 0, i.e. we do not consider the time value of money in this easy example. Obviously, at maturity of the options both portfolios are worth $max(S_T, 30)$, where S_T denotes the price of the underlying at expiration of the options. As the two portfolios do not have the same value today, there exists an arbitrage opportunity. The correct arbitrage strategy is to buy the securities in Portfolio B and to enter into a short position in the Call option (i.e. sell the Call option). This generates a cash-flow of EUR $2-1-30 = -29$. At maturity of the options either the Call or the Put will be exercised. The short Call and long Put position therefore leads to the stock being sold for EUR 30.-, which generates a cash flow of 30. The net profit of the arbitrageur is therefore a risk-free EUR 1.-. Arbitrage opportunities like the one in the example can not last for long. As an arbitrageur would try to massively buy the Put and massively short the Call, the forces of supply and demand would drive up the price of the Put option and drive down the price of the Call option, so that the arbitrage opportunity vanishes. The very existence of arbitrageurs ensures therefore that there are only very small arbitrage opportunities and that markets work efficiently.
In the example, the *Put-Call-Parity* was violated which says that the value of a European Put option can always be derived from the value of the corresponding European Call option by the equation $C + Ke^{-rT} = P + S_t$, where C and P denote the values of the Call and the Put option respectively, K is the strike price of the option, r the risk-free interest

rate, T the time to maturity of the options and S_t the current price oft the underlying. For further details about the Put-Call-Parity see Hull (2003), Chapter 8.4.

Brokers, Dealers and Market Makers

Before we can introduce brokers, dealers and market makers and their functions in the option market, we have to explain the way option prices are quoted at derivative exchanges. There are normally three quoted prices: The Bid price, the Ask price and the Last price. The *Bid price* is the price a buyer is willing to pay for a security. The highest quoted bid for a particular security is referred to as *best bid*. The *Ask price* or *offer price* is the price a seller is willing to accept for a security. The lowest quoted ask for a particular security is referred to as *best ask*. Finally, the *Last price* is the value the security was last traded for. The difference between best ask and best bid is called *Bid-Ask-spread*. Of course, a transaction may only be done if best bid $\geq$ best ask.

A *broker* is an individual or firm that charges a fee or commission for executing buy and sell orders submitted by an investor. An order is to be executed immediately by the broker. The investor can of course specify the least favourable price at which the order can be executed. These orders are called limit orders. A broker normally charges a certain fixed cost as commission plus a certain proportion of the total amount of the trade. If an option is exercised the commission is normally the same as it would be if the investor placed an order to buy or sell the underlying. As, due to the leverage effect, the total transaction volume of the trade when purchasing or selling the underlying is normally higher than when trading in options, the commissions may be higher for exercising an option than for selling it shortly before maturity. The commission system therefore tends to push retail investors in the direction of selling options rather than exercising them.

A *dealer* is an individual or firm willing to buy or sell securities for their own account. Banks often act as both brokers and dealers in the market. They earn commissions when executing orders on behalf of their clients and they try to explore arbitrage opportunities or hedge their positions when trading for their own account.

Option exchanges use *market makers* to facilitate trading in a particular security. A market maker is both a broker and dealer willing to accept the risk of holding a particular number of shares of a particular security in order to facilitate trading in that security. Each market maker competes for customer order flow by displaying bid and ask quotations for a guaranteed number of shares. Once an order is received, the market maker will immediately sell from its own inventory, buy on its own account or seek the other side of

the trade so that the trade can immediately be executed. Market makers ensure that both buy and sell orders can always be executed without any delay and therefore add liquidity to the market. They make their profits from the bid-ask spread. At the EUREX exchange there are currently 27 market makers for options on the XETRA DAX index.

2.2.2 Exchange-traded markets and Over-the-counter markets

In addition to trading at a derivative exchange, options can also be traded *over-the-counter*. In the over-the-counter market financial institutions, corporate treasurers and fund managers trade over the phone. The instruments that are traded over-the-counter are often structured products designed by financial institutions in order to meet the precise needs of their clients. This involves choosing underlyings, exercise dates or strike prices that are different from those offered at exchanges. Furthermore, the structure of these over-the-counter options may be different from the standard structure of Calls and Puts as introduced in the previous sections. Options with standard structure are usually called *plain vanilla options*, options with a non-standard structure are called *exotic options*.
The main disadvantage of the over-the-counter market is that the purchaser of the option is subject to some credit risk, since the writer of the option may default. At a derivative exchange the purchaser of the option is not exposed to this credit risk as the writer of the option is required to constantly maintain funds in a *margin account*. The size of the margin required depends on the circumstances. Both the exchange and the investor's broker want to be satisfied that the investor will not default if the option is exercised. In addition to this there exists an emergency fund maintained by the exchange which can be used if any member of the exchange defaults on an option obligation.

2.3 The EUREX exchange

As the data set we will analyze later in this thesis is intraday option price data of options traded at the EUREX exchange we will give a short overview of the specific trading conditions and contract specifications at this exchange. The EUREX exchange has become the world's largest futures and options exchange. It is jointly operated by Deutsche Börse AG and SWX Swiss Exchange and is based in Frankfurt/Main. A record volume of 801 million of derivative contracts were traded at EUREX in 2002.

2.3.1 Trading mechanisms and regulatory conditions

The Eurex market is a fully integrated electronic exchange whose trading participants are connected to the Eurex system via a communications network with currently 700 connected locations worldwide. All option contracts are exchange-issued options, in contrast to bank-issued options traded, for example, at the Euwax exchange in Stuttgart. For a detailed discussion and comparison of these competing exchanges see Bartram and Fehle (2003).

Option contracts at Eurex are standardized with respect to underlying, exercise style, expiration date and strike prices. There are always pairs of calls and puts and new contracts are created according to specific rules governing, for example, the addition of new strike prices and expiration dates.

Orders are automatically matched via an electronic order book and unmatched orders are typically filled with a Eurex market maker. For the market makers there are typically exchange-mandated maximum bid-ask spreads, a minimum volume that has to be guaranteed to be traded at the currently quoted Bid- and Ask prices and a minimum period for maintaining quotes.

Bartram and Fehle (2003) come to the conclusion that the security design/creation and liquidity provisions at Eurex may mainly attract large non-speculative institutional investors, whereas small speculative investors were likely to prefer the Euwax market.

2.3.2 Option contract specifications

In this section we briefly quote the most important contract specifications for options on the XETRA DAX traded at Eurex since we will analyze data for these particular options later. The complete contract specifications can be obtained directly from the Eurex exchange and they are also published on the Eurex exchange website:

www.eurexchange.com/products/specifications.

Contract value

EUR 5.- per DAX index point.

Quotation

In points, carried out one decimal place.

Minimum price movement (=tick size)

0.1 of a point, representing a value of EUR 0.50.

A *tick* is the smallest increment in which the price of a derivatives contract may trade. For a detailed discussion and modelling of the relationship between tick sizes, Bid-Ask spreads and quotation sizes see Harris (1994).

Last trading day
The third Friday of the expiration month, if that day is an exchange trading day; otherwise, on the exchange trading day immediately prior to that Friday. Trading ceases at the start of the intraday trading auction on the electronic trading system (XETRA) at 1:00 pm CET.

Exercise period
European style, i.e. an option may only be exercised on the last trading day of the options series up to the end of the post-trading period.

Settlement
Cash settlement, payable on the first exchange trading day immediately following the last trading day.

Final Settlement Price
The value of the DAX, determined on the basis of the collective prices of the shares contained in the DAX index as reflected in the intraday trading auction on the electronic system of the Frankfurt Stock Exchange (XETRA).

Chapter 3

Mathematical basics

3.1 Generalized Linear Models

Generalized Linear Models (GLMs) are used for modelling a non-normally distributed variable of interest (response) with a series of explanatory variables (regressors, predictors). In linear models the response is assumed to be normally distributed with a mean which is modelled as linear in the predictors.
In order to build a GLM McCulloch and Searle (2001) list three questions that have to be answered:

(i) What is the distribution of the response (for fixed values of the predictors and possibly after a transformation)?

(ii) What function of the mean will be modelled as linear in the predictors?

(iii) What will the predictors be?

Answers to questions number (i) and (ii) will determine the structure of the GLM, which we will further specify in the following paragraphs. In order to compare different models and to be able to consistently select predictors goodness-of-fit measures are needed which we will introduce in Chapter 3.1.4.

3.1.1 Components of a GLM

Distribution of the response variable Y

The vector $\mathbf{Y} = (Y_1, ..., Y_n)^t$ is assumed to consist of independent observations from a distribution with density of the form

$$f_{Y_i}(y_i, \gamma_i, \tau^2) = exp\left\{\frac{y_i\gamma_i - b(\gamma_i)}{\tau^2} + c(y_i, \tau)\right\} \tag{3.1}$$

with two functions $b : \mathbb{R} \to \mathbb{R}$ and $c : \mathbb{R}^2 \to \mathbb{R}$, which is called the "canonical form" of a GLM density.

Definition 3.1 (Exponential family) *A family of distributions* $\{P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ *is called k-parametric exponential family if there exist real valued functions* $c_i(\boldsymbol{\theta}), d(\boldsymbol{\theta})$ *on* Θ, T_i *and* S *on* $\mathbb{R}^n$ *with* $A \subset \mathbb{R}^n$, *so that the density of* $P_{\boldsymbol{\theta}}$ *can be written as*

$$p(\mathbf{y}, \boldsymbol{\theta}) = exp\left\{\sum_{i=1}^{k} c_i(\boldsymbol{\theta}) T_i(\mathbf{y}) + d(\boldsymbol{\theta}) + S(\mathbf{y})\right\} \cdot 1_A(\mathbf{y})$$

Note that for τ^2 known, the canonical GLM-density f_{Y_i} in (3.1) is a density from the exponential family where

$$\begin{aligned} c_i(\boldsymbol{\theta}) &= \frac{\gamma_i}{\tau^2} \\ T_i(\mathbf{y}) &= y_i \\ d(\boldsymbol{\theta}) &= -\sum_{i=1}^{n} \frac{b(\gamma_i)}{\tau^2} \\ S(\mathbf{y}) &= \sum_{i=1}^{n} c(y_i, \tau). \end{aligned}$$

For τ^2 unknown, this may not be the case.
Note that, conversely, most commonly-used distributions from the exponential family can be written in the form (3.1). Examples are the Binomial-, Poisson- and the Gamma distribution.
As we will have to deal with an extension of Poisson GLMs later, let us consider the case of the Poisson distribution more closely. The probability function of a Poisson-distributed random variable Y is given by:

$$f_Y(y, \lambda) = P(Y = y) = e^{-\lambda} \frac{\lambda^y}{y!}$$

This can be written as:

$$f_Y(y, \lambda) = exp\{-\lambda + y \cdot log\lambda - log(y!)\}$$

Defining $\gamma := log\lambda$ and $\tau := 1$ we get:

$$f_Y(y, \gamma, \tau) = exp\{\frac{y\gamma - e^\gamma}{\tau^2} - log(y!)\}$$

which is the canonical form of a GLM density with $b(\gamma) = e^\gamma$ and $c(y, \tau) = -log(y!)$.

The link function

When building a regression model one wants to relate the parameters of the distribution of the response y_i to the regressors $\mathbf{x}_i = (x_{i1}, ..., x_{ip})^t$, where p denotes the total number of regressors. In the case of the GLMs this is done by modelling a transformation of the mean

$$\mu_i := E[Y_i|\mathbf{x}_i]$$

(which is a function of γ_i) as a linear model in the predictors:

$$g(\mu_i) = \mathbf{x}_i^t \boldsymbol{\beta}$$

where $g : \mathbb{R} \to \mathbb{R}$ is a known function called the link function (since it links together the mean of Y_i and the linear form of the predictors), $\mathbf{x}_i$ the vector of regressors for the i-th observation, i.e. the i-th row of the Design matrix of the model, and $\boldsymbol{\beta}$ the parameter vector.

Definition 3.2 (Canonical link) *In a GLM with $\gamma_i = g(\mu_i)$, g is called the "canonical link" function.*

In the Poisson-GLM the canonical link is the log-link, i.e. $g : \mathbb{R} \to \mathbb{R}$ with $g(x) = log(x)$.

3.1.2 Some useful properties

In this section we proof some properties of a GLM which will be useful for parameter estimation later. Under regularity conditions (for details see e.g. Casella and Berger (1990)) the following equations hold with l being defined as the log-likelihood $l := log f_{Y_i}(y_i, \gamma_i, \tau)$:

$$E\left(\frac{\partial l}{\partial \gamma_i}\right) = 0 \tag{3.2}$$

$$E\left(\frac{\partial^2 l}{\partial \gamma_i^2}\right) + E\left(\left[\frac{\partial l}{\partial \gamma_i}\right]^2\right) = 0 \tag{3.3}$$

Proof:

$$\frac{\partial l}{\partial \gamma_i} = \frac{\partial log f}{\partial \gamma_i} = \frac{\frac{\partial f}{\partial \gamma_i}}{f}$$

$$\Rightarrow E\left(\frac{\partial l}{\partial \gamma_i}\right) = E\left(\frac{\frac{\partial f}{\partial \gamma_i}}{f}\right) = \int \frac{\partial f(y_i)}{\partial \gamma_i} \cdot \frac{1}{f(y_i)} \cdot f(y_i) dy_i \overset{\text{reg. conditions}}{=} \frac{\partial}{\partial \gamma_i}\left[\int f(y_i) dy_i\right] = 0$$

This proves Equation 3.2.

For the proof of Equation 3.3 consider:

$$E\left(\frac{\partial^2 l}{\partial \gamma_i^2}\right) = E\left(\frac{\partial(\frac{1}{f}\cdot\frac{\partial f}{\partial \gamma_i})}{\partial \gamma_i}\right) = E\left(\frac{1}{f}\cdot\frac{\partial^2 f}{\partial \gamma_i^2} + \frac{\partial f}{\partial \gamma_i}\cdot\frac{-\frac{\partial f}{\partial \gamma_i}}{f^2}\right) = E\left(\frac{\frac{\partial^2 f}{\partial \gamma_i^2}}{f}\right) - E\left[\left(\frac{\frac{\partial f}{\partial \gamma_i}}{f}\right)^2\right] =$$

$$= \int \frac{\frac{\partial^2 f(y_i)}{\partial \gamma_i^2}}{f(y_i)}\cdot f(y_i)dy_i - E\left[\left(\frac{\partial l}{\partial \gamma_i}\right)^2\right] =$$

$$\overset{\text{reg. conditions}}{=} \frac{\partial^2}{\partial \gamma_i^2}\left[\int f(y_i)dy_i\right] - E\left[\left(\frac{\partial l}{\partial \gamma_i}\right)^2\right] = -E\left[\left(\frac{\partial l}{\partial \gamma_i}\right)^2\right]$$

□

Note that for a GLM density in the canonical form

$$\frac{\partial l}{\partial \gamma_i} = \frac{y_i - \frac{\partial b}{\partial \gamma_i}}{\tau^2}$$

$$\overset{(3.2)}{\Longrightarrow} \quad \frac{E(Y_i) - \frac{\partial b}{\partial \gamma_i}}{\tau^2} = 0$$

$$\iff \quad \mu_i := E(Y_i) = \frac{\partial b}{\partial \gamma_i} \tag{3.4}$$

Furthermore,

$$\frac{\partial^2 l}{\partial \gamma_i^2} = \frac{-\frac{\partial^2 b}{\partial \gamma_i^2}}{\tau^2} \quad \overset{(3.3)}{\Longrightarrow} \quad \frac{-\frac{\partial^2 b}{\partial \gamma_i^2}}{\tau^2} + E\left[\left(\frac{y_i - \frac{\partial b}{\partial \gamma_i}}{\tau^2}\right)^2\right] = 0$$

$$\overset{(3.4)}{\Longleftrightarrow} \quad \frac{-\frac{\partial^2 b}{\partial \gamma_i^2}}{\tau^2} + \frac{Var(Y_i)}{\tau^4} = 0$$

$$\iff \quad Var(Y_i) = \tau^2 \cdot \frac{\partial^2 b}{\partial \gamma_i^2} := \tau^2 \cdot v(\mu_i) \tag{3.5}$$

where $v(\mu_i)$ is called the *variance function*.
Two other useful identities are:

$$\frac{\partial \gamma_i}{\partial \mu_i} = \left(\frac{\partial \mu_i}{\partial \gamma_i}\right)^{-1} = \left(\frac{\partial^2 b(\gamma_i)}{\partial \gamma_i^2}\right)^{-1} = \frac{1}{v(\mu_i)} \tag{3.6}$$

and

$$\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} = \frac{\partial \mu_i}{\partial g(\mu_i)}\cdot\frac{\partial g(\mu_i)}{\partial \boldsymbol{\beta}} = \left(\frac{\partial g(\mu_i)}{\partial \mu_i}\right)^{-1}\cdot\frac{\partial \mathbf{x_i}^t\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = \left(\frac{\partial g(\mu_i)}{\partial \mu_i}\right)^{-1}\cdot\mathbf{x}_i \tag{3.7}$$

With these identities it is now easy to derive the maximum-likelihood equations.

3.1.3 Parameter estimation

The likelihood equations

For the estimation of the parameter vector $\boldsymbol{\beta} \in \mathbb{R}^p$ of the GLM we will use the maximum-likelihood method. Let $l := log f_{\mathbf{Y}}(\mathbf{y}, \boldsymbol{\gamma}, \tau)$ denote the log-likelihood of the complete model and let n be the total number of observations. With the identities introduced in the previous section, we can now derive

$$\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{1}{\tau^2} \cdot \sum_{i=1}^{n} \left[y_i \frac{\partial \gamma_i}{\partial \boldsymbol{\beta}} - \frac{\partial b(\gamma_i)}{\partial \gamma_i} \cdot \frac{\partial \gamma_i}{\partial \boldsymbol{\beta}} \right] \\
&\overset{(3.4)}{=} \frac{1}{\tau^2} \cdot \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial \gamma_i}{\partial \boldsymbol{\beta}} \\
&= \frac{1}{\tau^2} \cdot \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial \gamma_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \\
&\overset{(3.6),(3.7)}{=} \frac{1}{\tau^2} \cdot \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{v(\mu_i) \cdot \frac{\partial g(\mu_i)}{\partial \mu_i}} \cdot \mathbf{x}_i \qquad (3.8) \\
&= \frac{1}{\tau^2} \cdot \sum_{i=1}^{n} (y_i - \mu_i) \cdot w_i \cdot \frac{\partial g(\mu_i)}{\partial \mu_i} \cdot \mathbf{x}_i
\end{aligned}$$

with

$$w_i := \left[v(\mu_i) \left(\frac{\partial g(\mu_i)}{\partial \mu_i} \right)^2 \right]^{-1}$$

In matrix notation this can be written as:

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{1}{\tau^2} \cdot \mathbf{X^t W \Delta} (\mathbf{y} - \boldsymbol{\mu})$$

where $\boldsymbol{\mu} := (\mu_1, ..., \mu_n)^t$, $\mathbf{X}$ is the design matrix defined by

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix},$$

$\mathbf{W} := diag(w_i)$ is a matrix with

$$W_{i,j} = \begin{cases} w_i & \text{if } i = j \\ 0 & otherwise \end{cases}$$

and $\mathbf{\Delta} := diag\left(\frac{\partial g(\mu_i)}{\partial \mu_i} \right)$.

Thus, the ML equations have the form

$$\mathbf{X^t W \Delta y} = \mathbf{X^t W \Delta} \mu$$

Typically these p equations are highly non-linear in $\boldsymbol{\beta}$.

Solving the ML equations: The Fisher-Scoring method

A general method for solving non-linear equations is the iterative Newton-Raphson algorithm. For a system of nonlinear equations $\mathbf{f}(\boldsymbol{\theta}) = (f_1(\boldsymbol{\theta}), ..., f_p(\boldsymbol{\theta}))^t = \mathbf{0}$ and an initializing estimate $\boldsymbol{\theta}^{(0)} \in \mathbb{R}^p$ of the solution set

$$\boldsymbol{\theta}^{(m+1)} := \boldsymbol{\theta}^{(m)} - [D\mathbf{f}(\boldsymbol{\theta}^{(m)})]^{-1}\mathbf{f}(\boldsymbol{\theta}^{(m)}),$$

where $D\mathbf{f}(\boldsymbol{\theta})$ denotes the $p \times p$ Jacobi-matrix of $\mathbf{f}(\boldsymbol{\theta})$, i.e.

$$D\mathbf{f}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial f_1}{\partial \theta_1} & \cdots & \frac{\partial f_1}{\partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_p}{\partial \theta_1} & \cdots & \frac{\partial f_p}{\partial \theta_p} \end{pmatrix}.$$

Under "normal" conditions (for details see Stoehr (1999), p. 302) $\boldsymbol{\theta}^{(m)}$ will at least quadratically converge to the solution of the equation system.
For the GLM we want to solve $\frac{\partial l}{\partial \boldsymbol{\beta}} = \mathbf{0}$. To be able to apply the Newton-Raphson algorithm we need the Hessian matrix of the log-likelihood of the model. The Hessian or second derivatives matrix is given by

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}^2} = \begin{pmatrix} \frac{\partial^2 l}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 l}{\partial \beta_1 \partial \beta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \beta_p \partial \beta_1} & \cdots & \frac{\partial^2 l}{\partial \beta_p \partial \beta_p} \end{pmatrix}$$

with

$$\frac{\partial^2 l}{\partial \beta_r \partial \beta_s} \overset{(3.8)}{=} \frac{\partial}{\partial \beta_s}\left[\frac{1}{\tau^2} \cdot \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{v(\mu_i) \cdot \frac{\partial g(\mu_i)}{\partial \mu_i}} \cdot \mathbf{x}_{ir}\right] =$$

$$= \frac{1}{\tau^2} \cdot \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial}{\partial \boldsymbol{\beta}_s}\left[\frac{1}{v(\mu_i) \cdot \frac{\partial g(\mu_i)}{\partial \mu_i}} \cdot \mathbf{x}_{ir}\right] + \frac{1}{\tau^2} \cdot \sum_{i=1}^{n} \frac{1}{v(\mu_i) \cdot \frac{\partial g(\mu_i)}{\partial \mu_i}} \cdot \mathbf{x}_{ir} \cdot \frac{\partial}{\partial \beta_s}(y_i - \mu_i) =$$

$$\overset{(3.7)}{=} \frac{1}{\tau^2} \cdot \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial}{\partial \boldsymbol{\beta}_s}\left[\frac{1}{v(\mu_i) \cdot \frac{\partial g(\mu_i)}{\partial \mu_i}} \cdot \mathbf{x}_{ir}\right] + \frac{1}{\tau^2} \cdot \sum_{i=1}^{n} \frac{1}{v(\mu_i) \cdot \frac{\partial g(\mu_i)}{\partial \mu_i}} \cdot \mathbf{x}_{ir} \left(-\frac{1}{\frac{\partial g(\mu_i)}{\partial \mu_i}} \cdot \mathbf{x}_{is}\right) =$$

$$= \frac{1}{\tau^2} \cdot \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial}{\partial \boldsymbol{\beta}_s}\left[\frac{1}{v(\mu_i) \cdot \frac{\partial g(\mu_i)}{\partial \mu_i}} \cdot \mathbf{x}_{ir}\right] - \frac{1}{\tau^2} \cdot \sum_{i=1}^{n} w_i x_{ir} x_{is}$$

Taking expectations yields

$$E\left(\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}\right) = -\frac{1}{\tau^2}\cdot\sum_{i=1}^{n} w_i x_{ir} x_{is}$$

or in matrix notation

$$E\left(-\frac{\partial^2 l}{\partial\boldsymbol{\beta}^2}\right) = \frac{1}{\tau^2}\mathbf{X^t W X} \qquad \in \mathbb{R}^{\,p\times p}.$$

The matrix $E\left(-\frac{\partial^2 l}{\partial\boldsymbol{\beta}^2}\right)$ is called *expected information matrix*.
The Fisher-Scoring method is an adaption of the Newton-Raphson algorithm which uses the expected information matrix instead of the observed matrix. The iterative step in the estimating procedure is therefore:

$$\boldsymbol{\beta}^{(m+1)} := \boldsymbol{\beta}^{(m)} + (\mathbf{X^t W X})^{-1}\mathbf{X^t W}\boldsymbol{\Delta}(\mathbf{y}-\boldsymbol{\mu}) \tag{3.9}$$

where $\mathbf{W}, \boldsymbol{\Delta}, \mu$ are evaluated at $\boldsymbol{\beta}^{(m)}$.
With $\mathbf{A} := (\mathbf{X^t W X})$ Equation (3.9) can be written as:

$$\mathbf{A}\beta^{(m+1)} = \mathbf{A}\beta^{(m)} + \tau^2\cdot\frac{\partial l}{\partial\boldsymbol{\beta}^{(m)}}$$

$$\begin{aligned}
(\mathbf{A}\beta^{(m+1)})_j &\overset{(3.8)}{=} \sum_{s=1}^{p}\mathbf{A}_{js}\beta_s^{(m)} + \sum_{i=1}^{n} w_i(y_i-\mu_i)\frac{\partial g(\mu_i)}{\partial\mu_i}\cdot x_{ij} \\
&= \sum_{s=1}^{p}\sum_{i=1}^{n} w_i x_{is} x_{ij}\boldsymbol{\beta}_s^{(m)} + \sum_{i=1}^{n} w_i(y_i-\mu_i)\frac{\partial g(\mu_i)}{\partial\mu_i}\cdot\mathbf{x}_{ij} \\
&= \sum_{i=1}^{n} w_i x_{ij}\left(\sum_{s=1}^{p} x_{is}\boldsymbol{\beta}_s^{(m)} + (y_i-\mu_i)\frac{\partial g(\mu_i)}{\partial\mu_i}\right) \\
&= \sum_{i=1}^{n} w_i x_{ij}\left(g(\mu_i) + (y_i-\mu_i)\frac{\partial g(\mu_i)}{\partial\mu_i}\right) \qquad (3.10)\\
&= \sum_{i=1}^{n} w_i x_{ij} z_i
\end{aligned}$$

with $z_i := g(\mu_i) + (y_i-\mu_i)\frac{\partial g(\mu_i)}{\partial\mu_i}$.
Yet, the expression $(\mathbf{A}\beta^{(m+1)})_j$ can also be written as:

$$(\mathbf{A}\beta^{(m+1)})_j = \sum_{s=1}^{p}\mathbf{A}_{js}\beta_s^{(m+1)}$$

$$= \sum_{s=1}^{p} \left(\sum_{i=1}^{n} w_i x_{is} x_{ij} \right) \beta_s^{(m+1)}$$
$$= \sum_{i=1}^{n} w_i x_{ij} g^*(\mu_i)$$

where $g^*(\mu_i) := g(\mu_i)\big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(m+1)}}$.

Therefore we get:

$$\sum_{i=1}^{n} w_i x_{ij} z_i = \sum_{i=1}^{n} w_i x_{ij} g^*(\mu_i) \qquad j = 1, ..., p \tag{3.11}$$

These equations are equivalent to the maximum-likelihood equations for the estimation of the parameter vector $\mathbf{b} \in \mathbb{R}^p$ in the following ordinary linear model with weights $\mathbf{w}$:

$$\begin{aligned} z_i &= \mathbf{x}_i^t \mathbf{b} + \varepsilon_i, \quad i = 1, ..., n \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \mathbf{V}) \\ \mathbf{V} &= \begin{pmatrix} \frac{1}{w_i} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{w_n} \end{pmatrix} \end{aligned}$$

In this model the maximum-likelihood equations are given by

$$\sum_{i=1}^{n} w_i x_{ij} z_i = \sum_{i=1}^{n} w_i x_{ij} \mathbf{x}_i^t \mathbf{b} \qquad j = 1, ..., p \tag{3.12}$$

and thus have the same form as the equations in (3.11).
For an arbitrary starting value $\hat{\boldsymbol{\beta}}^{(0)}$ the iterative algorithm for the estimation of $\boldsymbol{\beta}$ in the GLM can therefore be formulated as:

(i) With a current estimate $\hat{\boldsymbol{\beta}}^{(m)}$ for $\boldsymbol{\beta}$, calculate

$$\begin{aligned} \hat{\mu_i}^{(m)} &:= g^{-1}(\mathbf{x}_i^t \hat{\beta}^{(m)}) \\ z_i^{(m)} &:= g(\mu_i) + (y_i - \mu_i) \frac{\partial g(\mu_i)}{\partial \mu_i}\bigg|_{\mu_i = \hat{\mu_i}^{(m)}} \\ w_i^{(m)} &:= \left[v(\mu_i) \left(\frac{\partial g(\mu_i)}{\partial \mu_i} \right)^2 \right]^{-1} \Bigg|_{\mu_i = \hat{\mu_i}^{(m)}} \end{aligned}$$

with $v(\mu_i)$ as defined in (3.5) for $i = 1, ..., n$.

(ii) Build a linear regression model with response $\mathbf{z}^{(m)}$, design matrix $\mathbf{X}$ and weights $\mathbf{w}^{(m)}$. Estimate the parameter vector $\mathbf{b}$ of this model by solving the equations in (3.12) and set $\hat{\boldsymbol{\beta}}^{(m+1)} = \hat{\mathbf{b}}$. Go to (i) if $\|\boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)}\|$ is not small enough.

3.1.4 A goodness-of-fit measure

In this section we introduce the *deviance* as a goodness-of-fit measure and the *Partial Deviance Test* as a tool to identify significant regressors.
Let $l(\mathbf{y}, \hat{\boldsymbol{\mu}})$ be the log-likelihood of a GLM for a known τ^2 and maximized over $\boldsymbol{\beta}$ with $\hat{\mu}_i = g^{-1}(\mathbf{x}_i^t \hat{\beta})$. Furthermore, let $l(\mathbf{y}, \mathbf{y})$ denote the maximized log-likelihood of the saturated model where $\hat{\mu}_i = y_i$.

Define

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) := -2\tau^2[l(\mathbf{y}, \hat{\boldsymbol{\mu}}) - l(\mathbf{y}, \mathbf{y})]$$

as the *deviance* of the model. Obviously, the better the fit, the smaller $D(\mathbf{y}, \hat{\boldsymbol{\mu}})$. Therefore $D(\mathbf{y}, \hat{\boldsymbol{\mu}})$ can be considered as a goodness-of-fit measure for GLMs.
Since the deviance is equivalent to a likelihood-ratio-test statistic, it follows that

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) \xrightarrow{d} \chi^2_{n-q} \qquad \text{for} \qquad n \to \infty$$

i.e., asymptotically the deviance follows a χ^2 distribution with $n - q$ degrees of freedom where q is the number of parameters in the model and n denotes the total number of observations. McCullagh and Nelder (1989) point out, however, that the χ^2 approximation may not be very good in some cases even for large values of n.

In order to identify significant regressors one can apply the *Partial Deviance Test*.
Let $g(\mu_i) = \mathbf{x}_{i1}^t \boldsymbol{\beta}_1 + \mathbf{x}_{i2}^t \boldsymbol{\beta}_2$ be the full model with parameter vector $\boldsymbol{\beta}^t = (\boldsymbol{\beta}_1^t, \boldsymbol{\beta}_2^t) \in \mathbb{R}^{\,p_1+p_2}$ and deviance D_F and let $g(\mu_i) = \mathbf{x}_{i1}^t \boldsymbol{\beta}_1$ be the restricted model with parameter vector $\boldsymbol{\beta} = \boldsymbol{\beta}_1 \in \mathbb{R}^{\,p_1}$ and deviance D_R.
We want to test

$$H_0: \quad \boldsymbol{\beta}_2 = \mathbf{0} \qquad \text{against} \qquad H_1: \quad \boldsymbol{\beta}_2 \neq \mathbf{0}$$

An asymptotically appropriate criteria for this test is:

$$\text{Reject } H_0 \text{ with significance level } \alpha \iff D_R - D_F > \chi^2_{p_2,1-\alpha},$$

where $\chi^2_{p_2,1-\alpha}$ denotes the $1-\alpha$ quantile of a χ^2 distribution with p_2 degrees of freedom.
Using this test one can construct analysis-of-deviance tables and successively select significant regressors. Yet, due to the lack of an exact theory for the distribution of the deviance, McCullagh and Nelder (1989) point out that the assignment of a precise significance level is in general not possible. Note also that analysis-of-deviance tables always depend on the order of the regressors.

3.1.5 GLMs with the Negative Binomial Distribution

In the previous sections we have introduced the GLM and we have mentioned the example of the Poisson-GLM which is a natural starting point for count data regression. One important property of the Poisson distribution is equidispersion, i.e.

$$E(Y_t|\mathbf{x}_t) = Var(Y_t|\mathbf{x}_t) = \lambda_t.$$

For some real data examples this may be a too restrictive approach. In many real count data sets one can observe that the variance exceeds the mean, i.e.

$$E(Y_t|\mathbf{x}_t) < Var(Y_t|\mathbf{x}_t).$$

This is commonly called *overdispersion*. Cameron and Trivedi (1998) provide a detailed discussion of regression modelling of overdispersed count data. One possible approach to build a regression model for overdispersed count data consists of using the Negative Binomial distribution instead of the Poisson distribution for the variable of interest conditional on the regressors.
In this section we briefly introduce this class of models which we will also apply to our option data in Chapter 6.

Different authors use different parameterizations for the Negative Binomial distribution. In this thesis we will stick to the one used by Cameron and Trivedi (1998) and by Venables and Ripley (1999).

Definition 3.3 (Negative Binomial distribution) *A random variable Y has Negative Binomial distribution if its probability function is of the form*

$$f_Y(y;\lambda,\alpha) = \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y!}\left(\frac{\lambda}{\alpha+\lambda}\right)^y\left(\frac{\alpha}{\alpha+\lambda}\right)^\alpha \tag{3.13}$$

for $y = 0, 1, 2, ...$ and $\lambda, \alpha > 0$.

This density can also be written as:

$$f_Y(y;\gamma,\tau) = exp\left\{log\left(\frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y!}\right) + y \cdot log\left(\frac{\lambda}{\alpha+\lambda}\right) - \alpha \cdot log\left(1+\frac{\lambda}{\alpha}\right)\right\} \tag{3.14}$$

which is the common form of a GLM density (3.1) if α is known and where

$$\begin{aligned} \gamma &= log\left(\frac{\lambda}{\alpha+\lambda}\right) \\ \Longleftrightarrow \lambda &= \frac{\alpha \cdot e^\gamma}{1-e^\gamma}, \end{aligned}$$

$$\begin{aligned} \tau &= 1, \\ b(\gamma) &= \alpha \cdot log\left(1 + \frac{\lambda}{\alpha}\right), \\ c(y, \tau) &= log\left(\frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y!}\right). \end{aligned}$$

Note that

$$\begin{aligned} b(\gamma) &= \alpha \cdot log\left(1 + \frac{\lambda}{\alpha}\right) \\ &= \alpha \cdot log\left(1 + \frac{e^{\gamma}}{1 - e^{\gamma}}\right) \\ &= \alpha \cdot log\left(\frac{1}{1 - e^{\gamma}}\right). \end{aligned}$$

The expectation and variance of Y can now be calculated according to (3.4) and (3.5):

$$\begin{aligned} E(Y) = b'(\gamma) &= \alpha(1 - e^{\gamma})\frac{e^{\gamma}}{(1 - e^{\gamma})^2} \\ &= \alpha\frac{e^{\gamma}}{1 - e^{\gamma}} \\ &= \lambda \end{aligned}$$

$$\begin{aligned} Var(Y) = b''(\gamma) \cdot \tau^2 &= \alpha\frac{e^{\gamma}(1 - e^{\gamma}) + e^{\gamma}e^{\gamma}}{(1 - e^{\theta})^2} \\ &= \alpha\frac{e^{\gamma}}{1 - e^{\gamma}} + \alpha\frac{(e^{\gamma})^2}{(1 - e^{\gamma})^2} \\ &= \lambda + \frac{1}{\alpha}\lambda^2 \end{aligned}$$

The Negative Binomial-GLM (for a known parameter α) with the log-link is therefore given by:

$$Y_t|\mathbf{x}_t \sim NegBin(\lambda_t, \alpha) \tag{3.15}$$

$$\lambda_t = exp(\mathbf{x}_t^t\boldsymbol{\beta})$$

$$E(Y_t|\mathbf{x}_t) = \lambda_t$$

$$Var(Y_t|\mathbf{x}_t) = \lambda_t + \frac{1}{\alpha}\lambda_t^2 \tag{3.16}$$

If the parameter α is unknown, (3.14) is not a standard GLM density anymore. However, Venables and Ripley (1999) provide an S-Plus library for Negative Binomial models of this kind including the function *glm.nb* which incorporates maximum-likelihood estimation of the parameter α in the Negative Binomial model setting. The function provides

summaries and ANOVA tables which produce likelihood ratio tests for a sequence of fitted models.
Parameter estimation can also be done using Markov Chain Monte Carlo estimation in the WinBUGS software.

Due to the additional parameter α the Negative Binomial model allows for overdispersion in the count data and thus provides a more flexible approach for count data regression analysis than the Poisson-GLM. In fact, the Poisson distribution can be considered as a special case of the Negative Binomial distribution when $\alpha \to \infty$. To see this, note that

$$\frac{\Gamma(y+\alpha)}{\Gamma(\alpha)} = \frac{(y+\alpha-1)!}{(\alpha-1)!} = (y+\alpha-1)(y+\alpha-2)...(y+\alpha-y) = \prod_{j=0}^{y-1}(j+\alpha) \tag{3.17}$$

and that

$$\left(\frac{1}{1+\frac{\lambda}{\alpha}}\right)^{\alpha} = \left(1+\frac{\lambda}{\alpha}\right)^{-\alpha} \quad \overset{\alpha\to\infty}{\longrightarrow} \quad e^{-\lambda}. \tag{3.18}$$

The probability function (3.13) of the Negative Binomial distribution can be written as:

$$\begin{aligned} f_Y(y) = P(Y=y) &\overset{(3.17)}{=} \left(\prod_{j=0}^{y-1}(j+\alpha)\right)\frac{1}{y!}\left(\frac{1}{\alpha+\lambda}\right)^{y}\lambda^{y}\left(\frac{\alpha}{\alpha+\lambda}\right)^{\alpha} \\ &= \left(\prod_{j=0}^{y-1}\frac{j+\alpha}{\alpha+\lambda}\right)\frac{1}{y!}\lambda^{y}\left(\frac{\alpha}{\alpha+\lambda}\right)^{\alpha} \\ &= \left(\prod_{j=0}^{y-1}\frac{\frac{j}{\alpha}+1}{1+\frac{\lambda}{\alpha}}\right)\left(\frac{1}{1+\frac{\lambda}{\alpha}}\right)^{\alpha}\frac{1}{y!}\lambda^{y} \\ &\overset{(3.18)}{\longrightarrow} 1\cdot e^{-\lambda}\frac{1}{y!}\lambda^{y} \qquad \text{for } \alpha\to\infty. \end{aligned}$$

This is the probability function of the Poisson distribution.

3.2 Time series

In this section we provide a short introduction to the basic concepts of discrete-time series analysis. A time series is a set of observations $\{X_t\}$, each one being recorded at a specified time t. A discrete-time series is one in which the set T of times at which observations are made is a discrete set. One example of a discrete-time series are intraday price changes of a specific financial security, e.g. of an exchange-traded option where x_1 denotes the first price change in the time interval of interest, x_2 the second price change, etc. Since when modelling the price changes of our option data example in the following chapters we will have to take into account the time-series structure of the data, it will be useful to get a brief overview of the basic concepts concerning time series. We will also give an introduction to the most important class of time series models, the autoregressive-moving average or shortly ARMA-models. A detailed introduction to time series analysis in general and ARMA processes in particular can be found for example in Brockwell and Davis (1991).

3.2.1 Basic concepts and definitions

A basic concept of time series is stationarity, for which two different definitions must be distinguished.

Definition 3.4 (Strict stationarity) *The time series $\{X_t, t \in \mathbb{Z}\}$ is said to be strictly stationary if the joint distributions of $(X_{t_1}, .., X_{t_k})$ and $(X_{t_1+h}, .., X_{t_k+h})$ are the same for all positive integers k and for all $t_1, ..., t_k, h \in \mathbb{Z}$.*

Definition 3.5 (Weak stationarity) *The time series $\{X_t, t \in \mathbb{Z}\}$ is said to be stationary in its weak sense if*

(i) $E(X_t) = \mu \quad \forall t$

(ii) $Var(X_t) = \sigma < \infty \quad \forall t$

(iii) $Cov(X_r, X_s) = Cov(X_{r+t}, X_{s+t}) \quad \forall r, s, t \in \mathbb{Z}$

Weak stationarity means that expectation, variance and covariance structure of an observation X_t do not depend on the point of time t when the observation is recorded. This is a weaker requirement than the parity of joint distributions required for strict stationarity. In fact, if a time series X_t is strict stationary it follows that it is also stationary in the weak sense, whereas in general the converse of this statement is not true. For details see Brockwell and Davis (1991). From now on we will use the term stationarity for weak stationarity if not explicitly stated otherwise.

Stationary processes can be characterized by its first- and second-order moments. The *autocovariance function* summarizes the covariance structure of the time series.

Definition 3.6 (Autocovariance function) *If $\{X_t, t \in T\}$ is a stationary time series, then the autocovariance function $\gamma_X(\cdot,\cdot)$ of $\{X_t\}$ is defined by*

$$\gamma_X(r,s) := Cov(X_r, X_s) = E[(X_r - E(X_r))(X_s - E(X_s))], \quad r,s \in \mathbb{Z}.$$

The autocorrelation structure of the time series is summarized in the *autocorrelation function.*

Definition 3.7 (Autocorrelation function) *If $\{X_t, t \in T\}$ is a stationary time series, then the autocorrelation function $\rho_X(h)$ of $\{X_t\}$ is defined by*

$$\rho_X(h) := Corr(X_t, X_{t+h}) = \frac{\gamma_X(h,0)}{\gamma_X(0,0)} = \frac{\gamma_X(h)}{\gamma_X(0)}, \quad \forall t,h \in \mathbb{Z}.$$

To make notation easier, especially in the following paragraph about ARMA-processes, we will use the *backward shift operator B* which is defined by:

$$B^j X_t := X_{t-j}, \quad j \in \mathbb{Z}$$

We will also use $\phi(\cdot)$ and $\theta(\cdot)$ which denote the following polynomials of degree p and q respectively:

$$\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$$
$$\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q$$

Thus, $\phi(B)X_t$ stands for $X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p}$.

3.2.2 Autoregressive (AR), moving average (MA) and autoregressive-moving average (ARMA) processes

The class of autoregressive-moving average processes plays a key role in the modelling of time series. Brockwell and Davis (1991), page 77, state the fact that for any autocovariance function $\gamma(\cdot)$ such that $\lim_{h\to\infty}\gamma(h) = 0$, and for any integer $k > 0$ it is possible to find an ARMA process with autocovariance function $\gamma_X(h) = \gamma(h)$, $h = 0, 1, ..., k$ as the major reason for the outstanding role of ARMA processes.
We will first define autoregressive moving average-processes in general and introduce some basic properties and then discuss the special cases of pure autoregressive (AR)- and pure moving average (MA)-processes.

Before we can define the ARMA(p,q)-process we need one more definition:

Definition 3.8 (White noise process) *A stochastic process $\{Z_t, t \in T\}$ is said to be a white noise process with zero mean and variance σ^2, written*

$$\{Z_t\} \sim WN(0, \sigma^2)$$

if and only if $E(Z_t) = 0 \; \forall t \in T$ and the covariance function of $\{Z_t\}$ has the form

$$\gamma(h) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0. \end{cases}$$

With the definition of the white noise process we can now define the ARMA(p,q)-process.

Definition 3.9 (ARMA(p,q)-process) *The process $\{X_t, t \in \mathbb{Z}\}$ is said to be an ARMA(p,q)-process if for every t*

$$X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q} \tag{3.19}$$

or shortly

$$\phi(B) X_t = \theta(B) Z_t$$

where $\{Z_t\} \sim WN(0, \sigma^2)$. $\{X_t\}$ is called an ARMA(p,q)-process with mean μ if $\{X_t - \mu\}$ is an ARMA(p,q)-process.

In the following we only consider ARMA-processes which are *future-independent*, i.e. ARMA-processes with $Cov(X_{t-h}, Z_t) = 0$ for $h > 0$. Especially in applications not much is lost by this restriction since ARMA-processes with $Cov(X_{t-h}, Z_t) \neq 0$ for an $h > 0$ can be regarded as unnatural since in this case the X_t are correlated with white noise perturbances that will happen in the future.
Note that it is not at all evident if all future-independent processes of the form (3.19), i.e all future-independent ARMA-processes, are stationary. We will come back to this question later.

Remark: *Brockwell and Davis (1991) denote as ARMA-processes only those processes of the form (3.19) which are stationary. Evidently, if one follows this definition, all ARMA-processes are stationary (per definition). In this thesis we will use the more general definition stated above which can also be found in Fahrmeir et al. (1981).*

Two major properties an ARMA(p,q)-process may have are *causality* and *invertibility*.

Definition 3.10 (Causality) *An ARMA(p,q)-process defined by the equation (3.19) is said to be causal (or a causal function of $\{Z_t\}$) if there exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t \in \mathbb{Z}. \tag{3.20}$$

Definition 3.11 (Invertibility) *An ARMA(p,q)-process defined by the equation (3.19) is said to be invertible if there exists a sequence of constants* $\{\pi_j\}$ *such that* $\sum_{j=0}^{\infty} |\pi_j| < \infty$ *and*

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t \in \mathbb{Z}. \tag{3.21}$$

Again, the question arises what conditions the ARMA(p,q)-process must fulfill in order to be causal and/or invertible. Since the major focus of this thesis is not a theoretical discussion of time series analysis we will not discuss this question in detail, but refer to Brockwell and Davis (1991), Chapter 3, for further details.
Yet, we state the two following theorems that give necessary and sufficient conditions for an ARMA(p,q)-process to be causal and invertible respectively.

Theorem 3.12 *Suppose* $\{X_t\}$ *to be an ARMA(p,q)-process for which the polynomials* $\phi(\cdot)$ *and* $\theta(\cdot)$ *have no common zeros. Then* $\{X_t\}$ *is causal if and only if all zeros of* $\phi(z)$ *lie outside the unit circle, i.e.* $\phi(z) \neq 0, \forall z \in \mathbb{C}$ *such that* $|z| \leq 1$*. The coefficients* $\{\psi_j\}$ *in (3.20) are determined by the relation*

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| < 1$$

Proof: The proof can be found in Brockwell and Davis (1991), pp. 85.

Theorem 3.13 *Suppose* $\{X_t\}$ *to be an ARMA(p,q)-process for which the polynomials* $\phi(\cdot)$ *and* $\theta(\cdot)$ *have no common zeros. Then* $\{X_t\}$ *is causal if and only if all zeros of* $\theta(z)$ *lie outside the unit circle, i.e.* $\theta(z) \neq 0, \forall z \in \mathbb{C}$ *such that* $|z| \leq 1$*. The coefficients* $\{\pi_j\}$ *in (3.21) are determined by the relation*

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| < 1$$

Proof: The proof can be found in Brockwell and Davis (1991), pp. 87.

Before we continue the discussion of the general ARMA(p,q)-processes we will have a look at the special cases of pure MA(q)- and pure AR(p)-processes.

Definition 3.14 (MA(q)-process) *The process* $\{X_t, t \in \mathbb{Z}\}$ *is said to be an MA(q)-process if for every t*

$$X_t = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q} \tag{3.22}$$

or shortly

$$X_t = \theta(B)Z_t$$

where $\{Z_t\} \sim WN(0, \sigma^2)$.

Since $\phi(z) \equiv 1 \neq 0 \ \forall z \in \mathbb{C}$ the MA(q)-process is invertible if the zeros of $\theta(\cdot)$ lie outside the unit circle. Furthermore it is obvious that every MA(q)-process is causal with $\psi_j = \theta_j$ for $j = 0, ..., q$ with $\theta_0 := 1$ and $\psi_j = 0$ for $j > q$.
It is also quite easy to see that every MA(q)-process is stationary:

From (3.22) it follows that

$$E(X_t) = \sum_{j=0}^{q} \theta_j E(Z_{t-j}) = 0$$

and since for an integer h with $|h| \leq q$

$$\begin{aligned} X_t &= Z_t + \theta_1 Z_{t-1} + ... + \theta_h Z_{t-h} + ... + \theta_q Z_{t-q} \\ X_{t-h} &= Z_{t-h} + ... + \theta_{q-h} Z_{t-q} + ... + \theta_q Z_{t-q-h} \end{aligned}$$

and

$$E(Z_t Z_s) = \begin{cases} \sigma^2 & t = s \\ 0 & \text{otherwise} \end{cases}$$

it follows that with $\theta_0 := 1$

$$\gamma_X(h) = Cov(X_t, X_{t-h}) = E(X_t X_{t-h}) = \begin{cases} \sigma^2(\theta_0\theta_h + \theta_1\theta_{h+1} + ... + \theta_{q-h}\theta_q) & |h| \leq q \\ 0 & \text{otherwise} \end{cases}$$

Note that for $h = 0$

$$\gamma_X(0) = Var(X_t) = \sigma^2 \sum_{j=0}^{\infty} \theta_j^2 < \infty.$$

Thus, a MA(q)-process fulfills all the requirements for (weak) stationarity.

The concept of MA(q)-processes can be extended for $q \to \infty$ in the following way:

Definition 3.15 (MA(∞)) *Let* $\{X_t^{(q)}\}$ *denote an MA(q)-process of the form (3.22). If there exists a sequence* $\{\theta_j\}$ *with* $\sum_{j=0}^{\infty} |\theta_j| < \infty$ *such that*

$$X_t^{(q)} \xrightarrow{m.s.} \sum_{j=0}^{\infty} \theta_j Z_{t-j} := X_t \quad \text{as } q \to \infty \text{ and } \forall t \tag{3.23}$$

where $\xrightarrow{m.s.}$ *denotes mean square convergence,* $\{X_t\}$ *is called a moving average process of infinite order (MA(*∞*)).*

Recall that a series of random variables $X_1, X_2, ...$ is said to converge in mean square to the random variable X, written $X_n \xrightarrow{m.s.} X$ as $n \to \infty$, if

(i) $E(X_n^2) < \infty \; \forall n$

(ii) $\lim_{n\to\infty} E[(X_n - X)^2] = 0$

Also recall that if $X_n \xrightarrow{m.s.} X$ and $Y_n \xrightarrow{m.s.} Y$ then $E(X_n Y_n) \longrightarrow E(XY)$.

With this property of the mean square convergence at hand it is now straightforward to see that an MA(∞)-process $\{X_t\}$ is stationary with

$$E(X_t) = \lim_{q\to\infty} E(X_t^{(q)}) = 0$$

and

$$\gamma_X(h) = \lim_{q\to\infty} E(X_t^{(q)} X_{t+h}^{(q)}) = \sigma^2 \sum_{j=0}^{\infty} \theta_j \theta_{j+h}.$$

We can now draw our attention to pure autoregressive or shortly AR(p)-processes:

Definition 3.16 (AR(p)-process) *The process $\{X_t, t \in \mathbb{Z}\}$ is said to be an AR(p)-process if for every t*

$$X_t - \phi_1 X_{t-1} - ... - \phi_p X_{t-p} = Z_t \tag{3.24}$$

or shortly

$$\phi(B) X_t = Z_t$$

where $\{Z_t\} \sim WN(0, \sigma^2)$. $\{X_t\}$ is called an AR(p)-process with mean μ if $\{X_t - \mu\}$ is an AR(p)-process.

Since for every AR(p)-process $\theta(z) \equiv 1 \neq 0 \; \forall z \in \mathbb{C}$, it is obvious that AR(p)-processes are always invertible with $-\phi_i = \pi_i$, $i = 1, ..., p$. According to Theorem 3.12 an AR(p) process is causal if and only if the zeros of $\phi(\cdot)$ lie outside the unit circle. Yet, the question whether an AR(p)-process is stationary or not still has to be answered. The following theorem specifies the relationship between causality and stationarity of an AR(p)-process:

Theorem 3.17 *For an AR(p)-process $\{X_t\}$ with $\phi(B)X_t = Z_t$ the following assertions are equivalent:*

(i) $\{X_t\}$ is stationary

(ii) The zeros of $\phi(\cdot)$ lie outside the unit circle

(iii) $\{X_t\}$ *is causal*

Proof:
(ii)$\Longrightarrow$*(iii)* follows directly from Theorem 3.13 if one takes into account that $\phi(\cdot)$ and $\theta(\cdot)$ cannot have common zeros since in an AR(p)-process $\theta(z) \equiv 1$.

To see that *(iii)*$\Longrightarrow$*(i)* note that if $\{X_t\}$ is causal, then per definition there exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t \in \mathbb{Z}.$$

This is exactly the representation (3.23) of an MA(∞)-process with $\psi_j = \theta_j$ for $j = 0, 1, 2, ...$ Since MA(∞)-processes are always stationary, it follows that a causal AR(p)-process is stationary.

We still have to show *(i)*$\Longrightarrow$*(ii)*.
First consider a stationary AR(1)-process , i.e. a process $\{X_t\}$ with $X_t = \phi_1 X_{t-1} + Z_t$. Since we only consider processes with $Cov(X_{t-h}, Z_t) = 0$ for $h > 0$ it follows that

$$\gamma_X(0) = Var(X_t) = Var(\phi_1 X_{t-1} + Z_t) = \phi_1^2 Var(X_{t-1}) + \sigma^2 = \phi_1^2 \gamma_X(0) + \sigma^2$$

which is equivalent to

$$\gamma_X(0) = \frac{\sigma^2}{1 - \phi_1^2}.$$

Since $\{X_t\}$ was assumed to be stationary, $0 \leq \gamma_X(0) < \infty$ and it follows that the inequality $-1 < \phi_1 < 1$ must hold.
Now note that if $|\phi_1| < 1 \implies \left|\frac{1}{\phi_1}\right| > 1$ and also note that $\frac{1}{\phi_1}$ is the solution of the equation

$$\phi(z) = 1 - \phi_1 z = 0,$$

which proves that for a stationary AR(1)-process the zero(s) of $\phi(\cdot)$ lie outside the unit circle.

We now consider the general case. For this we consider the equation

$$\phi^{(q)}(z) := 1 - \phi_1 z - ... - \phi_q z^q = 0 \tag{3.25}$$

for an integer $q > 1$.
Let z^* be any solution of this equation.
We assume at this point that $z^* \in \mathbb{R}$, so that

$$\phi^{(q)}(z) = (z - z^*)\phi^{(q-1)}(z) = (1 - \frac{1}{z^*} z^1)(-z^*)\phi^{(q-1)}(z) \tag{3.26}$$

with a polynomial $\phi^{(q-1)}(z) = \phi'_0 + \phi'_1 z + ... + \phi'_{q-1} z^{q-1}$ of degree $q-1$ whose coefficients ϕ'_i, $i = 0, ..., q-1$, do not need to be specified further.

If $z^* \notin \mathbb{R}$ one can at least separate this complex solution together with its conjugate complex, so that $\phi^{(q)}(z) = \phi^{(2)}(z)\phi^{(q-2)}(z)$ where $\phi^{(2)}(z)$ and $\phi^{(q-2)}(z)$ are polynomials of degrees 2 and $q-2$ respectively. In this case one would also have to show the assertion separately for an AR(2)-process so that the following argumentation works. This is straightforward and quite similar to the proof for an AR(1)-process. For details see Fahrmeir et al. (1981), page 220.

The process $Y_t := (-z^*)\phi^{(q-1)}(B)X_t$ is a linear combination of present and past observations of the stationary process X_t and is therefore stationary. With (3.26) the equation $\phi(B)X_t = Z_t$ can be rewritten as:

$$(1 - \frac{1}{z^*}B^1)Y_t = Z_t, \tag{3.27}$$

i.e. Y_t is a stationary AR(1)-process. We have already shown that for an AR(1)-process $\{\tilde{X}_t\}$ of the form $(1 - \phi_1 B^1)\tilde{X}_t = Z_t$ stationarity implies that $|\phi_1| < 1$. Thus, it follows that for the stationary AR(1)-process given by (3.27) $|z^*| > 1$ must hold.

Recalling that z^* was any solution of the equation (3.25) the proof is completed.

□

We finally come back to the question whether an ARMA(p,q)-process is stationary. Since as previously shown all MA(q)-processes are stationary, it is perfectly plausible that this question can be answered by the AR(p) part of the ARMA(p,q)-process alone, i.e. an ARMA(p,q)-process is stationary if and only if its AR(p) part is stationary.
Moreover, since all AR(p)-processes are invertible, an ARMA(p,q)-process is invertible if and only if its MA(q) part is invertible.

We conclude the brief introduction to time series analysis with a discussion of one particular process which we will need in the following chapters: The Gaussian AR(1)-process.

Definition 3.18 (Zero-mean Gaussian AR(1)-process) *A time series $\{u_t, t \in \mathbb{Z}\}$ is called (zero-mean) Gaussian AR(1) process if for every t*

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad \textit{with} \qquad \varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2).$$

The process $\{u_t, t \in \mathbb{Z}\}$ is called Gaussian AR(1) process with mean μ if $\{u_t - \mu, t \in \mathbb{Z}\}$ is a zero-mean Gaussian AR(1)-process.

Since

$$u_t = \rho u_{t-1} + \varepsilon_t = \rho \cdot (\rho u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \rho^2 u_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t = \ldots = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}$$

it follows that

$$E(u_t) = \sum_{j=0}^{\infty} \rho^j \cdot E(\varepsilon_{t-j}) = 0, \tag{3.28}$$

$$\gamma_u(0) := Var(u_t) = E(u_t^2) = E\left[\left(\sum_{j=0}^{\infty} \rho^j \cdot \varepsilon_{t-j}\right)^2\right] = \sum_{j=0}^{\infty} (\rho^2)^j \cdot E(\varepsilon_{t-j}^2) = \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{3.29}$$

and

$$\begin{aligned} Cov(u_t, u_{t-h}) &= E(u_t u_{t-h}) \\ &= E\left(\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i} \cdot \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-h-j}\right) \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \rho^i \rho^j E(\varepsilon_{t-i} \cdot \varepsilon_{t-h-j}). \end{aligned}$$

Since

$$E(\varepsilon_{t-i} \cdot \varepsilon_{t-h-j}) = \begin{cases} \sigma_\varepsilon^2 & i = h + j \\ 0 & otherwise \end{cases}$$

it follows that

$$\begin{aligned} Cov(u_t, u_{t-h}) &= \sum_{j=0}^{\infty} \rho^{h+j} \rho^j \sigma_\varepsilon^2 \\ &= \rho^h \cdot \sigma_\varepsilon^2 \cdot \sum_{j=0}^{\infty} (\rho^j)^2 \\ &= \rho^h \cdot \sigma_\varepsilon^2 \cdot \frac{1}{1-\rho^2} \\ &= \rho^h \cdot \gamma_u(0). \end{aligned} \tag{3.30}$$

This means that the autocovariance function $\gamma_u(h)$ and the autocorrelation function $\rho_u(h)$ for a Gaussian AR(1)-process have the form:

$$\gamma_u(h) = \rho^h \cdot \gamma_u(0) = \rho^h \cdot \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{3.31}$$

$$\rho_u(h) = \frac{\gamma_u(h)}{\gamma_u(0)} = \frac{\rho^h \cdot \gamma_u(0)}{\gamma_u(0)} = \rho^h \tag{3.32}$$

3.3 Markov Chain Monte Carlo methods

In this section we introduce Markov Chain Monte Carlo (MCMC) methods which can be used for parameter estimation in complex statistical models where other estimation methods such as maximum-likelihood estimation are not feasible. We will use MCMC methods within the WinBUGS software later for the modelling of our option data.
In order to understand how MCMC parameter estimation works it is necessary to have the basic concepts concerning both Bayesian Inference and Markov Chains at hand. In the first paragraphs of this section we give a short overview of the most important concepts related to these two topics so that we can discuss MCMC estimation techniques in the following. A detailed introduction to MCMC methods in general can be found in Gamerman (1997) or Gilks et al. (1996). For a detailed discussion of Bayesian Inference and of Markov Chain theory see for example Tanner (1996) and Resnick (1992) respectively.

3.3.1 Bayesian Inference

In the classical approach parameters in a statistical model are fixed unknown quantities, whereas in Bayesian Inference they are considered to be random variables.
Let $\boldsymbol{\theta}$ be the parameter vector of a statistical model and let $p(\boldsymbol{\theta})$ be its density or probability function that summarizes the knowledge about $\boldsymbol{\theta}$, obtained e.g. by previous experiments, experts' assumptions, etc. $p(\boldsymbol{\theta})$ is commonly called the *prior-distribution* of $\boldsymbol{\theta}$.
Then, given a random sample $\mathbf{x} = (x_1, ..., x_T)^t$ with joint density or probability function $f(\mathbf{x}|\boldsymbol{\theta})$, one can update the knowledge about $\boldsymbol{\theta}$ by using the new information contained in $\mathbf{x}$ and calculating the conditional density of $\boldsymbol{\theta}$ given $\mathbf{x}$ denoted by $p(\boldsymbol{\theta}|\mathbf{x})$. The density or probability function $p(\boldsymbol{\theta}|\mathbf{x})$ is called the *posterior-distribution* of $\boldsymbol{\theta}$.
From Bayes' theorem it follows that $p(\boldsymbol{\theta}|\mathbf{x})$ can be calculated in the following way:

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{f(\mathbf{x})} \propto f(\mathbf{x}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}) \tag{3.33}$$

where

$$f(\mathbf{x}) = \int_{-\infty}^{\infty} f(\mathbf{x}, \boldsymbol{\theta}) \quad d\boldsymbol{\theta} = \int_{-\infty}^{\infty} f(\mathbf{x}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}) \quad d\boldsymbol{\theta}$$

is the marginal distribution of $\mathbf{x}$.

Note that in many complex models it is not possible to get an analytic closed form for the posterior $p(\boldsymbol{\theta}|\mathbf{x})$ since the calculation of $f(\mathbf{x})$ might require high-dimensional integration which is in many cases infeasible. However, since

$$posterior \propto likelihood \cdot prior$$

$p(\boldsymbol{\theta}|\mathbf{x})$ can always be specified up to a constant of proportionality.
Generally speaking, MCMC methods provide tools for approximate sampling from the posterior distribution of the parameters in complex models where this is analytically and numerically not tractable otherwise.

3.3.2 Markov Chains with discrete state space

Let $\{\theta^{(t)}, t \in T\}$ be a stochastic process with a discrete state space S.

Definition 3.19 (Markov Chain) *The process $\{\theta^{(t)}, t \in T\}$ is called a Markov Chain if given the present, past and future states are independent, i.e.*

$$P(\theta^{(t+1)} = x_{t+1}|\theta^{(t)} = x_t, \theta^{(t-1)} = x_{t-1}, ..., \theta^{(1)} = x_1) = P(\theta_{(t+1)} = x_{t+1}|\theta^{(t)} = x_t) \tag{3.34}$$
$$\forall \quad x_1, ..., x_{t+1} \in S.$$

A Markov Chain is said to be homogeneous if the probabilities in (3.34) do not depend on t.

For a homogeneous Markov Chain $\{\theta^{(t)}, t \in T\}$ with discrete state space $S = (x_1, ..., x_n)$ define the *transition probability* $P(x_i, x_j)$ as

$$P(x_i, x_j) := P(\theta^{(t+1)} = x_j|\theta^{(t)} = x_i)$$

and the *transition matrix*

$$\mathbf{P} = \begin{pmatrix} P(x_1, x_1) & \cdots & P(x_1, x_n) \\ \vdots & \ddots & \vdots \\ P(x_n, x_1) & \cdots & P(x_n, x_n) \end{pmatrix}.$$

Moreover, define the *t-step transition probability* for the Markov Chain by

$$P(\theta^{(t)} = x_j|\theta^{(0)} = x_i).$$

Proposition 3.20 (Higher-order transition prob.) *For all $s \geq 0$ and x_i, x_j in the state space*

$$P(\theta^{(s)} = x_j|\theta^{(0)} = x_i) = p_{ij}^{(s)} \tag{3.35}$$

where $(p_{ij}^{(s)}) = \mathbf{P}^s$ and $\mathbf{P}^s = \underbrace{P \cdot ... \cdot P}_{s \text{ times}}$ with $\cdot$ denoting usual matrix multiplication.

Proof: The formula is obviously true for $s = 0, 1$. Now suppose that (3.35) is also true for s=0,1,...,S. Then,

$$\begin{aligned} P(\theta^{(S+1)} = x_j | \theta^{(0)} = x_i) &= \sum_{k=1}^{n} \frac{P(\theta^{S+1} = x_j, \theta^{(1)} = x_k, \theta^{(0)} = x_i)}{P(\theta^{(0)} = x_i)} \\ &= \sum_{k=1}^{n} \frac{P(\theta^{(S+1)} = x_j | \theta^{(1)} = x_k, \theta^{(0)} = x_i) \cdot P(\theta^{(1)} = x_k, \theta^{(0)} = x_i)}{P(\theta^{(0)} = x_i)} \end{aligned}$$

Since $\{\theta^{(t)}, t \in T\}$ is a homogeneous Markov Chain this is equal to

$$\sum_{k=1}^{n} P(\theta^{(S)} = x_j | \theta^{(0)} = x_k) \cdot P(\theta^{(1)} = x_k | \theta^{(0)} = x_i),$$

which by the induction hypothesis yields

$$\sum_{k=1}^{n} p_{ik}^{(1)} \cdot p_{kj}^{(S)} = p_{ij}^{(S+1)}$$

as required.

□

The component form

$$p_{ij}^{(s+t)} = \sum_{k=1}^{n} p_{ik}^{(s)} \cdot p_{kj}^{(t)}$$

of the matrix identity

$$\mathbf{P}^{(s+t)} = \mathbf{P}^s \cdot \mathbf{P}^t$$

is usually called *Chapman-Kolmogorov equation*.

Now define for $x, y \in S$:

$$\begin{aligned} T_y &:= inf\{t \geq 1 : \theta^{(t)} = y\} \\ \rho_{xy} &:= P(T_y < \infty | \theta^{(0)} = x) \\ m_y &:= E(T_y | \theta^{(0)} = y) \\ d_y &:= \text{largest common divisor of } \{t \geq 1 : P(\theta^{(t)} = y | \theta^{(0)} = y)\} \end{aligned}$$

Then, the states of a Markov Chain with discrete state space can be classified in the following way: A state $y \in S$ is said to be

$$\text{recurrent} \quad :\Longleftrightarrow \quad \rho_{yy} = 1$$

$$\begin{aligned} \text{positive recurrent} &:\Longleftrightarrow y \quad \text{recurrent and } m_y < \infty \\ \text{null recurrent} &:\Longleftrightarrow y \quad \text{recurrent and } m_y = \infty \\ \text{transient} &:\Longleftrightarrow \rho_{yy} < 1 \\ \text{absorbing} &:\Longleftrightarrow P(y,y) = 1 \\ \text{aperiodic} &:\Longleftrightarrow d_y = 1 \end{aligned} \tag{3.36}$$

A state y is said to be accessible from another state x, denoted by $x \to y :\Longleftrightarrow \rho_{xy} \geq 0$. The Markov Chain is called *irreducible* if $x \to y$ and $y \to x$ for every pair $x, y \in S$. Furthermore, it is called (positive) recurrent/aperiodic if all states are (positive) recurrent/aperiodic. The phrase *positive recurrent, aperiodic and irreducible* is commonly subsumed under the name *ergodic*.

An important question concerning Markov Chains is the existence of a stationary equilibrium or invariant distribution π which satisfies

$$\sum_{x \in S} \pi(x) P(x,y) = \pi(y) \quad \forall y \in S \tag{3.37}$$

or in matrix notation

$$\pi^t = \pi^t \cdot \mathbf{P} \tag{3.38}$$

Once the chain reaches a stage where a distribution π satisfying (3.37) is the distribution of the chain, the chain retains this distribution for all subsequent steps.
The following proposition states conditions for the existence and uniqueness of a stationary distribution.

Proposition 3.21 (Existence of a stationary distribution) *If a Markov Chain with discrete state space is positive recurrent and irreducible, there exists a unique stationary distribution π as defined in (3.37) with*

$$\pi(y) = \frac{1}{m_y} \quad \forall y \in S. \tag{3.39}$$

Proof: The proof can be found in Resnick (1992), page 121.

The following theorem states the relationship between the stationary distribution and the limiting distribution of a Markov Chain with discrete state space which will be a key concept for the discussion of MCMC methods in the following paragraphs.

Theorem 3.22 (Stationary and limiting distribution) *Suppose a Markov Chain with discrete state space is irreducible and aperiodic and that a stationary solution π as defined in (3.37) exists.*
Then,

(i) The Markov Chain is positive recurrent.

(ii) π is a limit distribution, i.e $\lim_{t\to\infty} P(\theta^{(t)} = y|\theta^{(0)} = x) = \pi(y) \quad \forall x, y \in S.$

(iii) For all $y \in S$: $\pi(y) > 0$.

(iv) The stationary distribution is unique.

Proof: The proof can be found in Resnick (1992), pp 128.

Note that this theorem also provides a useful practical test for an irreducible, aperiodic chain to be positive recurrent. This is the case if a stationary distribution exists which can be calculated by solving the linear equation system (3.38).

3.3.3 Markov Chains with continuous state space

So far we have only considered Markov Chains with a discrete state space. If the state space S of a homogeneous Markov Chain is continuous *transition distributions* defined by

$$P(x, y) := P(\theta^{(t+1)} \leq y|\theta^{(t)} = x) \quad \forall x, y \in S.$$

can be considered. The corresponding density

$$p(x, y) = \frac{\partial P(x, y)}{\partial y}$$

is called *transition kernel*.
Let $\pi^{(t)}$ denote the marginal density at any step $t \geq 0$, then $\forall y \in S$, $\pi^{(t)}(y)$ can be obtained from the previous step by

$$\pi^{(t)}(y) = \int_{-\infty}^{\infty} p(x, y)\pi^{(t-1)}(x) \quad dx$$

A stationary distribution π of the Markov Chain must, in the case of a continuous state space, fulfill the following condition for all $y \in S$:

$$\pi(y) = \int_{-\infty}^{\infty} p(x, y)\pi(x) \quad dx \tag{3.40}$$

In order to study convergence and limiting results consider a set $A \subset S$, the hitting time T_A, a distribution ν and define

$$\rho_{xA} := P(T_A < \infty | \theta^{(0)} = x).$$

The Markov Chain is irreducible if there exists at least one distribution ν that ensures

$$\rho_{xA} > 0 \quad \text{if} \quad \nu(A) > 0 \quad \forall x \in S.$$

The definitions for aperiodicity and positive recurrence are similar to those for Markov Chains with discrete state space in (3.36) if y is replaced by sets A with $\nu(A) > 0$.
With these definitions all previously stated convergence results for the discrete case remain valid in the continuous case. Similar to the discrete case, a positive recurrent, aperiodic and irreducible Markov Chain with continuous state space is called *ergodic*.

Now let $\{\boldsymbol{\theta}^{(t)}, t \in T\}$ be an ergodic (vector valued) Markov Chain with state space $S \in \mathbb{R}^d$, transition kernel $p(x, y)$ and initial distribution $\pi^{(0)}$. To generate values from this Markov Chain, sample $\boldsymbol{\theta}^{(0)}$ from $\pi^{(0)}$ and then, in the t-th step, $\boldsymbol{\theta}^{(t)}$ from $p(\boldsymbol{\theta}^{(t-1)}, \cdot)$. As t gets large the sampled values will have a distribution approximately similar to the limiting distribution π of the Markov Chain. Thus, by generating values of the Markov Chain one can obtain an approximate sample from π. This approach can be used to obtain approximate samples from a complex posterior distribution $p(\boldsymbol{\theta}|\mathbf{x})$ in Bayesian Inference by assuming that the posterior distribution is the stationary/limiting distribution of a Markov Chain. This general approach is called *Markov Chain Monte Carlo method*.

Now let $f : S \mapsto \mathbb{R}^d$ be a non-negative or bounded function with domain S. Then, the following theorem, which is a Markov Chain version of the law of large numbers, holds:

Theorem 3.23 (LLN for Markov Chains) *If a Markov Chain $\{\boldsymbol{\theta}^{(t)}, t \in T\}$ is ergodic and $E_\pi(f(\boldsymbol{\theta})) < \infty$ for the unique limiting distribution π and a non-negative or bounded function f, then*

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} f(\boldsymbol{\theta}^{(t)}) = E_\pi(f(\boldsymbol{\theta}))$$

almost surely.

Proof: The proof can be found in Resnick (1992), page 123.

This theorem can be used in order to justify $\frac{1}{T} \sum_{t=1}^{T} f(\boldsymbol{\theta}^{(t)})$ as an estimate for $E_\pi(f(\boldsymbol{\theta}))$.

3.3.4 The Gibbs Sampler and the Metropolis-Hastings Algorithm

A widely used MCMC sampling scheme is the *Gibbs sampler* which can be used for constructing a Markov Chain with pre-specified limiting distribution π when the full conditionals

$$\pi_i(\theta_i) := \pi(\theta_i|\boldsymbol{\theta}_{-i}) \quad \text{where} \quad \boldsymbol{\theta}_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_d)^t$$

are known and can be sampled from. A detailed discussion of the Gibbs sampler together with its historical development can be found for example in Gamerman (1997).
The Gibbs sampling algorithm works as follows:

(i) Choose initial values $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, ..., \theta_d^{(0)})^t$

(ii) Obtain a new value $\boldsymbol{\theta}^{(j)}$ through successive generation of values

$$\begin{aligned}
\theta_1^{(j)} &\sim \pi(\theta_1|\theta_2^{(j-1)}, ..., \theta_d^{(j-1)}) \\
\theta_2^{(j)} &\sim \pi(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)} ..., \theta_d^{(j-1)}) \\
&\vdots \\
\theta_d^{(j)} &\sim \pi(\theta_d|\theta_1^{(j)}, ..., \theta_{d-1}^{(j)})
\end{aligned}$$

(iii) Change the counter j to $j+1$ and return to the previous step until convergence of the Markov Chain is reached.

When convergence is reached $\boldsymbol{\theta}^{(j)}$ and all subsequent values $\boldsymbol{\theta}^{(j+k)}$ for $k \geq 1$ can be considered realizations of π.
Obviously, the Gibbs sampler defines a homogeneous Markov Chain since in the update step the transition probabilities only depend on the sampled values of the previous step. The transition kernel from $\boldsymbol{\theta}$ to $\boldsymbol{\phi}$ is given by:

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}) = \prod_{i=1}^{d} \pi(\phi_i|\phi_1, ..., \phi_{i-1}, \theta_{i+1}, ..., \theta_d).$$

It still has to be shown that π is the limiting distribution of the Markov Chain generated by the Gibbs sampler.
If a Markov Chain with transition kernel $p(\boldsymbol{\theta}, \boldsymbol{\phi})$ has a (unique) limiting distribution π^∞, the chain must be irreducible and the stationarity condition (3.40) must be fulfilled by π^∞ since a limiting distribution is always a stationary distribution.
Irreducibility must be verified by checking that $P(\boldsymbol{\theta}, A) > 0$ for all sets A with a positive

posterior probability.
Now suppose that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^t, \boldsymbol{\theta}_2^t)^t$ and that $\pi^\infty(\boldsymbol{\theta}_1)$ and $\pi^\infty(\boldsymbol{\theta}_2)$ are the marginal limiting densities for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ respectively. The limiting densities for $\boldsymbol{\theta}_i$ conditional on $\boldsymbol{\theta}_j$ are denoted by $\pi^\infty(\boldsymbol{\theta}_i|\boldsymbol{\theta}_j)$ for $i, j \in \{1, 2\}$, $i \neq j$. The transition kernel is given by

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}) = \pi(\boldsymbol{\phi}_1|\boldsymbol{\theta}_2) \cdot \pi(\boldsymbol{\phi}_2|\boldsymbol{\phi}_1).$$

Since

$$\int\int \pi(\boldsymbol{\phi}_2|\boldsymbol{\phi}_1) \cdot \pi^\infty(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 d\boldsymbol{\phi}_2 = \int \pi(\boldsymbol{\phi}_2|\boldsymbol{\phi}_1) \cdot \underbrace{\left(\int \pi^\infty(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 \right)}_{=1} d\boldsymbol{\phi}_2 = 1$$

π^∞ can be rewritten as:

$$\begin{aligned} \pi^\infty(\boldsymbol{\theta}_2) &= \int\int \pi(\boldsymbol{\phi}_2|\boldsymbol{\phi}_1) \cdot \pi^\infty(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) \cdot \pi^\infty(\boldsymbol{\theta}_2) \quad d\boldsymbol{\theta}_1 d\boldsymbol{\phi}_2 \\ &= \int\int \pi(\boldsymbol{\phi}_2|\boldsymbol{\phi}_1) \cdot \pi^\infty(\boldsymbol{\theta}) \quad d\boldsymbol{\theta}_1 d\boldsymbol{\phi}_2 \end{aligned} \tag{3.41}$$

Thus,

$$\begin{aligned} \pi^\infty(\boldsymbol{\phi}_1) &= \int \pi^\infty(\boldsymbol{\phi}) \quad d\boldsymbol{\phi}_2 \\ &= \int \left(\int\int p(\boldsymbol{\theta}, \boldsymbol{\phi}) \cdot \pi^\infty(\boldsymbol{\theta}) \quad d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 \right) d\boldsymbol{\phi}_2 \\ &= \int\int\int \pi(\boldsymbol{\phi}_1|\boldsymbol{\theta}_2) \cdot \pi(\boldsymbol{\phi}_2|\boldsymbol{\phi}_1) \cdot \pi^\infty(\boldsymbol{\theta}) \quad d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 d\boldsymbol{\phi}_2 \\ &\overset{(3.41)}{=} \int \pi(\boldsymbol{\phi}_1|\boldsymbol{\theta}_2) \cdot \pi^\infty(\boldsymbol{\theta}_2) \quad d\boldsymbol{\theta}_2. \end{aligned} \tag{3.42}$$

The only conditional distribution $\pi(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2)$ which can satisfy (3.42) is the limiting conditional distribution $\pi^\infty(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2)$, i.e. $\pi(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2) = \pi^\infty(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2)$ must hold.
Thus, the limiting distribution of the Markov Chain generated by the Gibbs sampler must have the same full conditionals as the posterior distribution π. Since under "mild conditions" (for details see Gamerman (1997)) the set of all full conditionals determine the joint distribution, the Markov Chain converges to the posterior π and the Gibbs sampler draws values from the posterior distribution after convergence.

Another possibility to construct a Markov Chain with a given stationary/limiting distribution π is the *Metropolis-Hastings algorithm*, which we also introduce shortly. This algorithm is based on research papers by Metropolis et al. (1953) and Hastings (1970) and again discussed in detail in Gamerman (1997).

Consider a Markov Chain where the transition kernel $p(\boldsymbol{\theta}, \boldsymbol{\phi})$ consists of an arbitrary kernel $q(\boldsymbol{\theta}, \boldsymbol{\phi})$ and an acceptance probability $\alpha(\boldsymbol{\theta}, \boldsymbol{\phi})$ such that for $\theta, \boldsymbol{\phi} \in S$

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}) = q(\boldsymbol{\theta}, \boldsymbol{\phi}) \cdot \alpha(\boldsymbol{\theta}, \boldsymbol{\phi}) \quad \text{for} \quad \boldsymbol{\theta} \neq \boldsymbol{\phi} \tag{3.43}$$

and

$$p(\boldsymbol{\theta}, \boldsymbol{\theta}) = 1 - \int q(\boldsymbol{\theta}, \boldsymbol{\phi}) \cdot \alpha(\boldsymbol{\theta}, \boldsymbol{\phi}) \quad d\boldsymbol{\phi}.$$

The acceptance probability $\alpha(\boldsymbol{\theta}, \boldsymbol{\phi})$ is given by:

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\phi}) = min \left\{ 1, \frac{\pi(\boldsymbol{\phi}) \cdot q(\boldsymbol{\phi}, \boldsymbol{\theta})}{\pi(\boldsymbol{\theta}) \cdot q(\boldsymbol{\theta}, \boldsymbol{\phi})} \right\} \tag{3.44}$$

It is straightforward to see that the transition kernel defined by (3.43) and (3.44) satisfies

$$\pi(\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}, \boldsymbol{\phi}) = \pi(\boldsymbol{\phi}) \cdot p(\boldsymbol{\phi}, \boldsymbol{\theta}) \tag{3.45}$$

which is usually called the *reversibility condition*.
If a distribution/density π satisfies (3.45) for an irreducible chain, then the chain is pos. recurrent and since

$$\int \pi(\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}, \boldsymbol{\phi}) \quad d\boldsymbol{\theta} \overset{(3.45)}{=} \int \pi(\boldsymbol{\phi}) \cdot p(\boldsymbol{\phi}, \boldsymbol{\theta}) \quad d\theta = \pi(\boldsymbol{\phi})$$

π also satisfies the stationarity condition (3.40). It is therefore the stationary and limiting distribution of the Markov Chain.

Note that it is not necessary to know the normalizing constant of the posterior distribution π in order to implement the Metropolis-Hastings algorithm which can be formulated in the following way:

(i) Choose an arbitrary initial value $\boldsymbol{\theta}^{(0)}$

(ii) In the j-th step move the chain to a new value $\boldsymbol{\phi}$ generated from $q(\boldsymbol{\theta}^{(j-1)}, \cdot)$

(iii) Evaluate the acceptance probability $\alpha(\boldsymbol{\theta}^{(j-1)}, \boldsymbol{\phi})$ given by (3.44)

(iv) If the move is accepted then $\boldsymbol{\theta}^{(j)} = \boldsymbol{\phi}$. If the move is not accepted then set $\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^{(j-1)}$

(v) Return to the second step until convergence is reached.

For both the Gibbs sampler and the Metropolis-Hastings algorithm it is crucial to know when convergence is reached so that the values of the generated Markov Chain can be considered as realizations of the posterior distribution. We discuss some of the convergence diagnostics methods proposed in the literature in the following paragraph.

3.3.5 Convergence diagnostics

One of the most important implementation issues concerning all MCMC techniques is the question of how many initial replications have to be discarded in order to avoid the possibility of bias in the ergodic average estimation caused by the effect of the starting values. This initial period, from the first iteration up to the point of time when the chain is assumed to have reached convergence is commonly called *burn-in*. Unfortunately, for almost all MCMC settings a priori theoretical techniques to determine the rate of convergence of the chain, the necessary burn-in and the required run length to obtain a pre-specified estimation precision are not available. Therefore, it is indispensable to carry out some form of output-based statistical analysis to assess convergence.
In addition to the most widely used informal methods we briefly introduce three of the more elaborate methods proposed in the literature for which S-Plus language code is available within the CODA package. For a more extensive overview of the convergence diagnostics methods proposed in the literature see Brooks and Roberts (1998) and Cowles and Carlin (1996).

Informal methods

Maybe the easiest way to check convergence is the so-called "thick pen method" which indicates convergence of a single chain based on a graphical analysis of the trajectory of the chain. A trajectory of the chain exhibiting the same qualitative behaviour through iterations after an initial burn-in is an indication of convergence. In practice this means that if after the initial burn-in all subsequent values lie within a certain band, the chain is likely to have converged.
If one runs m parallel chains one can plot a histogram of the m values of the n-th iterates and compare this histogram to the one of the m values of the (n+k)-th iterates for a large enough k, typically between 10 and 50. Convergence is then accepted if the histograms can not be distinguished.

The Gelman-Rubin-statistic

This method was originally developed by Gelman and Rubin (1992) and reconsidered in a simplified version by Gamerman (1997) and by Gilks et al. (1996). To monitor convergence according to this method one needs to run $m \geq 2$ parallel chains of length $2n$ with different starting values which are clearly overdispersed with respect to the posterior distribution. The first n iterates are discarded and only the last n retained for further analysis. Then,

if one wants to estimate $\psi = f(\theta)$ calculate

$$B := \frac{n}{m-1} \sum_{i=1}^{m} (\bar{\psi}_i - \bar{\psi})^2$$

where

$$\bar{\psi}_i := \frac{1}{n} \sum_{t=n+1}^{2n} \psi_i^{(t)}$$

and

$$\bar{\psi} := \frac{1}{m} \sum_{i=1}^{m} \bar{\psi}_i \qquad .$$

$\frac{B}{n}$ is the empirical variance between the m sequence-means, denoted by $\bar{\psi}_i$ respectively. Furthermore, calculate the mean of the m within-sequence variances

$$W := \frac{1}{m} \sum_{i=1}^{m} s_i^2$$

where

$$s_i^2 := \frac{1}{n-1} \sum_{t=n+1}^{2n} (\psi_i^{(t)} - \bar{\psi}_i)^2.$$

Now define

$$\hat{\sigma}_\psi^2 := \left(1 - \frac{1}{n}\right) W + \frac{1}{n} B$$

and note that under the assumption that all $n \cdot m$ values of ψ were drawn from the posterior distribution (i.e. under the assumption of convergence) and defining $\sigma_\psi^2 := Var(\psi)$ both W and $\hat{\sigma}_\psi^2$ are unbiased estimators for σ_ψ^2. To check this assertion note that

$$E(W) = \frac{1}{m} \sum_{i=1}^{m} E(s_i^2) = \sigma_\psi^2$$

and

$$\begin{aligned} E(\hat{\sigma}_\psi^2) &= \sigma_\psi^2 - \frac{1}{n} \cdot \sigma_\psi^2 + E\left[\frac{1}{m-1} \sum_{i=1}^{m} (\bar{\psi}_i - \bar{\psi})^2\right] \\ &= \sigma_\psi^2 - \frac{1}{n} \cdot \sigma_\psi^2 + Var(\bar{\psi}_i) \\ &= \sigma_\psi^2 - \frac{1}{n} \cdot \sigma_\psi^2 + \frac{1}{n} \cdot Var(\psi) \\ &= \sigma_\psi^2. \end{aligned}$$

If, however, the chains have not yet converged, the overdispersed starting values will still have a significant influence on the trajectories and force $\hat{\sigma}_\psi^2$ to overestimate σ_ψ^2. Meanwhile,

the mean of the within-sequence variances W will tend to underestimate σ_ψ^2 because each chain will not have adequately traversed the complete parameter space and will therefore have less variability. The convergence of the chain can now be monitored by estimating the "potential scale reduction factor"

$$\hat{R} := \sqrt{\frac{\hat{\sigma}_\psi^2}{W}} \qquad .$$

As the simulation converges the estimated potential scale reduction factor R declines to 1, which means that the parallel chains are essentially overlapping. Gilks et al. (1996) consider a value of $\hat{R} = 1.1$ or 1.2 as small enough to indicate convergence.

Geweke's Time series diagnostic method

Functionals of positive recurrent Markov Chains are a special case of stationary time series. So, time series methods can be used to assess convergence.
Consider again $\psi = f(\boldsymbol{\theta})$ which we want to estimate. The trajectory of the simulated Markov Chain $\psi^{(1)}, \psi^{(2)}, ...$ defines a time series. Geweke (1992) suggests that if the chain has converged to its stationary distribution, the empirical first and second order moments of two subsequences of the chain should behave similarly. Recall that this was the definition of weak stationarity of a time series given in Section 3.2 . Now consider a simulation with a burn-in period of m and a total of $m + n$ iterations. Choose two subsequences $\{\psi^{(t)} : t = m + 1, ..., m + n_a\}$ and $\{\psi^{(t)} : t = m + n - n_b + 1, ..., m + n\}$ and define

$$\bar{\psi_a} := \frac{1}{n_a} \sum_{j=m+1}^{m+n_a} \psi^{(j)}$$

$$\bar{\psi_b} := \frac{1}{n_b} \sum_{j=m+n-n_a+1}^{m+n} \psi^{(j)}$$

As n gets large and the ratios $\frac{n_a}{n}$ and $\frac{n_b}{n}$ remain fixed, the standardized difference $Z_n(m)$ defined by

$$Z_n(m) := \frac{\bar{\psi_a} - \bar{\psi_b}}{\sqrt{\widehat{Var}(\bar{\psi_a}) + \widehat{Var}(\bar{\psi_b})}} \xrightarrow{d} N(0, 1) \tag{3.46}$$

as $n \to \infty$ as a consequence of the Central Limit Theorem. In (3.46)

$$\begin{aligned} \widehat{Var}(\bar{\psi_a}) &= \frac{1}{n_a} \hat{S}^a(0) \\ \widehat{Var}(\bar{\psi_b}) &= \frac{1}{n_b} \hat{S}^b(0) \end{aligned}$$

where $\hat{S}^a(0)$ and $\hat{S}^b(0)$ denote spectral sensitivity estimates for $\{\psi^{(t)} : t = m+1, ..., m+n_a\}$ and $\{\psi^{(t)} : t = m + n - n_b + 1, ..., m + n\}$ respectively. For further details see Geweke (1992).
Large absolute differences $|Z_n(m)|$ indicate lack of convergence and consequently, the burn-in m should be increased. For a fixed n and for fixed ratios $\frac{n_a}{n}$ and $\frac{n_b}{n}$ (Geweke (1992) suggests values of 0.1 and 0.5 for the two ratios respectively) $Z_n(m)$ can be plotted against the number of discarded iterations m in order to get a graphical indication of the length of the required burn-in period.
A drawback of this method is the fact that convergence is also indicated if the chain is merely trapped in a subregion of the parameter space and has not converged yet.

The Raftery and Lewis convergence rate estimator

The last convergence diagnostics tool we state here (without elaborating all the details) is the method proposed by Raftery and Lewis (1992).
Suppose one wants to estimate a particular quantile of the posterior distribution of $\psi = f(\theta)$, i.e. one wants to estimate u such that $P(\psi \leq u) = q$ for some pre-described p, typically 0.025, so that given the estimate $\hat{u}$, $P(\psi \leq \hat{u})$ lies within ±r of the true value with a certain probability p.
Raftery and Lewis (1992) suggest to run a MCMC simulation for an initial burn-in of m iterations and to record a further n iterations which are thinned by storing only every s-th iteration. They also provide a method to determine the values of m, n and s. The basic idea is to define

$$Z^{(t)} := I_{(\psi^{(t)} \leq u)}$$

where I denotes the indicator function. The process $Z^{(t)}$ is a binary process derived from a Markov Chain, but will in general not satisfy the Markov property (3.34). However, one can expect the dependencies between iterations to fall off quickly and it is therefore reasonable to approximate the process $\{Z_s^{(t)} := Z^{1+(t-1)s}\}$ for a large enough s by a Markov Chain. For details on how to select s see Raftery and Lewis (1992). The transition matrix $\mathbf{P}$ of $\{Z_s^{(t)}\}$ is estimated by

$$\hat{p}_{ij} = \frac{\sharp\{t : Z_s^{(t)} = j, Z_s^{(t-1)} = i\}}{\sharp\{t : Z_s^{(t-1)} = i\}} \quad i, j = 0, 1. \tag{3.47}$$

The limiting distribution of a binary Markov Chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

is $\pi = \frac{1}{p+q}\begin{pmatrix} q \\ p \end{pmatrix}$ and the n-step transition matrix is given by (see Resnick (1992), pp. 75):

$$\mathbf{P}^n = \frac{(1-p-q)^n}{p+q}\begin{pmatrix} p & -p \\ -q & q \end{pmatrix} + \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix}$$

The burn-in m^* should be chosen in such a way that after m^* iterations the marginal distribution of $Z_s^{(m^*)}$ is close to the limiting distribution, and since

$$\begin{aligned} |P(Z_s^{(m^*)} = 1 | Z_s^{(0)} = j) - \pi_1| &= \frac{|1-p-q|^{m^*} \cdot p}{p+q} \quad j = 0, 1 \\ |P(Z_s^{(m^*)} = 0 | Z_s^{(0)} = j) - \pi_0| &= \frac{|1-p-q|^{m^*} \cdot q}{p+q} \quad j = 0, 1 \end{aligned}$$

m^* has to satisfy:

$$\frac{|1-p-q|^{m^*} \cdot max(p,q)}{p+q} < \varepsilon$$

for a small enough ε, e.g. $\varepsilon = 0.01$, which is equivalent to

$$m^* \geq \frac{log\left(\frac{\varepsilon \cdot (p+q)}{max(p,q)}\right)}{log(|1-p-q|)}$$

where $p = \hat{p}_{01}$ and $q = \hat{p}_{10}$ estimated by (3.47).

Although this convergence diagnostics method is easy to implement it has to be pointed out that it relies upon three levels of approximation: The Markov approximation, the choice of s and the estimation of $\hat{\mathbf{P}}$ in (3.47).

Since all of the previously discussed methods to assess convergence of a MCMC simulation are based on some form of approximation and can therefore only indicate convergence without mathematical accuracy one should not entirely rely on a single method, but always consider several of the previously introduced techniques.

Chapter 4

Regression models for time series of counts

If the variable of interest in a count data regression model comes from a time series it is most unlikely that neighbouring observations are independent which is assumed in an ordinary GLM. In the literature two classes of models for time-dependent count data have been distinguished (see Cox (1993) or Zeger (1988)): Observation driven models where the conditional distribution of Y_t is specified as a function of past observations $y_{t-1}, y_{t-2}, ...$ and parameter-driven models where autocorrelation as well as possible overdispersion in the data is introduced through a latent process. In this chapter we will first discuss a parameter-driven model considered, for example, by Zeger (1988) and by Chan and Ledolter (1995). They introduce a Poisson-GLM with a latent process in the mean which we will also apply to our option data example in the following chapters. As an example of an observation-driven model we will then discuss a Poisson-GLARMA model and relate it to the parameter-driven models discussed before. Some details about GLARMA models can be found in Shephard (1995) and in Davis et al. (2003). For a more extensive overview of the different models for time-dependent count data that have been proposed in the literature see Davis et al. (1999).

4.1 Poisson-GLMs with a latent process in the mean

4.1.1 General model framework

The general model framework is given by

$$Y_t|\alpha_t, \mathbf{x_t} \sim Poi(\alpha_t \cdot \mu_t)$$

where $\{\alpha_t\}$ is a non-negative stationary time series with $E(\alpha_t) = 1$ for all t and

$$log(\mu_t) = \mathbf{x}_t^t \boldsymbol{\beta} \tag{4.1}$$

The observations Y_t conditional on the regressors $\mathbf{x}_t$ and on the latent process α_t are assumed to be independent. We denote the (homogeneous) variance of the latent process $\{\alpha_t\}$ by $\sigma_\alpha^2 := Var(\alpha_t)$, its autocovariance function by $\gamma_\alpha(k) := E[(\alpha_t - 1)(\alpha_{t+k} - 1)]$ for $k = 0, 1, 2, ...$ and the autocorrelation function by $\rho_\alpha(k) := \frac{\gamma_\alpha(k)}{\sigma_\alpha^2}$.
Note that in this model

$$E(Y_t|\mathbf{x}_t) = E(E(Y_t|\alpha_t, \mathbf{x}_t)) = E(\alpha_t) \cdot exp(\mathbf{x}_t^t \boldsymbol{\beta}) = exp(\mathbf{x}_t^t \boldsymbol{\beta}),$$

similar to the expectation of $Y_t|\mathbf{x}_t$ in an ordinary GLM. We now show, however, how overdispersion and autocorrelation are introduced into the model through the latent process $\{\alpha_t\}$. We therefore calculate the variance, covariance and autocorrelation of the variable of interest conditional upon the regressors $\mathbf{x}_t$ (but not on the unobservable latent process $\{\alpha_t\}$).

Conditional variance of Y_t given $\mathbf{x}_t$:

$$\begin{aligned} Var(Y_t|\mathbf{x}_t) &= E(Var(Y_t|\mathbf{x}_t, \alpha_t)) + Var(E(Y_t|\mathbf{x}_t, \alpha_t)) \\ &= \mu_t + Var(\alpha_t \cdot \mu_t) \\ &= \mu_t + Var(\alpha_t) \cdot \mu_t^2 \\ &= \mu_t + \sigma_\alpha^2 \cdot \mu_t^2 \end{aligned} \tag{4.2}$$

Since $Var(Y_t|\mathbf{x}_t) \geq E(Y_t|\mathbf{x}_t)$ this model accounts for overdispersion in the data.

Conditional covariance of Y_t and Y_{t+k} given $\mathbf{x}_t$ and $\mathbf{x}_{t+k}$ for $k \geq 1$:

$$\begin{aligned} Cov(Y_t, Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}) &= E[(Y_t - \mu_t)(Y_{t+k} - \mu_{t+k})|\mathbf{x}_t, \mathbf{x}_{t+k}] \\ &= E(Y_t \cdot Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}) - E(Y_t|\mathbf{x}_t) \cdot E(Y_{t+k}|\mathbf{x}_{t+k}) \\ &= E(E(Y_t \cdot Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}, \alpha_t, \alpha_{t+k})) - \mu_t \cdot \mu_{t+k} \\ &= \mu_t \cdot \mu_{t+k} \cdot E(\alpha_t \cdot \alpha_{t+k}) - \mu_t \cdot \mu_{t+k} \\ &= \mu_t \cdot \mu_{t+k} \cdot [E(\alpha_t \cdot \alpha_{t+k}) - 1] \\ &= \mu_t \cdot \mu_{t+k} \cdot \gamma_\alpha(k) \end{aligned} \tag{4.3}$$

Conditional autocorrelation between Y_t and Y_{t+k} given $\mathbf{x}_t$ and $\mathbf{x}_{t+k}$:

$$\begin{aligned} Cor(Y_t, Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}) &= \frac{\mu_t \cdot \mu_{t+k} \cdot \gamma_\alpha(k)}{\sqrt{[\mu_t + \mu_t^2\sigma_\alpha^2] \cdot [\mu_{t+k} + \mu_{t+k}^2\sigma_\alpha^2]}} \\ &= \frac{\rho_\alpha(k)}{\sqrt{[1 + \frac{1}{\mu_t\sigma_\alpha^2}] \cdot [1 + \frac{1}{\mu_{t+k}\sigma_\alpha^2}]}} \end{aligned}$$

Note that if $\rho_\alpha(k) \neq 0 \Longrightarrow Cor(Y_t, Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}) \neq 0$, i.e. autocorrelation in the latent process induces autocorrelation into the variable of interest.
Note also that $|\rho_\alpha(k)| \geq |Cor(Y_t, Y_{t+k}|\mathbf{x}_t)|$ for all $t = 1, .., T$ and for all $k \in \mathbb{N}$, i.e. the observed autocorrelation in the count process Y_t will typically underestimate the autocorrelation in the latent process $\{\alpha_t\}$.

Since Zeger (1988) uses moment estimators for σ_α^2 and for the autocorrelation function $\rho_\alpha(k)$ of the latent process, he does not need to specify the exact structure of the latent process. However, he mentions a Gaussian AR(1) structure for $u_t := log(\alpha_t)$ as a possible approach. This specification is explicitly given in a model discussed by Chan and Ledolter (1995) which is a slight modification of the model setting (4.1) and which we will discuss in a more detailed way in the following section.

4.1.2 The special case of a Gaussian AR(1) latent process

In the general setting of (4.1) the Poisson GLM with an AR(1) latent process in the mean is given by:

$$\begin{aligned} Y_t &\sim Poi(\lambda_t), \\ \lambda_t &= exp(u_t + \mathbf{x}_t^t\boldsymbol{\beta}), \\ u_t = -\frac{\sigma_u^2}{2} \cdot (1-\rho) + \rho u_{t-1} + \varepsilon_t \qquad &\text{with} \qquad \varepsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma_\varepsilon^2) \end{aligned} \tag{4.4}$$

where again the observations Y_t conditional on the regressors $\mathbf{x}_t$ and on the latent process $\{\alpha_t\}$ are assumed to be independent.
Note that as in this model

$$\begin{aligned} u_t &= -\frac{\sigma_u^2}{2} \cdot (1-\rho) + \rho u_{t-1} + \varepsilon_t \\ &= -\frac{\sigma_u^2}{2} \cdot (1-\rho) + \rho \cdot (-\frac{\sigma_u^2}{2} \cdot (1-\rho) + \rho u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= -\frac{\sigma_u^2}{2} \cdot (1-\rho)(1+\rho) + \rho^2 u_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t \\ &= \dots \end{aligned}$$

$$
\begin{aligned}
&= -\frac{\sigma_u^2}{2} \cdot (1-\rho) \cdot \sum_{j=0}^{\infty} \rho^j + \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j} \\
&= -\frac{\sigma_u^2}{2} + \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}
\end{aligned}
$$

it follows that

$$E(u_t) = -\frac{\sigma_u^2}{2} + \sum_{j=0}^{\infty} \rho^j \cdot E(\varepsilon_{t-j}) = -\frac{\sigma_u^2}{2}$$

and

$$\sigma_u^2 := Var(u_t) = E[(u_t - E(u_t))^2] = E\left[\left(\sum_{j=0}^{\infty} \rho^j \cdot \varepsilon_{t-j}\right)^2\right] = \sum_{j=0}^{\infty} (\rho^2)^j \cdot E(\varepsilon_{t-j}^2) = \frac{\sigma_\varepsilon^2}{1-\rho^2}.$$

Thus, $u_t \sim N(-\frac{\sigma_u^2}{2}, \sigma_u^2)$. Note that the condition $E(\alpha_t) = E(exp(u_t)) = 1$, which was one of the model specifications in (4.1), is satisfied since the expectation of a random variable $X = exp(Z)$ where $Z \sim N(\mu, \sigma^2)$ (i.e. of a log-normally distributed random variable X) can be calculated as (see for example Johnson et al. (1994)):

$$E(X) = exp\left(\mu + \frac{\sigma^2}{2}\right). \tag{4.5}$$

Furthermore

$$Var(X) = e^{2\mu + 0.5\sigma^2}(e^{\sigma^2} - 1), \tag{4.6}$$

which we mention here for further reference.

We have already calculated the variance, covariance and autocorrelation of Y_t conditional only on the regressors and we have seen how through the autocovariance function of the latent process overdispersion and autocorrelation is introduced into the model. We will now derive the relationship between the autocovariance function $\gamma_\alpha(k)$ of the latent process $\{\alpha_t\}$ and the autocovariance function $\gamma_u(k)$ of the process $u_t = log(\alpha_t)$ as previously defined.
We have

$$
\begin{aligned}
\gamma_\alpha(k) &= E[(\alpha_t - 1)(\alpha_{t+k} - 1)] \\
&= E[(exp(u_t) - 1)(exp(u_{t+k}) - 1)] \\
&= E[exp(u_t + u_{t+k})] - E[exp(u_t)] - E[exp(u_{t+k})] + 1 \\
&= E[exp(u_t + u_{t+k})] - 1
\end{aligned}
$$

Now note that $u_t + u_{t+k}$ is (as a sum of two normally distributed random variables) normally distributed with $E(u_t + u_{t+k}) = E(u_t) + E(u_{t+k}) = -\sigma_u^2$ and

$$Var(u_t + u_{t+k}) = Var(u_t) + Var(u_{t+k}) + 2 \cdot Cov(u_t, u_{t+k}) = 2(\sigma_u^2 + \gamma_u(k))$$

Thus, $exp(u_t + u_{t+k})$ is log-normally distributed with

$$E[exp(u_t + u_{t+k})] = exp\left(-\sigma_u^2 + \frac{1}{2} \cdot 2(\sigma_u^2 + \gamma_u(k))\right) = exp(\gamma_u(k))$$

which yields

$$\gamma_\alpha(k) = exp(\gamma_u(k)) - 1$$

and in particular for $k = 0$:

$$\sigma_\alpha^2 = \gamma_\alpha(0) = e^{\sigma_u^2} - 1.$$

Since in a stationary Gaussian AR(1) process $\{u_t\}$ according to (3.32) the autocovariance function can be calculated as $\gamma_u(k) = \rho^k \cdot \gamma_u(0) = \rho^k \cdot \sigma_u^2$ the conditional variance and autocorrelation of Y_t in Model (4.4) given the regressors can be calculated using (4.2) and (4.3):

$$Var(Y_t|\mathbf{x}_t) = \mu_t + (e^{\sigma_u^2} - 1)\mu_t^2$$

$$Cor(Y_t, Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}) = \frac{\mu_t \cdot \mu_{t+k} \cdot (e^{\rho^k \cdot \sigma_u^2} - 1)}{\sqrt{[\mu_t + \mu_t^2(e^{\sigma_u^2} - 1)] \cdot [\mu_{t+k} + \mu_{t+k}^2(e^{\sigma_u^2} - 1)]}}$$

A slight modification of Model (4.4) was considered by Chan and Ledolter (1995) who specify their model by:

$$Y_t \sim Poi(\lambda_t), \tag{4.7}$$

$$\lambda_t = exp(u_t + \mathbf{x}_t^t \boldsymbol{\beta}),$$

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad \text{with} \qquad \varepsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma_\varepsilon^2)$$

Recall from Chapter 3.2 that in this case

$$u_t = \rho u_{t-1} + \varepsilon_t = \rho \cdot (\rho u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \rho^2 u_{t-2} + \rho \varepsilon_{t-1} + \varepsilon_t = \ldots = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j},$$

$$E(u_t) = \sum_{j=0}^{\infty} \rho^j \cdot E(\varepsilon_{t-j}) = 0$$

and

$$\sigma_u^2 := Var(u_t) = E(u_t^2) = E\left[\left(\sum_{j=0}^{\infty} \rho^j \cdot \varepsilon_{t-j}\right)^2\right] = \sum_{j=0}^{\infty} (\rho^2)^j \cdot E(\varepsilon_{t-j}^2) = \frac{\sigma_\varepsilon^2}{1 - \rho^2}.$$

Thus, in this model $u_t \sim N(0, \sigma_u^2)$, $E(exp(u_t)) = exp(\frac{\sigma_u^2}{2})$ and

$$\begin{aligned} E(Y_t|\mathbf{x}_t) &= = E(E(Y_t|u_t,\mathbf{x}_t)) \\ &= = E(exp(u_t))\cdot exp(\mathbf{x}_t^t\boldsymbol{\beta}) \\ &= exp\left(\frac{\sigma_u^2}{2}+\mathbf{x}_t^t\boldsymbol{\beta}\right) \\ &=: exp\left(\frac{\sigma_u^2}{2}\right)\cdot\mu_t \end{aligned} \tag{4.8}$$

Although the Chan and Ledolter-Model (4.7) is slightly different from the Model (4.1)/(4.4) originally considered by Zeger it is straightforward to compare parameter estimates for the two model settings.
If one defines

$$\begin{aligned} \alpha_t &:= \frac{exp(u_t)}{E(exp(u_t))} \\ &= \frac{exp(u_t)}{exp\left(\frac{\sigma_u^2}{2}\right)} \end{aligned} \tag{4.9}$$

all conditions of Model (4.1) are satisfied. Since $E(Y_t|\mathbf{x}_t) = exp\left(\frac{\sigma_u^2}{2}+\mathbf{x}_t^t\boldsymbol{\beta}\right)$ in the model (4.7) and $E(Y_t|\mathbf{x}_t) = exp(\mathbf{x}_t^t\boldsymbol{\beta})$ in the Model (4.1) one simply needs to add $\frac{\hat{\sigma}_u^2}{2}$ to the intercept estimate in (4.7) in order to obtain comparable results for the estimates of the parameter vector $\boldsymbol{\beta}$. For u_t and α_t as given in (4.7) and (4.9) respectively, using that $u_t + u_{t+k} \sim N(0, 2\cdot(\sigma_u^2 + Cov(u_t, u_{t-k}))$ and using the formula for the expectation of the log-normal distribution (4.5), we get:

$$\begin{aligned} \gamma_\alpha(k) &= E[(\alpha_t-1)(\alpha_{t+k}-1)] \\ &= \frac{E[exp(u_t+u_{t+k})]}{E[exp(u_t)]\cdot E[exp(u_{t+k})]} - 1 \\ &= \frac{e^{\frac{1}{2}\cdot 2(\sigma_u^2+\gamma_u(k))}}{e^{\sigma_u^2}} - 1 \\ &= e^{\gamma_u(k)} - 1 \end{aligned} \tag{4.10}$$

where $\gamma_u(k) = Cov(u_t, u_{t+k}) = \rho^k\sigma_u^2$ is the autocovariance function of the standard AR(1)-process $\{u_t\}$ in (4.7). It follows that

$$\sigma_\alpha^2 = \gamma_\alpha(0) = e^{\sigma_u^2} - 1,$$

$$\rho_\alpha(1) = \frac{\gamma_\alpha(1)}{\gamma_\alpha(0)} \overset{(4.10)}{=} \frac{e^{\rho\sigma_u^2}-1}{e^{\sigma_u^2}-1}. \tag{4.11}$$

Thus, after estimating σ_u^2 and ρ in model (4.7) one can calculate $\hat{\sigma}_\alpha^2 = e^{\hat{\sigma}_u^2} - 1$ and $\hat{\rho}_\alpha(1) = \frac{e^{\hat{\rho}\hat{\sigma}_u^2}-1}{e^{\hat{\sigma}_u^2}-1}$. Then, these results can be compared to estimates for the variance and

autocorrelation of the latent process $\{\alpha_t\}$ in the Model (4.1) considered by Zeger (1988).

In order to give a complete discussion of the Model (4.7) we also state the conditional variance $Var(Y_t|\mathbf{x}_t)$ and the conditional autocovariance $Cov(Y_t, Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k})$ in this case where again $\mu_t = exp(\mathbf{x}_t^t\boldsymbol{\beta})$.

$$\begin{aligned} Var(Y_t|\mathbf{x}_t) &= E(Var(Y_t|\mathbf{x}_t, u_t)) + Var(E(Y_t|\mathbf{x}_t, u_t)) \\ &= exp\left(\frac{\sigma_u^2}{2}\right) \cdot \mu_t + Var(exp(u_t) \cdot \mu_t) \\ &= exp\left(\frac{\sigma_u^2}{2}\right) \cdot \mu_t + Var(exp(u_t)) \cdot \mu_t^2 \\ &= exp\left(\frac{\sigma_u^2}{2}\right) \cdot \mu_t + exp(0.5\sigma_u^2)(exp(\sigma_u^2) - 1) \cdot \mu_t^2 \end{aligned} \tag{4.12}$$

$$\begin{aligned} Cov(Y_t, Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}) &= E(Y_t \cdot Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}) - E(Y_t|\mathbf{x}_t) \cdot E(Y_{t+k}|\mathbf{x}_{t+k}) \\ &= E(E(Y_t \cdot Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}, u_t, u_{t+k})) - \mu_t \cdot \mu_{t+k} \cdot exp(\sigma_u^2) \\ &= \mu_t \cdot \mu_{t+k} \cdot E(exp(u_t) \cdot exp(u_{t+k})) - \mu_t \cdot \mu_{t+k} \cdot exp(\sigma_u^2) \\ &= \mu_t \cdot \mu_{t+k} \cdot [E(exp(u_t) \cdot exp(u_{t+k})) - exp(\sigma_u^2)] \\ &\overset{(4.10)}{=} \mu_t \cdot \mu_{t+k} \cdot exp(\sigma_u^2) \cdot (exp(\gamma_u(k)) - 1) \\ &= \mu_t \cdot \mu_{t+k} \cdot exp(\sigma_u^2) \cdot (exp(\rho^k \sigma_u^2) - 1) \end{aligned} \tag{4.13}$$

These results are similar to the ones obtained in the Model specifications (4.1)/(4.4) if one takes into account the difference in the intercepts of the model settings.

4.2 Poisson-GLARMA models

As an example for an observation driven regression model for time-dependent count data we now introduce the Poisson-GLARMA model as discussed for example by Davis et al. (2003). Although it belongs to the class of observation-driven models it is closely related to the parameter-driven models discussed in the previous section.

4.2.1 General model framework

Before we can discuss the general model framework of a Poisson-GLARMA model, we need to define the term *martingale difference sequence*.

Definition 4.1 (Martingale difference sequence) *A stochastic process $\{e_t, t \in T\}$ is said to be a martingale difference sequence relative to the process $\{X_t, t \in T\}$ if*

$$E(e_{t+1}|X_s, -\infty < s \leq t) = 0 \text{ for all } t \in T.$$

Let Y_t for $t = 1, 2, ...$ denote the observations of the variable of interest and assume that given the past history up to $t-1$, i.e. given the σ-algebra $\mathcal{F}_{t-1} = \sigma\{Y_s, s \leq t-1\}$, Y_t is Poisson-distributed with mean λ_t, denoted by

$$Y_t|\mathcal{F}_{t-1} \sim Poi(\lambda_t).$$

Moreover, define e_t as

$$e_t := \frac{Y_t - \lambda_t}{\sqrt{\lambda_t}}. \tag{4.14}$$

Since

$$E(e_t|Y_{t-1}) = E(e_t|\mathcal{F}_{t-1}) = \frac{E(Y_t|\mathcal{F}_{t-1}) - \lambda_t}{\sqrt{\lambda_t}} = 0$$

$\{e_t\}$ is a martingale difference sequence relative to $\{Y_t\}$.

Now let $\{U_t, t \geq 0\}$ be an ARMA(p,q)-process as defined in (3.19) with noise given by the martingale difference sequence $\{e_t\}$, i.e.

$$U_t = \phi_1 U_{t-1} + ... + \phi_p U_{t-p} + e_t + \theta_1 e_{t-1} + ... + \theta_q e_{t-q}$$

where the polynomials $\phi(\cdot)$ and $\theta(\cdot)$ have no common zeros and all their zeros lie outside the unit circle.
Then, as shown in Chapter 3.2, U_t is causal and there exists a sequence $\{\psi_j, j = 1, 2, ...\}$ with $\sum_{j=0}^{\infty} |\psi_j| < \infty$ such that

$$U_t = \sum_{j=0}^{\infty} \psi_j e_{t-j} = e_t + \sum_{j=1}^{\infty} \psi_j e_{t-j} := e_t + \tilde{U}_t$$

where the coefficients ψ_j are determined by comparing coefficients in

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)} = \frac{1 - \theta_1 z - ... - \theta_q z^q}{1 - \phi_1 z - ... - \phi_p z^p}$$

Since $E(e_t) = 0$ for all t, $\tilde{U}_t$ can be regarded as a predictor for U_t given the past $\mathcal{F}_{t-1}$. It is therefore reasonable to define the general Poisson-GLARMA(p,q) model by:

$$Y_t|\mathcal{F}_{t-1} \sim Poi(\lambda_t).$$

$$log(\lambda_t) = \mathbf{x}_t^t\boldsymbol{\beta} + \tilde{U}_t = \mathbf{x}_t^t\boldsymbol{\beta} + \sum_{j=1}^{\infty}\psi_j e_{t-j} \tag{4.15}$$

Note that since the distribution of $Y_t|\mathcal{F}_{t-1}, \mathbf{x}_t$ depends on past martingale differences

$$e_{t-i} = \frac{y_{t-i} - \lambda_{t-i}}{\sqrt{\lambda_{t-i}}},$$

for $i = 1, 2, ...$ it is correct to say that Poisson-GLARMA models belong to the class of observation-driven models where the conditional distribution of Y_t is specified as a function of past observations $y_{t-1}, y_{t-2}, ...$

In the model fitting stage, the question arises how to deal with e_t and $\tilde{U}_t$ for $t \leq 0$ and how to compute $\tilde{U}_t$ for $t > 0$.
Since there are no observations for $t \leq 0$ it is reasonable to set $e_t = 0$ and $\tilde{U}_t = 0$ for $t \leq 0$. For $t > 0$, $\tilde{U}_t$ as a predictor for U_t given $\mathcal{F}_{t-1}$ can be computed recursively using the ARMA recursion:

$$\begin{aligned}\tilde{U}_t &= \phi_1 U_{t-1} + ... + \phi_p U_{t-p} + \theta_1 e_{t-1} + ... + \theta_q e_{t-q} \\ &= \phi_1(\tilde{U}_{t-1} + e_{t-1}) + ... + \phi_p(\tilde{U}_{t-p} + e_{t-p}) + \theta_1 e_{t-1} + ... + \theta_q e_{t-q}\end{aligned}$$

In the following paragraph we develop some properties of the Poisson-GLARMA model and relate it to the parameter-driven model introduced in the previous section.

4.2.2 Relationship between Poisson-GLARMA models and Poisson-GLMs with a latent process

We have already shown that given $\mathcal{F}_{t-1}$ the martingale differences e_t have expectation 0, i.e.

$$E(e_t) = E(E(e_t|\mathcal{F}_{t-1})) = 0 \quad \text{for } t \geq 1$$

The conditional variance of e_t given the past $\mathcal{F}_{t-1}$ for $t \geq 1$ can be calculated as:

$$Var(e_t|\mathcal{F}_{t-1}) = \frac{Var(Y_t|\mathcal{F}_{t-1})}{\lambda_t} = \frac{\lambda_t}{\lambda_t} = 1$$

From the martingale difference property of e_t is follows directly that for $t > s$

$$E(e_t e_s) = E(E(e_t e_s|\mathcal{F}_{t-1})) = E(e_s \cdot E(e_t|\mathcal{F}_{t-1})) = 0$$

and thus, for $t \neq s$

$$Cov(e_t, e_s) = 0.$$

With these properties at hand it is easy to calculate expectation, variance and covariances of $W_t := log(\lambda_t)$:

$$E(W_t) = E\left(\mathbf{x}_t^t\boldsymbol{\beta} + \sum_{j=1}^{\infty}\psi_j e_{t-j}\right) = \mathbf{x}_t^t\boldsymbol{\beta} + \sum_{j=1}^{\infty}\psi_j E(e_{t-j}) = \mathbf{x}_t^t\boldsymbol{\beta},$$

$$Var(W_t) = Var\left(\sum_{j=1}^{\infty}\psi_j e_{t-j}\right) = E\left[\left(\sum_{j=1}^{\infty}\psi_j e_{t-j}\right)^2\right] = \sum_{j=1}^{\infty}\psi_j^2 E(e_{t-j}^2) = \sum_{j=1}^{\infty}\psi_j^2,$$

$$Cov(W_t, W_{t+h}) = E\left(\sum_{j=1}^{\infty}\psi_j e_{t-j}\cdot\sum_{j=1}^{\infty}\psi_j e_{t+h-j}\right) = \sum_{j=1}^{\infty}\psi_j\psi_{j+h}E(e_{t-j}^2) = \sum_{j=1}^{\infty}\psi_j\psi_{j+h}$$

Note that while $W_t = \mathbf{x}_t^t\boldsymbol{\beta} + \sum_{j=1}^{\infty}\psi_j e_{t-j}$ has mean $\mathbf{x}_t^t\boldsymbol{\beta}$ the mean of $\lambda_t = e^{W_t}$ can not be simply calculated as $exp\{\mathbf{x}_t^t\boldsymbol{\beta}\}$.
However, if the sequence $\{\tilde{U}_t = \sum_{j=1}^{\infty}\psi_j e_{t-j}, t = 1, 2, ...\}$ can be approximated by a Gaussian stationary sequence $\{U_t'\}$ with zero mean and variances and covariances matched to those of $\{\sum_{j=1}^{\infty}\psi_j e_{t-j}, t = 1, 2, ...\}$, i.e. matched to those of $\{W_t, t = 1, 2, ...\}$, then

$$W_t \approx \mathbf{x}_t^t\boldsymbol{\beta} + U_t' \quad \sim N(\mathbf{x}_t^t\boldsymbol{\beta}, \sum_{j=1}^{\infty}\psi_j^2).$$

Thus, in this case λ_t would be approximately log-normally distributed with

$$E(\lambda_t) = E(e^{W_t}) \approx e^{\mathbf{x}_t^t\boldsymbol{\beta} + \frac{1}{2}\sum_{j=1}^{\infty}\psi_j^2}$$

While the distribution of e_t is of course not normal, the linear combination $\sum_{j=1}^{\infty}\psi_j e_{t-j}$ may have a distribution which can be closely approximated by a sequence of correlated normal random variables. If this is the case, $\tilde{U}_t$ can be considered as a proxy for a latent process in the mean of a Poisson-GLM and the Poisson-GLARMA model is approximately equivalent to the parameter-driven models discussed in the previous section.

4.3 Comparison of models and estimation methods with a well-known data example of monthly Polio incidence in the U.S.

In this section we apply the latent process Poisson-regression model (4.1) to a data example that was originally presented by Zeger (1988) and has been used by various other authors as reference. For the estimation of the parameters of the model we will use Markov Chain Monte Carlo (MCMC)-estimation with the WinBUGS software package.

The data set we consider consists of the monthly number of cases of poliomyelitis in the U.S. during the years 1970-1983 as reported by the Center for Disease Control. Figure 4.1 shows the data for January 1970 until December 1983. The question of interest when modelling the polio data is whether there is evidence for a decreasing time trend.

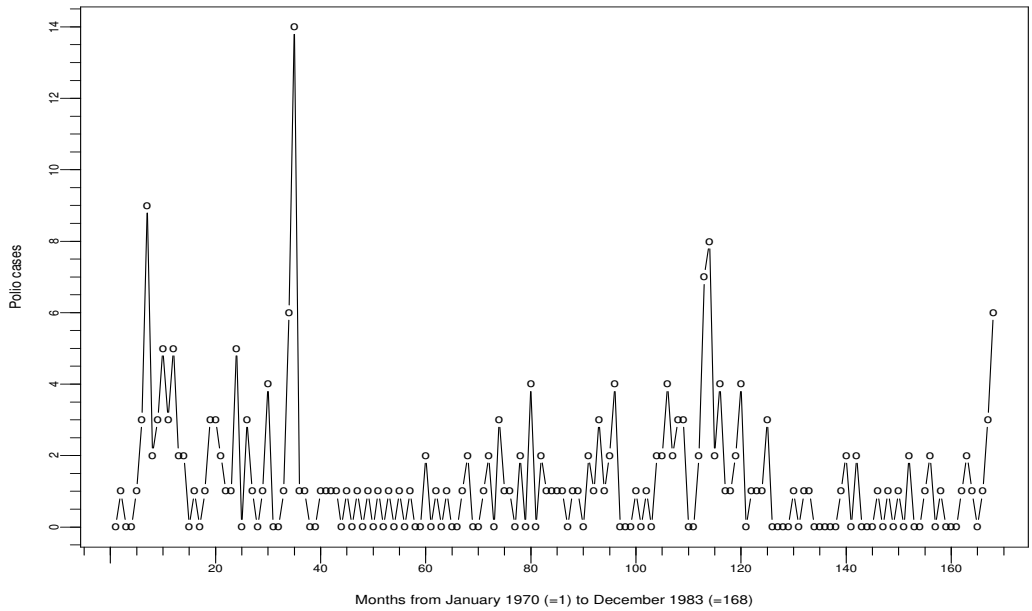

Figure 4.1: Monthly number of polio cases in the U.S. 1970-1983.

In addition to the intercept term $x_{t1} := 1$ and to the trend regressor $x_{t2} := \frac{t}{1000}$ there are four trigonometric regressors $x_{t3} := sin\left(\frac{2\pi t}{12}\right)$, $x_{t4} := cos\left(\frac{2\pi t}{12}\right)$, $x_{t5} := sin\left(\frac{2\pi t}{6}\right)$ and $x_{t6} := cos\left(\frac{2\pi t}{6}\right)$ which account for annual (x_{t3}, x_{t4}) and semi-annual (x_{t5}, x_{t6}) seasonalities in the data. Like Zeger (1988) and most of the other authors, we center the data around January 1976 by subtracting $\mathbf{x}_{73}$ from the regressors $\mathbf{x}_t$ for $t = 1, ..., 168$ (the observation 73 corresponds to January 1976).
The model we want to fit to the (centered) polio data is therefore given by:

$$Y_t \sim Poi(\lambda_t)$$

$$\lambda_t = \alpha_t \cdot exp\{\mathbf{x}_t^t \boldsymbol{\beta}\} = exp\{u_t\} \cdot exp\{\mathbf{x}_t^t \boldsymbol{\beta}\}$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2) \tag{4.16}$$

where given the latent variable α_t (and given the regressors) the variables Y_t are independent.

For parameter estimation determining the likelihood of the model is a natural starting point. For this purpose note that

$$p(y_t|\alpha_t, ..., \alpha_1, y_{t-1}, ..., y_1) = p(y_t|\alpha_t)$$

and

$$p(\alpha_{t+1}|\alpha_t, ..., \alpha_1, y_t, ..., y_1) = p(\alpha_{t+1}|\alpha_t).$$

The joint distribution of the variables $\{Y_t, t = 1, ..., T\}$ and the latent process $\{\alpha_t, t = 1, ..., T\}$ can be computed as:

$$\begin{aligned} p(y_1, ..., y_T, \alpha_1, ..., \alpha_T) &= p(y_T|\alpha_T, ..., \alpha_1, y_{t-1}, ..., y_1) \cdot p(\alpha_T, ..., \alpha_1, y_{t-1}, ..., y_1) \\ &= p(y_T|\alpha_T) \cdot p(\alpha_T|\alpha_{T-1}, ..., \alpha_1, y_{t-1}, ..., y_1) \cdot \\ &\quad \cdot p(\alpha_{T-1}, ..., \alpha_1, y_{t-1}, ..., y_1) \\ &= p(y_T|\alpha_T) \cdot p(\alpha_T|\alpha_{T-1}) \cdot p(y_{t-1}, ..., y_1, \alpha_{T-1}, ..., \alpha_1) \\ &= ... \\ &= \left(\prod_{t=1}^{T} p(y_t|\alpha_t)\right) \cdot \left(\prod_{t=2}^{T} p(\alpha_t|\alpha_{t-1})\right) \cdot p(\alpha_1) \end{aligned} \tag{4.17}$$

Now let $\varphi_{log}(x; \mu, \sigma)$ denote the probability density function of the log-Normal distribution with parameters μ and σ, i.e.

$$\varphi_{log}(x; \mu, \sigma) := \frac{1}{\sigma x \sqrt{2\pi}} \cdot exp\left\{\frac{-(log(x) - \mu)^2}{2\sigma^2}\right\}.$$

As previously discussed in Section 4.1., α_1 is log-normally distributed with parameters $\mu = 0$ and $\sigma = \frac{\sigma_\varepsilon^2}{1-\rho^2}$. Since $u_t|u_{t-1} \sim N(\rho u_{t-1}, \sigma_\varepsilon^2)$, $\alpha_t|\alpha_{t-1}$ is also log-normally distributed with parameters $\mu = \rho \cdot log(\alpha_{t-1})$ and $\sigma = \sigma_\varepsilon^2$.
According to (4.17) the likelihood function $f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\beta}, \rho, \sigma_\varepsilon)$ of Model (4.16) can then be written as the n-fold integral

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\beta}, \rho, \sigma_\varepsilon) &= \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \left(\prod_{t=1}^{T} f_{Poi}(y_t|\alpha_t)\right) \cdot \\ &\quad \cdot \left(\prod_{t=2}^{T} \varphi_{log}(\alpha_t; \rho \cdot log(\alpha_{t-1}), \sigma_\varepsilon^2)\right) \cdot \varphi_{log}(\alpha_1; 0, \frac{\sigma_\varepsilon^2}{1-\rho^2})\, d\alpha_1 ... d\alpha_T \\ &= \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \left(\prod_{t=1}^{T} exp\left\{-\alpha_t e^{\mathbf{x}_t^t \boldsymbol{\beta}}\right\} \cdot \frac{\left(\alpha_t e^{\mathbf{x}_t^t \boldsymbol{\beta}}\right)^{y_t}}{y_t!}\right) \cdot \end{aligned}$$

$$\cdot \left(\prod_{t=2}^{T} \varphi_{log}(\alpha_t; \rho \cdot log(\alpha_{t-1}), \sigma_\varepsilon^2) \right) \cdot \varphi_{log}(\alpha_1; 0, \frac{\sigma_\varepsilon^2}{1-\rho^2})\, d\alpha_1 ... d\alpha_T$$

$$= \frac{1}{\prod_{t=1}^{T}(y_t!)} \cdot \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} exp\left\{ \sum_{t=1}^{T} (y_t \cdot \mathbf{x}_t^t \boldsymbol{\beta} - \alpha_t e^{\mathbf{x}_t^t \boldsymbol{\beta}}) \right\} \left(\prod_{t=1}^{T} {\alpha_t}^{y_t} \right) \cdot$$

$$\cdot \left(\prod_{t=2}^{T} \varphi_{log}(\alpha_t; \rho \cdot log(\alpha_{t-1}), \sigma_\varepsilon^2) \right) \cdot \varphi_{log}(\alpha_1; 0, \frac{\sigma_\varepsilon^2}{1-\rho^2})\, d\alpha_1 ... d\alpha_T$$

Since the likelihood $f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\beta}, \rho, \sigma_\varepsilon)$ can not be written in closed form, direct maximum-likelihood estimation of the parameters is not feasible. However, several estimation techniques have been proposed in the literature.

Davis et al. (1999) show that for a large variety of regressor functions the simple GLM-estimates of $\boldsymbol{\beta}$ based on the algorithm discussed in Section 3.1 (that is, assuming that there is no latent process $\{\alpha_t\}$ in the mean) is a consistent estimator for $\boldsymbol{\beta}$ even when autocorrelation is present. They show that, asymptotically, the usual GLM estimate $\hat{\boldsymbol{\beta}}_{GLM}$ is normally distributed with expectation $\boldsymbol{\beta}$ and a covariance matrix that depends on the covariance structure of the latent process.

Zeger (1988) proposes estimating equations based on a time series analogue of quasi-likelihood as discussed for example by McCullagh (1983), while Chan and Ledolter (1995) use a Monte Carlo Expectation Maximization (MCEM) algorithm. For the MCEM algorithm the latent process $u_t = log(\alpha_t)$ must be explicitly specified as a Gaussian AR(1)-process whereas Zeger's approach allows for a more general specification of the latent process with the first and second order moments.

Davis et al. (1999) fit two Poisson-GLARMA models of the form (4.15) to the polio data for which direct maximum-likelihood estimation using the Newton-Raphson algorithm is possible. They fit a Poisson-GLARMA(6,0)- and a Poisson-GLARMA(0,6)-model where in both cases any AR- or MA-terms that are not significant at the 5%-level are dropped and the model is refitted. Finally ϕ_1, ϕ_5 and $\theta_1, \theta_2, \theta_5$ remain as significant parameters of the AR- and MA-terms respectively.

A table which offers an overview of the estimation results obtained by the previously mentioned authors is given at the end of this section (Table 4.2).

For our MCMC estimation in WinBUGS we have to choose priors for the parameters of the model. Since we want to keep the priors rather uninformative we choose

$$\begin{aligned} \rho &\sim Unif[-1,1] \\ \beta_i &\sim N(0,100) \quad \text{for } i = 1,...,6 \\ \tau_\varepsilon := \frac{1}{\sigma_\varepsilon^2} &\sim \Gamma(0.5, 0.5) \end{aligned}$$

We run 3 parallel chains (with overdispersed starting values) for each parameter and after an initial burn-in period of 2,000 iterations we record a further 8,000 iterations in each chain. Since autocorrelations within the chains are small, there is no need to thin the recorded iterations, so that we obtain a total of 24,000 iterations for the estimation. Figures 4.2 to 4.6 show the trajectories of the simulated chains for each parameter, as well as the plots of the estimated Gelman-Rubin potential scale reduction factor $\hat{R}$ as introduced in Chapter 3.3 . Both the trajectories of the chains and the Gelman-Rubin-statistic indicate that convergence is reached very quickly. The necessary burn-in according to the convergence diagnostics method of Raftery and Lewis (1992) is always estimated as < 100 for every parameter, which further justifies our choice of the burn-in and the MCMC sampling procedure in this data set. The results of the MCMC sampling are shown in Table 4.1.

Parameter	mean	std.err.	2.5%	median	97.5%
β_1 (Intercept)	0.022	0.276	-0.526	0.025	0.552
β_2 (Trend)	-3.201	3.350	-9.374	-3.331	3.873
β_3 (Annual cosine)	0.160	0.150	-0.138	0.161	0.454
β_4 (Annual sine)	-0.487	0.166	-0.817	-0.487	-0.162
β_5 (Semi-annual cos)	0.418	0.131	0.162	0.417	0.677
β_5 (Semi-annual sin)	-0.0150	0.129	-0.265	-0.016	0.239
ρ	0.656	0.160	0.298	0.675	0.909
σ^2_ε	0.325	0.145	0.106	0.303	0.667

Table 4.1: MCMC sampling results for the polio data example based on the last 8000 iterations of each chain.

Although the mean estimate of the trend term is negative, one can (due to the high standard error of the estimate) not come to the conclusion that there is a statistically significant decreasing trend in the number of monthly polio cases. This is in line with the results obtained by Zeger (1988) and by Davis et al. (1999) who, at least at the 5%-level, can not provide statistical evidence for a decreasing trend either. Due to a smaller standard error of their trend estimate, Chan and Ledolter (1995) come to the contrary result that the decreasing trend is indeed significant. Table 4.2 gives an overview of the estimation results obtained by these authors compared to our results of Table 4.1. Note that in order to be able to directly compare the estimates of the latent process models, intercept estimates of the model (4.16) used by Chan and Ledolter (1995) and by us had to be adjusted and $\hat{\sigma}^2_\alpha$, $\hat{\rho}_\alpha(1)$ had to be calculated as previously described in Section 4.1 .

	Poi-GLM		Parameter-driven models						Obs.-driven models			
			Zeger		Chan/Led.		WinBUGS		Glarma(6,0)		Glarma(0,6)	
Parameter	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
Intercept	0.21	0.08	0.17	0.13	0.42		0.31		0.14	0.12	0.13	0.11
Trend	-4.80	1.40	-4.35	2.68	-4.62	1.38	-3.20	3.35	-3.83	2.26	-3.93	2.18
Ann. cos.	-0.15	0.10	-0.11	0.16	0.15	0.09	0.16	0.15	-0.10	0.11	-0.10	0.12
Ann. sine	-0.53	0.11	-0.48	0.17	-0.50	0.12	-0.49	0.17	-0.51	0.13	-0.53	0.14
S-ann. cos	0.17	0.10	0.20	0.14	0.44	0.10	0.42	0.13	0.23	0.13	0.21	0.12
S-ann. sin	-0.43	0.10	-0.41	0.14	-0.04	0.10	-0.02	0.13	-0.40	0.12	-0.39	0.12
$\hat{\sigma}^2_\alpha$	-	-	0.77		0.54		0.77		-	-	-	-
$\hat{\rho}_\alpha(1)$	-	-	0.77		0.88		0.59		-	-	-	-
$\hat{\phi}_1$	-	-	-	-	-	-	-	-	0.23	0.05	-	-
$\hat{\phi}_5$	-	-	-	-	-	-	-	-	0.11	0.05	-	-
$\hat{\theta}_1$	-	-	-	-	-	-	-	-	-	-	0.22	0.06
$\hat{\theta}_2$	-	-	-	-	-	-	-	-	-	-	0.13	0.05
$\hat{\theta}_5$	-	-	-	-	-	-	-	-	-	-	0.09	0.04

Table 4.2: Comparison of estimates of different models and estimation techniques for the polio data example.

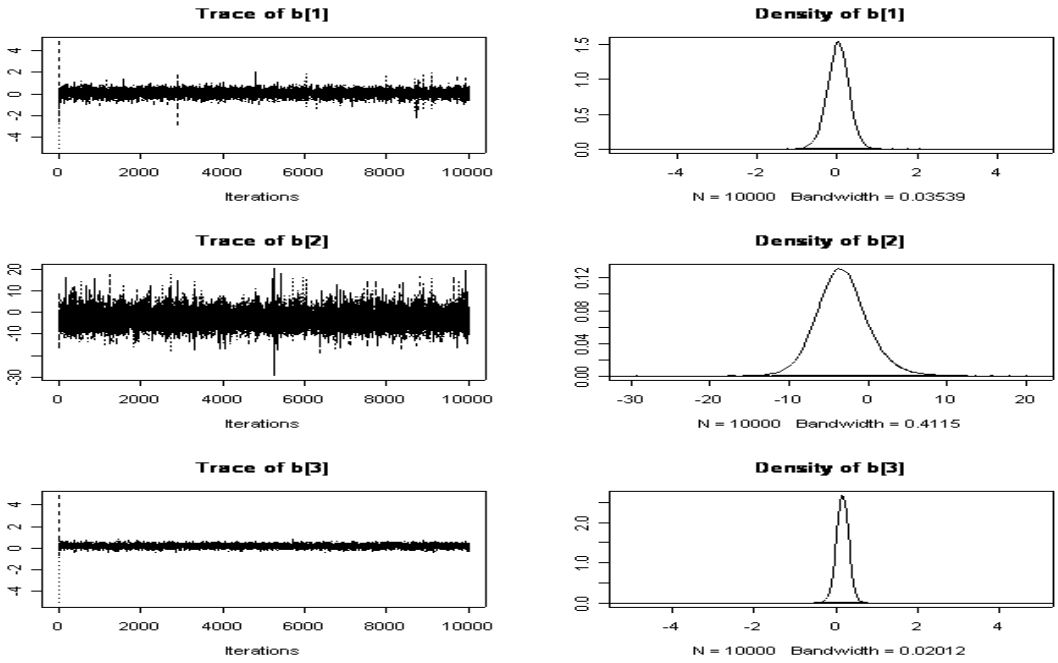

Figure 4.2: MCMC iterations and density estimates for $\beta_1, ..., \beta_3$ (b[1],...,b[3]) for the model (4.16).

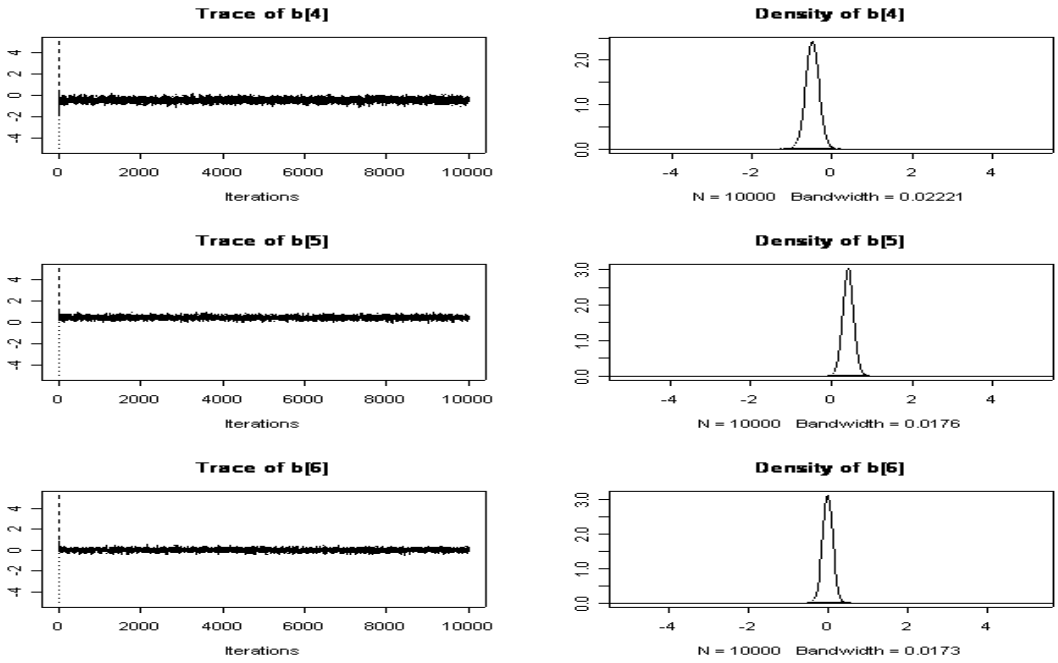

Figure 4.3: MCMC iterations and density estimates for $\beta_4, ..., \beta_6$ (b[4],...,b[6]) for the model (4.16).

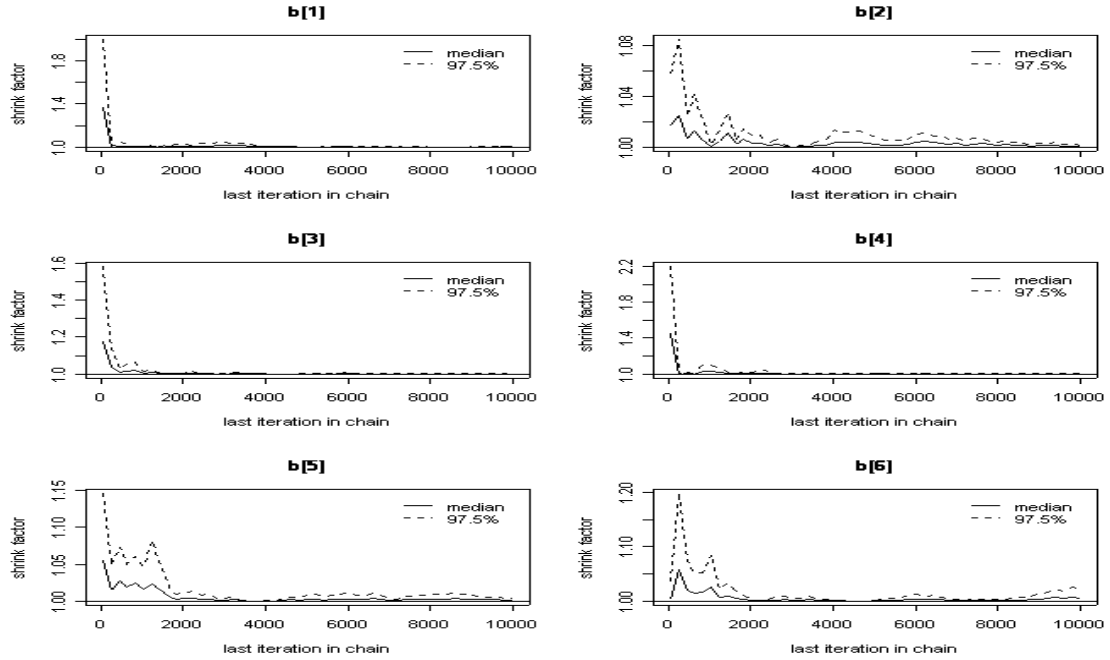

Figure 4.4: Gelman-Rubin convergence diagnostics for $\beta_1, ..., \beta_6$ (b[1],...,b[6]).

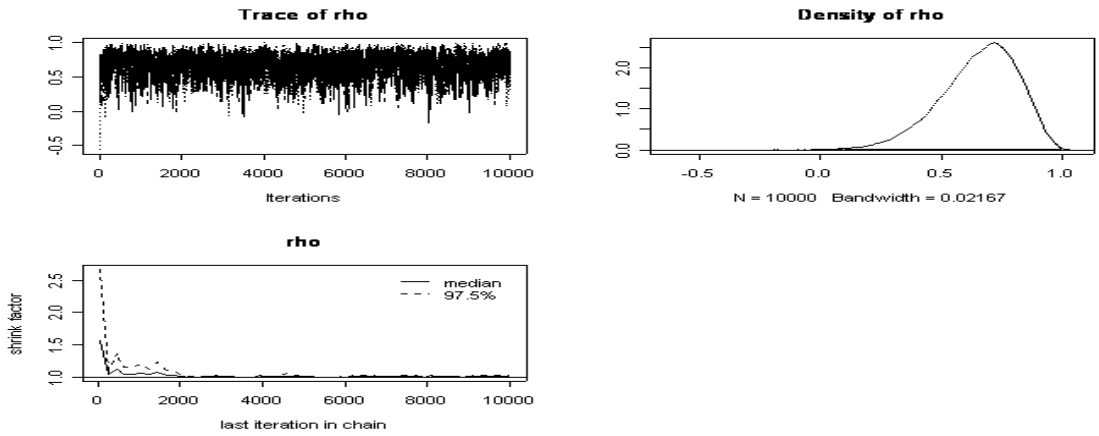

Figure 4.5: MCMC iterations, density estimates and Gelman-Rubin convergence diagnostics for ρ (rho) in the model (4.16).

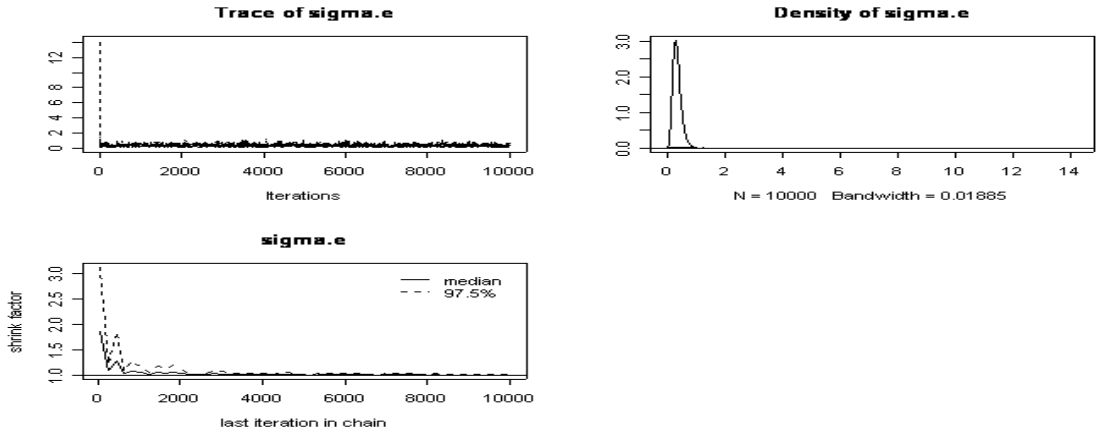

Figure 4.6: MCMC iterations, density estimates and Gelman-Rubin convergence diagnostics for σ_ε^2 (sigma.e) in the model (4.16).

Chapter 5

Data description and explorative data analysis

5.1 Available data and a representative data example

The available data set for this thesis consists of intraday quote-by-quote data, also called tick-by-tick data, for a large number of equity index options on four of the world's major equity indices: The XETRA DAX index, the DJ EURO STOXX 50 index, the S&P 500 index and the NIKKEI 225 index. The XETRA DAX index comprises the 30 most important German stocks. The top 50 stocks of the Euro-region are represented in the DJ Euro STOXX 50. The S&P 500 index consists of the top 500 US stocks chosen for market size, liquidity and group representation, and the NIKKEI 225 is the leading index of the Japanese stock market. The intraday data for the underlyings - the indices - is also available. The data set consists of a total of 55 pairs of Call and Put options on the XETRA DAX, differing in strike prices and expiration months. The strike prices range from 1000 to 7000 points, expiration months are March 2003 and June 2003. For the DJ Euro STOXX 50 a total number of 164 pairs of Call and Put options is available with strike prices from 1500 to 4400 and expiration months between July 2003 and March 2004. Furthermore, 26 pairs of options for the S&P 500 index and 107 option pairs for the NIKKEI 225 are available. All options considered in this thesis are European options.
The data was obtained from Reuters.

As a representative example of the behaviour of option prices we will study in this thesis the price changes of a Call option on the XETRA DAX index with strike price 2600 points and expiry month March 2003 between 10-Feb-2003 and 21-March-2003, the expiration day of the option. For this particular option the number of missing data values is reasonably small and due to its strike price the option was one of the most actively traded

options during the time period we consider.

5.2 General data background

Since we have chosen an option on the XETRA DAX index as an example for further investigation, it will be interesting to look at the price development of the DAX index first. Figure 5.1 shows the chart of the DAX index from 10-Feb-2003 to 09-July-2003 based on tick-by-tick data. Note that weekends are excluded from the time axis.

Figure 5.1: Tick-by-tick data for XETRA DAX index.

Figures 5.2 and 5.3 show the tick-by-tick charts of some options on the XETRA DAX index from 10-Feb-2003 to 21-March-2003, the last trading day of all of the selected options.
Figure 5.2 shows the last prices of some selected Call options with different strike prices including the option we have chosen as a representative example and which we will analyze more closely later. Figure 5.3 shows the last prices of some selected Put options. Please note that in these option charts weekends are not excluded from the time axis.

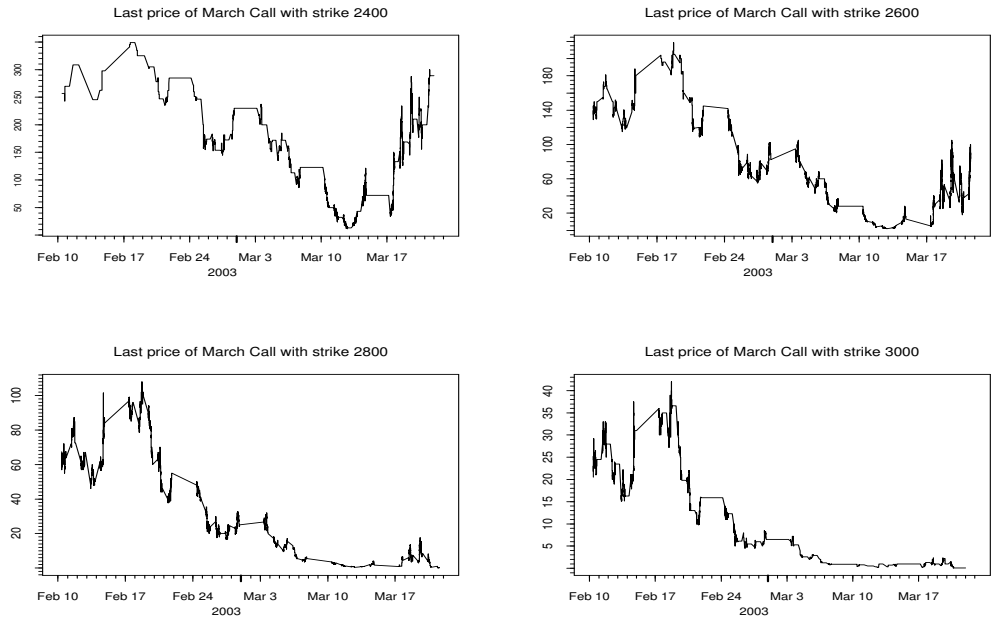

Figure 5.2: Tick-by-tick data for some selected Call options on XETRA DAX with different strike prices.

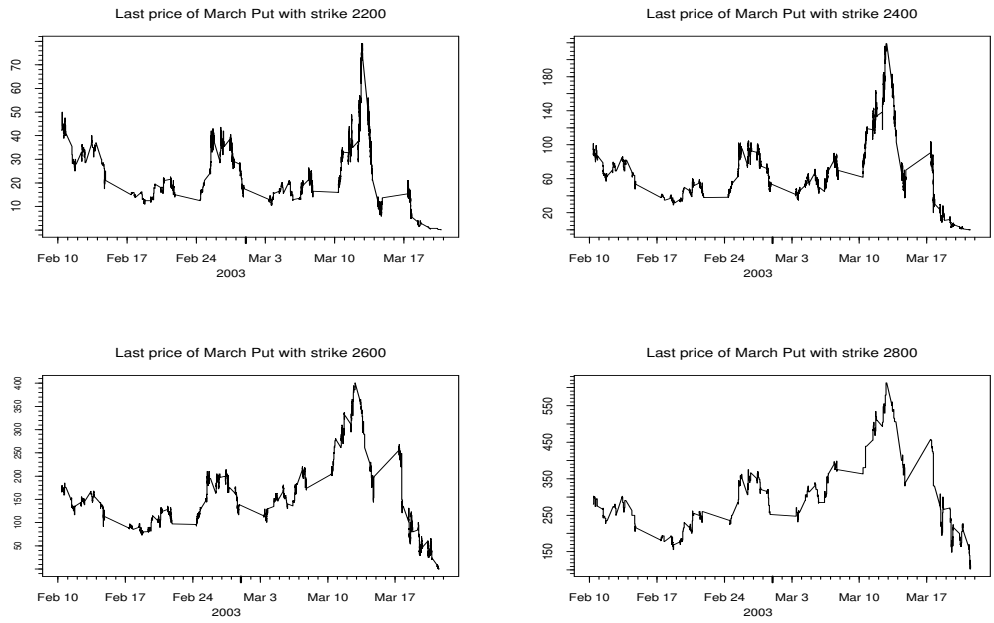

Figure 5.3: Tick-by-tick data for some selected Put options on XETRA DAX with different strike prices.

As already mentioned in Chapter 2, at maturity of the option, the 21-March-2003 in the data example, the option value is equal to its intrinsic value. Since options on the XETRA DAX index traded at EUREX are settled in cash, the buyer of the option receives an amount of cash equal to the intrinsic value from the writer when the option is exercised. One can easily see in Figures 5.2 and 5.3 that the Call options with strike prices 2400 and 2600 and the Put option with strike price 2800 were in-the-money at the time of maturity as the value of these options was positive at that time. The DAX index was quoted at 2686.46 points at 1 pm on 21-March-2003 when trading for the March options ceased.

The number of transaction price changes of a security that occur during a certain period of time can be considered as some measure of trading activity for that particular security. Figure 5.4 shows the number of non-zero price changes per hour of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003 plotted against the trading hours from 10-Feb-2003 10 am - 11 am to the last trading hour of the option, 21-March-2003 12 am - 1 pm. It can be observed that the trading activity and, as a most likely consequence, the short-term volatility of the option rose along the time axis. One possible reason for this behaviour may be the commission system described in Chapter 2 that tends to push retail investors in the direction of selling the options shortly before maturity rather than exercising them.

In the following chapters we will analyze the tick-by-tick non-zero transaction price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003 and relate these changes to a series of explanatory variables. In the following sections we therefore introduce and exploratively analyze these variables.

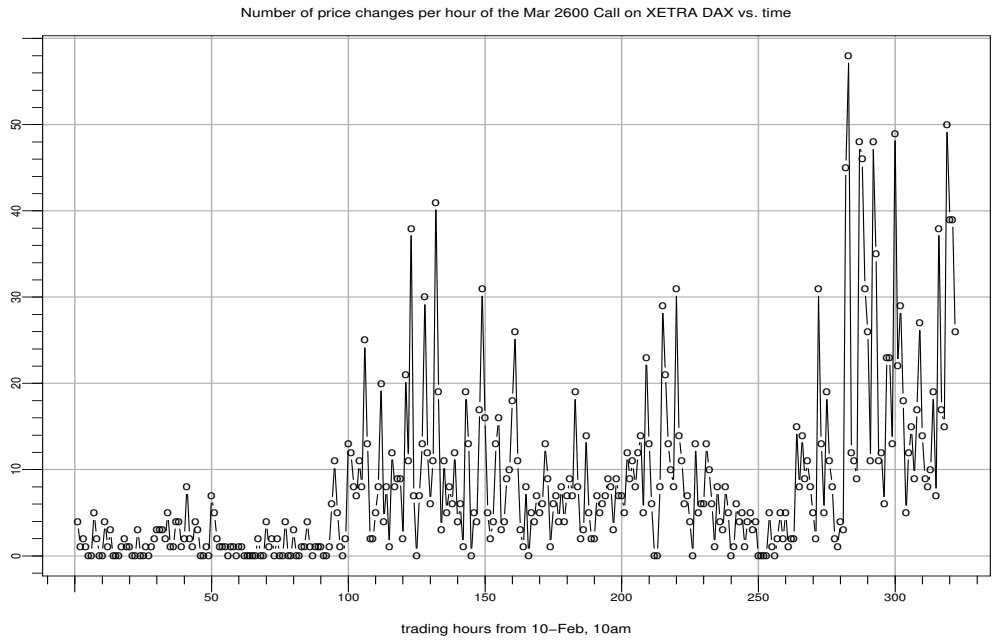

Figure 5.4: Number of non-zero transaction price changes per hour of the March Call on XETRA DAX with strike price 2600.

5.3 The variable of interest: Absolute option price changes

In order to get a first impression of the structure of the non-zero transaction price changes of the option it is useful to plot them against time. This plot is shown in Figure 5.5. Figure 5.6 shows the histogram of these price changes. The value of the price changes is shown in ticks where each tick represents a value of 0.1 index points or EUR 0.50.

In this thesis we do not model positive and negative price changes separately, but from now on consider only their absolute values. The symmetric structure of the positive and negative price changes which can be observed in Figure 5.6 justifies this approach. A possibility to take into account the direction of the price changes in the econometric modelling is discussed in Liesenfeld and Pohlmeier (2003) who decompose the process of the price changes into an autoregressive multinomial component for the direction of the price change and a dynamic count data component for the size of the price changes. Since in this thesis we consider absolute price changes we only need the dynamic count data component, for which we have already introduced several models in Chapter 4.

Since the dynamics of the price changes at the beginning and at the end of the trading day may differ from those during the day, we exclude the price changes before 10am and after 7pm from further investigation. This is in line with other authors (e.g. Liesenfeld and Pohlmeier (2003)) who proceed in a similar way.

Having a closer look at the data it can be observed that multiples of 10 ticks and to a lesser extent multiples of 5 ticks occur much more frequently than other values. An explanation for this may be the traders' preference for 'round' values. 10 ticks represent exactly 1 index point and 5 ticks represent 0.5 index points.
In order to avoid the problems this fact may cause in the modelling, we group the data to classes of 10 ticks, i.e. a value of 0 of our new grouped absolute price change variable means that there was an absolute price change between 1 and 10 ticks, a value of 1 stands for an absolute price change of 11-20 ticks, etc. Figure 5.7 shows the histogram of the absolute price changes grouped in this way. From now on we will simply use the term "absolute option price changes" for the non-zero transaction price changes of the option grouped as previously described.
Let t denote the time of a price change. For $t = 1, 2, ..., T$ let Y_t denote the absolute value of the price change after grouping as described before. In the data example of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003 we finally get a total of $T = 2419$ observations.

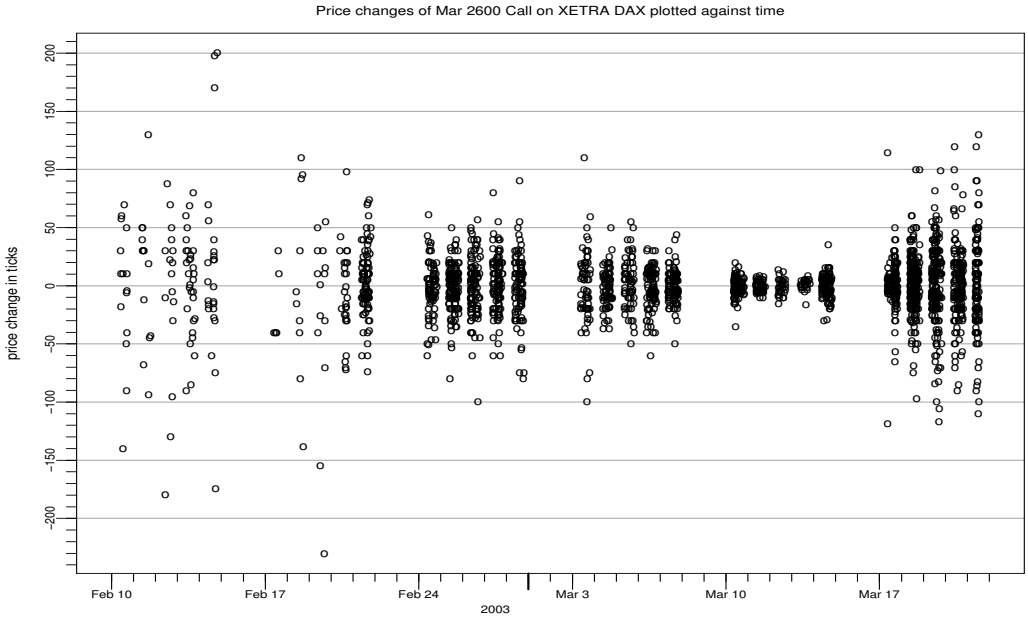

Figure 5.5: Non-zero transaction price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003.

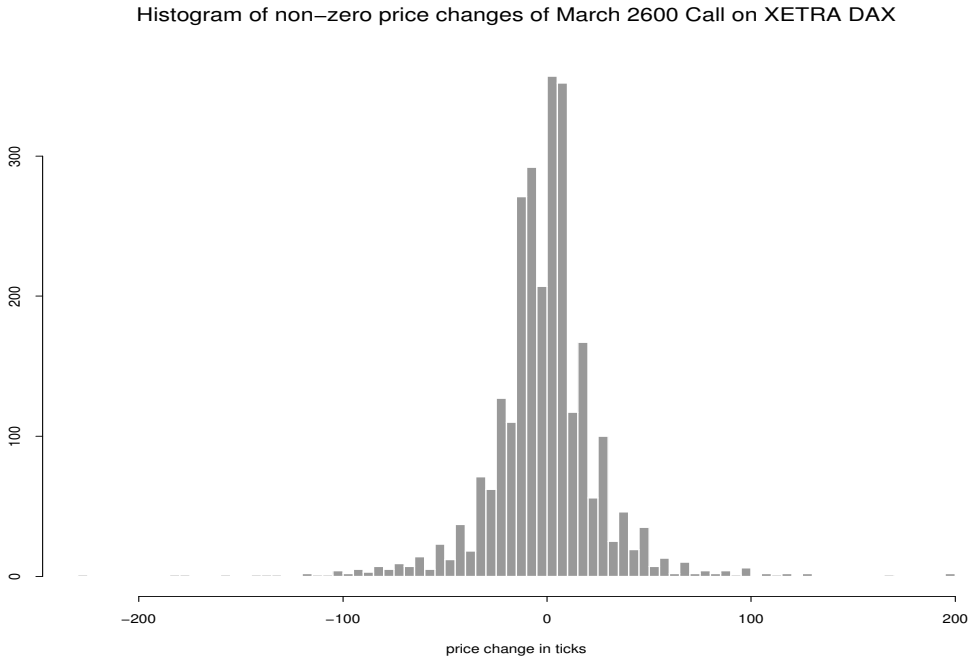

Figure 5.6: Histogram of non-zero transaction price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003.

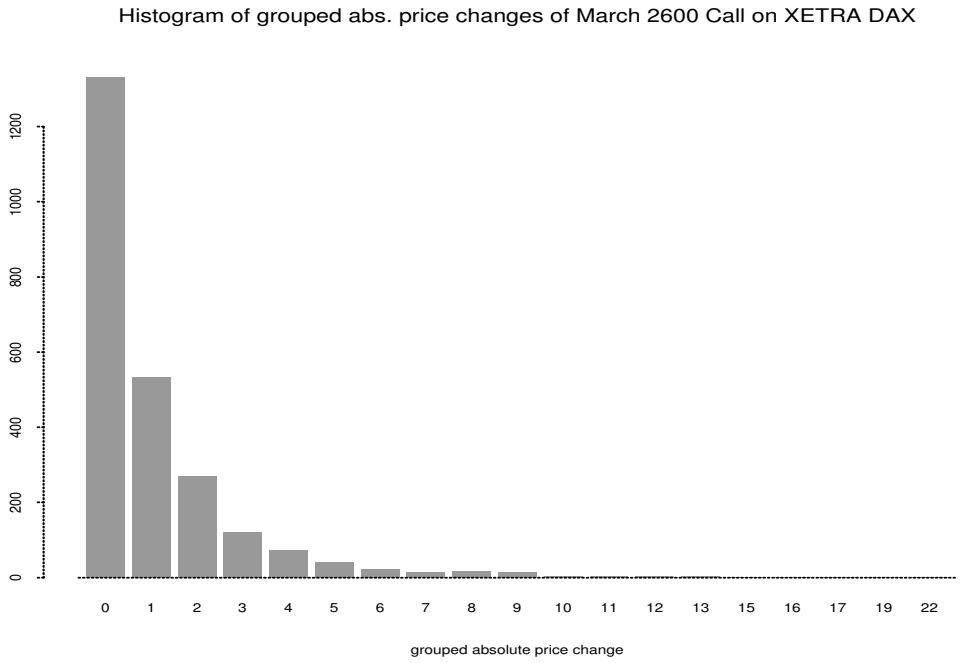

Figure 5.7: Histogram of grouped transaction price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003.

5.4 Explanatory variables

In this section we discuss a series of explanatory variables that may have influence on Y_t as defined in the previous section. In order to better understand the definition of the variables we will give a concrete data example for illustration purposes.
The following table shows consecutive quotations of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003 on 21-March-2003 a few minutes before trading for this option ceased at 1 pm. Non-zero transaction price changes (i.e. non-zero changes of the last price) of the option occurred at the points of time τ_1, τ_2 and τ_3. The corresponding values of the variable of interest are $Y_t = 7$ (since the price change at the point of time τ_1 is equal to 8 points or 80 ticks, which yields a value of 7 for our variable of interest grouped as described in the previous section), $Y_{t+1} = 2$ and $Y_{t+2} = 1$ where the t-th price change of the option in the time period we consider in this thesis occurred at the point of time τ_1.

Date & time of quot.	Last price (in pts.)	Bid (in pts.)	Ask (in pts.)	
2003-03-21 12:55:21	83	83.3	90	
2003-03-21 12:56:10	83	84.7	91	
2003-03-21 12:56:20	91	85.3	91	τ_1 =2003-03-21 12:56:20
2003-03-21 12:56:22	91	91	92.8	
2003-03-21 12:56:26	91	86.2	92.8	
2003-03-21 12:56:31	91	85.3	92.8	
2003-03-21 12:56:34	91	85.8	92.8	
2003-03-21 12:56:44	91	84.8	92.8	
2003-03-21 12:56:50	91	83.8	92.8	
2003-03-21 12:56:54	91	88	92.8	
2003-03-21 12:57:04	88	90	90.8	τ_2 =2003-03-21 12:57:04
2003-03-21 12:57:06	90	83	90.8	τ_3 =2003-03-21 12:57:06
2003-03-21 12:57:36	90	83	92	

With this data example as an illustration we now introduce possible explanatory variables and define notations that we will maintain throughout the following chapters. Histograms of all of the possible explanatory variables are attached at the end of this section.

Last quoted Bid-Ask spread

This variable, which we denote bas_t, is defined as the Bid-Ask-spread, i.e. Ask price - Bid price, of the quotation directly preceding the price-changing transaction. In the example bas_t has the value $6.3 = 91 - 84.7$.

Number of quotations between two consecutive price changes

This variable, denoted by NQ_t, counts the number of new quotations between the price changes with indices $t-1$ and t. In the example the value of NQ_{t+1} is 7 as there are 7 quotations between the two consecutive changes of the last price of the option.

Time lag between two consecutive price changes

Denoted by $DeltaTm_t$, this explanatory variable measures the time between two consecutive price changes of the option. In the example $DeltaTm_{t+1}$ is equal to 44 seconds, the time between the two price changes. This yields a value of 0.733 [minutes] for $DeltaTm_{t+1}$.

Intrinsic value of the option at the time of the transaction

We have already defined the intrinsic value of a Call option at the point of time τ as $max(S_\tau - K, 0)$ in Chapter 2. We now define the variable IV_t in a slightly different way by

$$IV_t := S_\tau - K$$

where τ denotes the point of time of the price change t. This definition accounts for 'how far' the Call option is out-of-the-money and therefore provides more detailed information than the usual definition of the intrinsic value. In the example IV_t has the value $88.37 = 2688.37 - 2600$ since the XETRA DAX index was quoted at 2688.37 points at the point of time τ_1.

Absolute price change of underlying since previous price change of option

This variable, denoted by $UnCh_t$, is simply defined as the absolute value of the difference between the price of the underlying at the point of time of the price change t and the price of the underlying at the point of time of the price change $t-1$.

Remaining time to maturity

This variable is denoted by TTM_t. It measures the time span from the point of time at which the t-th price change occurred and the time of maturity of the option. The option in the example matures on 21-March 13:00:00. Thus, for example, TTM_{t+2} is equal to 2min 54sec which yields a value of 0.0020139 [in days] for TTM_{t+2}.

In order to get an overview of these explanatory variables we show their histograms in Figure 5.8. The main range of the values of the Bid-Ask spread is, not taking into account some outlying larger values, between 0 and 10 index points. The range of the

number of new quotations between two consecutive option price changes is between 0 and 300, again discarding some outlying larger values. The histogram of the variable $DeltaTm$ shows that the times between consecutive price changes are rather short with most values smaller than 5 [minutes] and virtually no values larger than 40 [minutes]. This is not a surprise since options on the XETRA DAX index are among the most actively traded options at the EUREX exchange. The histogram of the explanatory variable IV shows that most of the price changes occurred while the option was out-of-the-money. The absolute price changes of the underlying DAX index between two consecutive option price changes lie mainly between 0 and 20 index points and finally, the histogram of the variable TTM confirms that the trading activity for this option rose significantly shortly before maturity of the option. We have already seen this in Figure 5.5.

To get a first impression of the relationship between these variables and the variable of interest one can take a look at the pairs plots of Figure 5.9. It can be observed that, as a very general rule, the larger each of the explanatory variables, the larger the absolute option price changes. In the case of the explanatory variable IV higher values of the explanatory variable are also associated with a higher variance of the absolute option price changes. We will discuss these relationships in much more detail in Chapter 6. Furthermore, the plots show that there is some correlation between the variables NQ and $DeltaTm$ which might cause problems of multicollinearity in the regression modelling in the following chapters. The plots for the relationship between the intrinsic value of the option IV and the time to maturity TTM replicate the chart of the underlying XETRA DAX index during the time period we consider. This is, of course, simply a consequence of the definition of these variables.

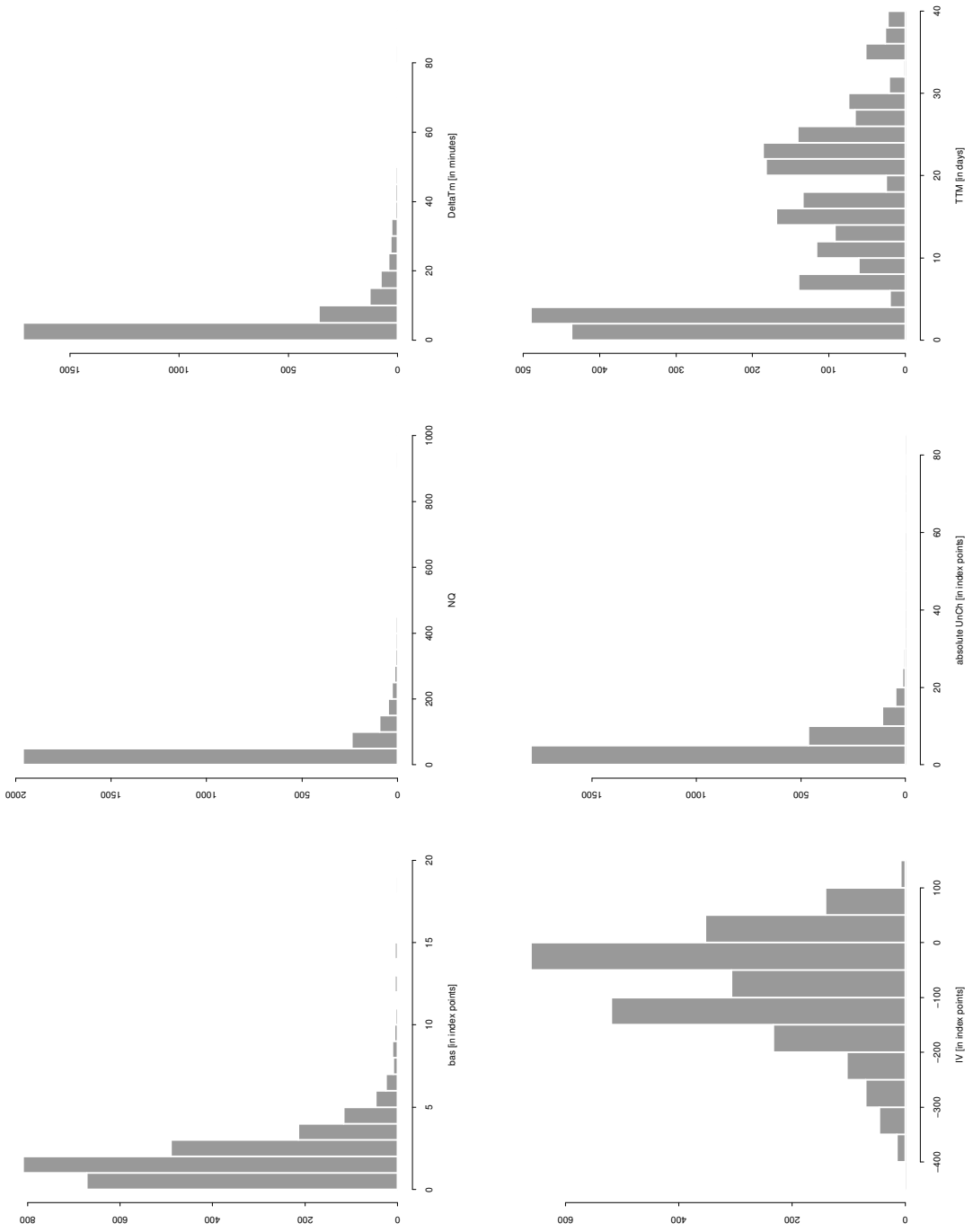

Figure 5.8: Histograms of explanatory variables.

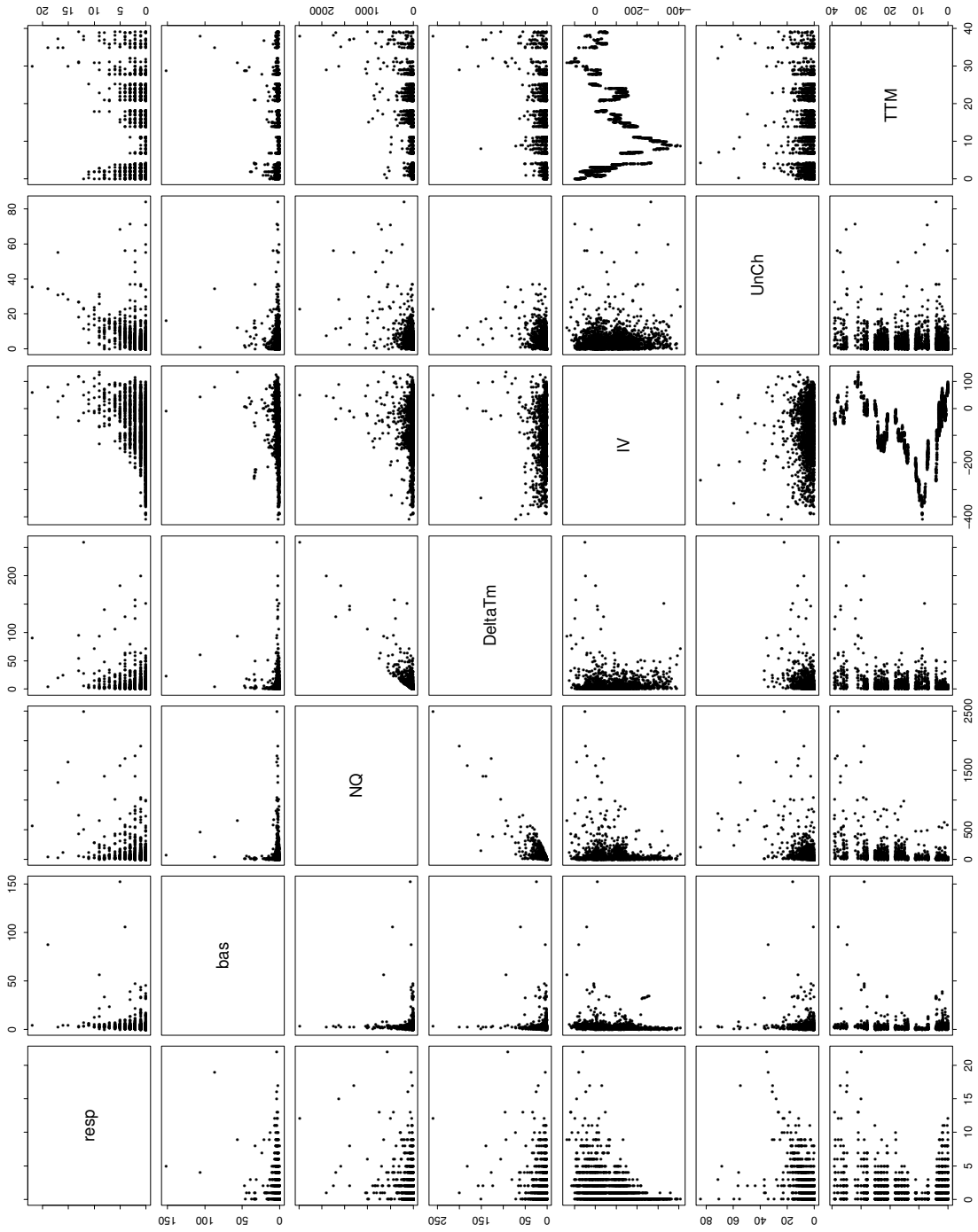

Figure 5.9: Pairs-plot of variable of interest (=resp) and explanatory variables.

Chapter 6

Regression modelling of option price changes

We finally fit a series of regression models of different complexity to the option data presented in the previous chapter.
As already mentioned in Chapter 3, when building a regression model the regressors that enter into the model have to be determined, possibly after transformations and possibly taking into account interaction effects between the regressors. We first fit an ordinary Poisson-GLM in Section 6.1 and use standard techniques to determine main- and interaction effects that have a significant influence on the absolute price changes of our option example. We also investigate whether transformations of the explanatory variables (e.g. taking logarithms) improve the model fit. In order to be able to compare modelling results we then maintain the transformations of the regressors for the fit of the more complex regression models in Sections 6.2 to 6.4.

6.1 An ordinary Poisson-GLM

6.1.1 Building the model

In order to get a first impression of the influence and significance of the explanatory variables introduced in the previous chapter we divide each explanatory variable into 10%-quantiles and calculate the logarithm of the empirical mean of the corresponding observations of the variable of interest (the absolute option price changes) for each quantile separately. The results are shown in Figure 6.1.

The plots indicate that all explanatory variables have indeed some influence on the variable of interest. Further quantitative methods will be needed to assess the statistical

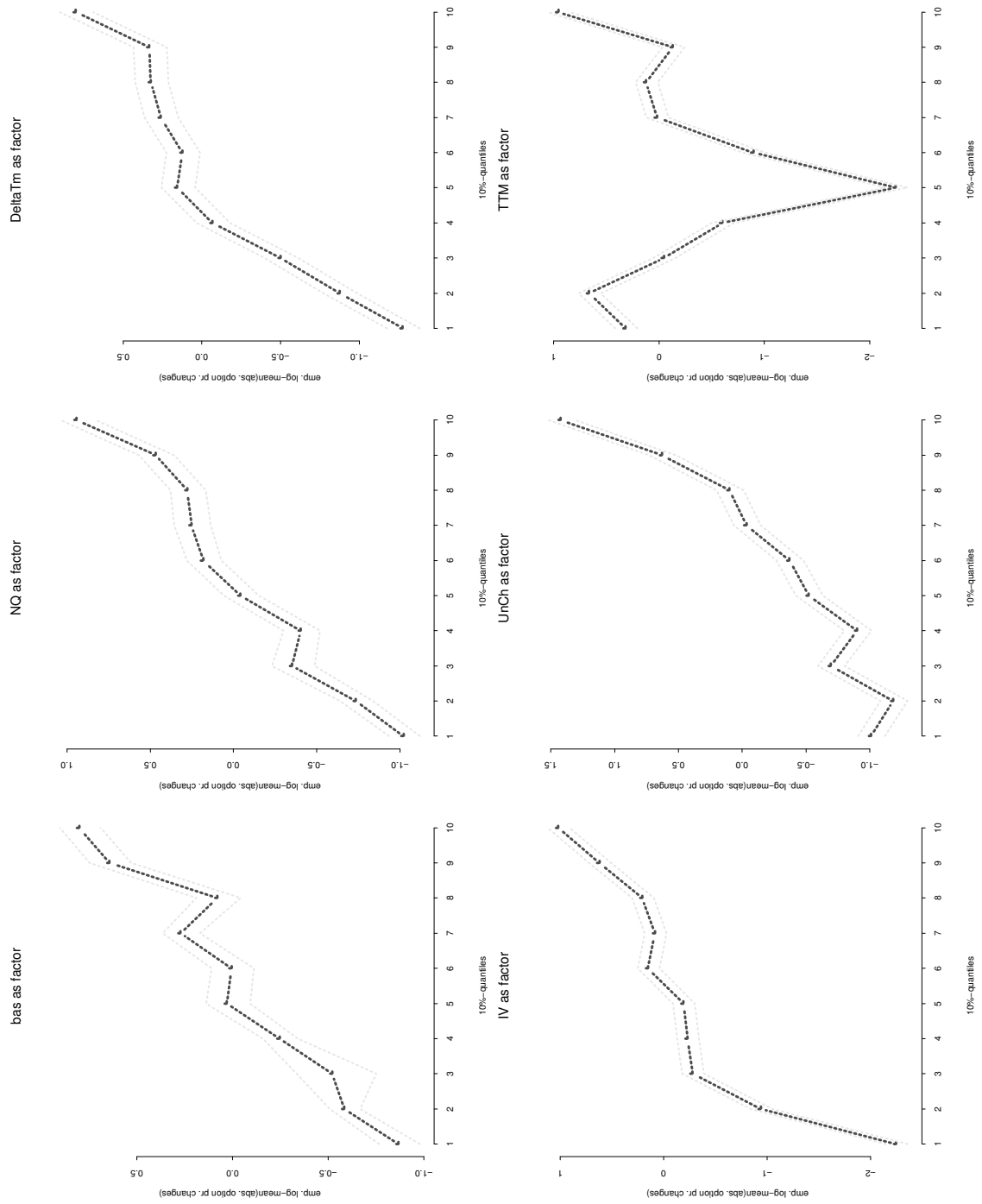

Figure 6.1: Main effects of explanatory variables on the empirical log-mean of the absolute price changes of the Call option on the XETRA DAX index with strike price 2600 and expiration month March 2003 (with 90%-confidence intervals).

significance of these effects. In addition to this, the non-linearity of the lines indicates that transformations of the explanatory variables, e.g. taking the logarithm, may lead to a better fit.
In the following we investigate log-, sqrt- and polynomial transformations for each explanatory variable. This is done by fitting simple Poisson-GLMs with only one explanatory variable. For each explanatory variable we fit four models of the form

$$Y_t \sim Poi(\lambda_t)$$

with

1. $\lambda_t = exp(\beta_0 + x_t\beta_1)$
2. $\lambda_t = exp(\beta_0 + log(x_t)\beta_1)$
3. $\lambda_t = exp(\beta_0 + \sqrt{x_t}\beta_1)$
4. $\lambda_t = exp(\beta_0 + x_t\beta_1 + x_t^2\beta_2)$

where Y_t denotes the t-th absolute option price change and x_t denotes the t-th observation of one particular explanatory variable. In the case of the logarithm it has to be ensured that $x_t > 0$ for all t. This can be done by adding a large enough constant c to x_t for $t = 1, ..., 2419$. The choice of c is of course arbitrary at this point. The same proceeding can be applied to ensure that $x_t \geq 0$ for all t in the case of the square root-transformation. We evaluate the transformations by comparing residual deviances of the models. It can be observed that $log(UnCh + c_1)$, $log(NQ + c_2)$, $log(DeltaTm)$, $log(bas)$, a polynomial of degree 2 for TTM and IV without any transformations lead to the best models. We use $c_1 = 80$, $c_2 = 1$. Once again, we point out that this particular choice for c_1 and c_2 does not have any physical motivation. The use of other values leads to very similar results in the modelling so that there is no reason to assume that the choice of c_1 and c_2 significantly influences the model building process. From now on we will use these transformations of the explanatory variables for the further modelling of the absolute option price changes.

As a next step in the model building process we take into account possible interaction effects between the regressors. Figure 6.2 shows some selected interactions effects (that will turn out to be significant in the further modelling) between the explanatory variables. In these plots all explanatory variables were divided into quartiles and for each quartile the empirical log-mean of the corresponding absolute option price changes was calculated, again separately according to the corresponding quartiles of the second explanatory variable. If there were no interaction effects between the explanatory variables the lines in the plots would have to show a parallel pattern.
We have to investigate the significance of a total of 6 possible explanatory variables and 15 possible interaction terms. We carry out a stepwise regressor selection using the *step-*

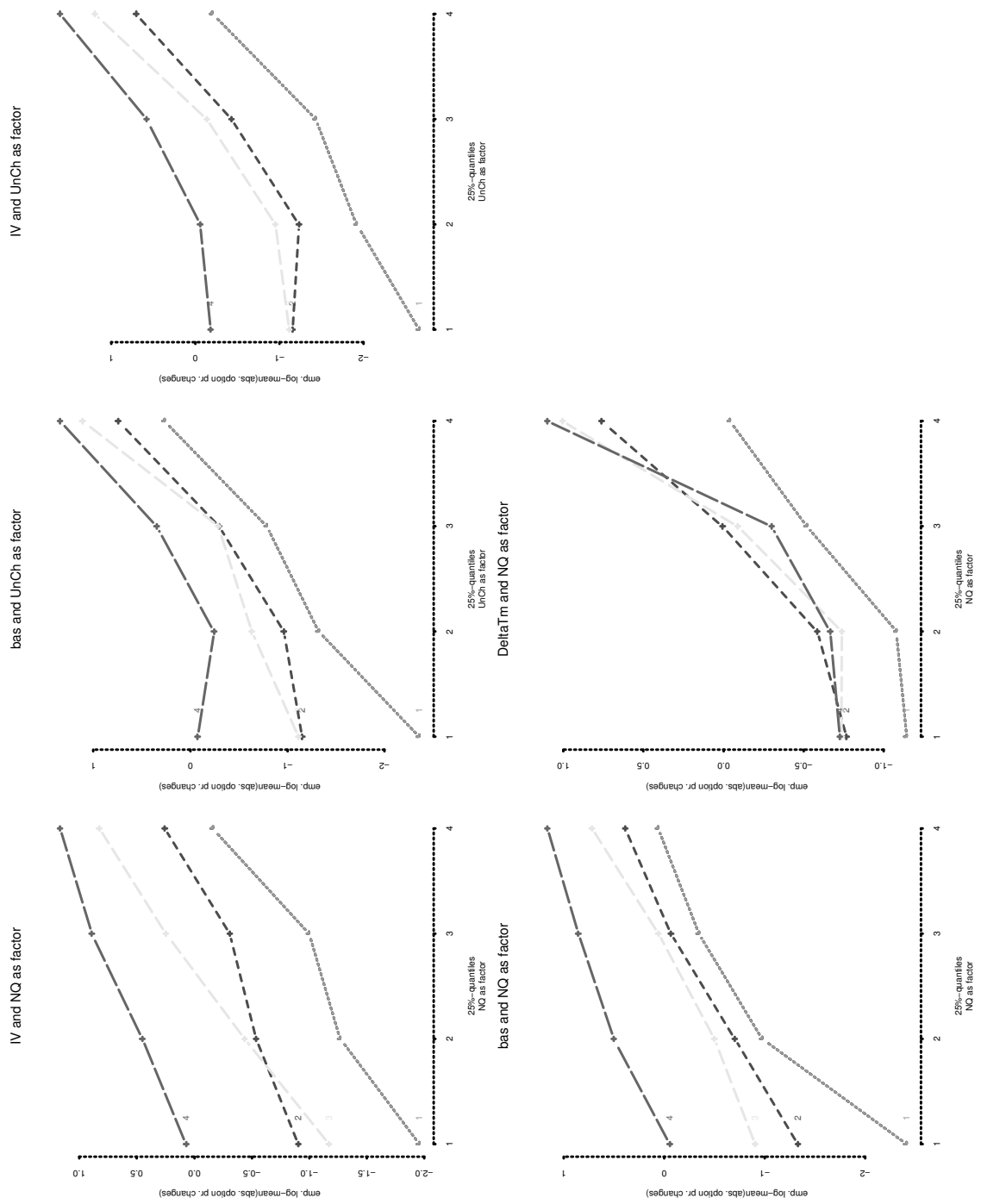

Figure 6.2: Selected interaction effects between explanatory variables for the absolute price changes of the Call option on the XETRA DAX index with strike price 2600 and expiration month March 2003.

function in S-Plus to determine significant regressors and their possible interactions. All explanatory variables with the exception of the remaining time to maturity TTM turn out to be significant and thus enter into the model. In order to improve the numerical stability of the estimation processes (especially for the more sophisticated models in the following sections) we standardize all regressors by subtracting the empirical mean and dividing through the empirical standard deviation of each regressor. Table 6.1 shows the estimation results of the final model. Note that for the last significant interaction term, the interaction between NQ and $DeltaTm$, Figure 6.2 shows no plot since not all of the 16 possible quartile-combinations between these two explanatory variables contain data.

	Parameter	Est.	Std. error	t-value
main effects	β_1 (Intercept)	-0.51	0.03	-16.03
	β_2 (stand. $log(UnCh + c_1)$)	0.99	0.04	23.95
	β_3 (stand. IV)	-2.50	1.41	-1.77
	β_4 (stand. $log(NQ + c_2)$)	0.29	0.07	4.36
	β_5 (stand. $log(bas)$)	9.64	0.93	10.37
	β_6 (stand. $log(DeltaTm)$)	10.06	1.41	7.15
interaction effects	β_7 (stand. $IV : log(NQ + c_2)$)	-0.28	0.06	-4.38
	β_8 (stand. $log(bas) : log(UnCh + c_1)$)	-9.29	0.93	-9.98
	β_9 (stand. $IV : log(UnCh + c_1)$)	3.46	1.43	2.42
	β_{10} (stand. $log(bas) : log(NQ + c_2)$)	-0.13	0.05	-2.73
	β_{11} (stand. $log(DeltaTm) : log(UnCh + c_1)$)	-9.91	1.42	-7.00
	β_{12} (stand. $log(NQ + c_2) : log(DeltaTm)$)	-0.14	0.05	-2.94

Table 6.1: Estimation results of the Poisson-GLM for the absolute price changes of the Call option on the XETRA DAX index with strike price 2600 and expiration month March 2003, obtained after stepwise regressor selection using standardized regressors ($c_1 = 80$, $c_2 = 1$).

6.1.2 Illustration and discussion of results

Although we will discuss the goodness-of-fit of the model in a more detailed way and in comparison to the other models in Chapter 7, we mention at this point that the residual deviance as introduced in Chapter 3.1 is 2418 on 2407 degrees of freedom in the Poisson-GLM of Table 6.1. Since the null deviance (i.e. the deviance of a Poisson-GLM with the Intercept as only regressor) has a value of 5592 on 2418 degrees of freedom, it can already be concluded that the overall fit of the Poisson-GLM is reasonably good.

Since it is not straightforward to interpret the estimation results in a model with variable transformations and interaction effects we draw fitted regression surfaces (Figure 6.3) where each plot shows the expected absolute price changes of the option as a function of two of the explanatory variables on a linear scale between the respective 10%- and 90%-quantiles. The expected absolute option price change is calculated according to the results of Table 6.1 with all other explanatory variables set to their median values.

Figure 6.3 shows that, as a very general rule, it can be said that higher values of all of the explanatory variables ($UnCh$, IV, NQ, bas and $DeltaTm$) lead to higher expected absolute option price changes. This is in line with the empirical results shown in Figure 6.1. Particularly high absolute option price changes can be expected when the absolute price change of the underlying (since the previous option price change) is high and if at the same time the option is clearly in-the-money (i.e. its intrinsic value is > 0). The plot for these two explanatory variables shows a clearly non-linear pattern with high expected absolute option price changes for high values of IV and $UnCh$. The plots moreover confirm that the explanatory variable $UnCh$ has the greatest impact on the expected absolute option price change, which is not a surprise taking into account economic considerations. We will discuss economic interpretations of the modelling results in more detail in Chapter 8.

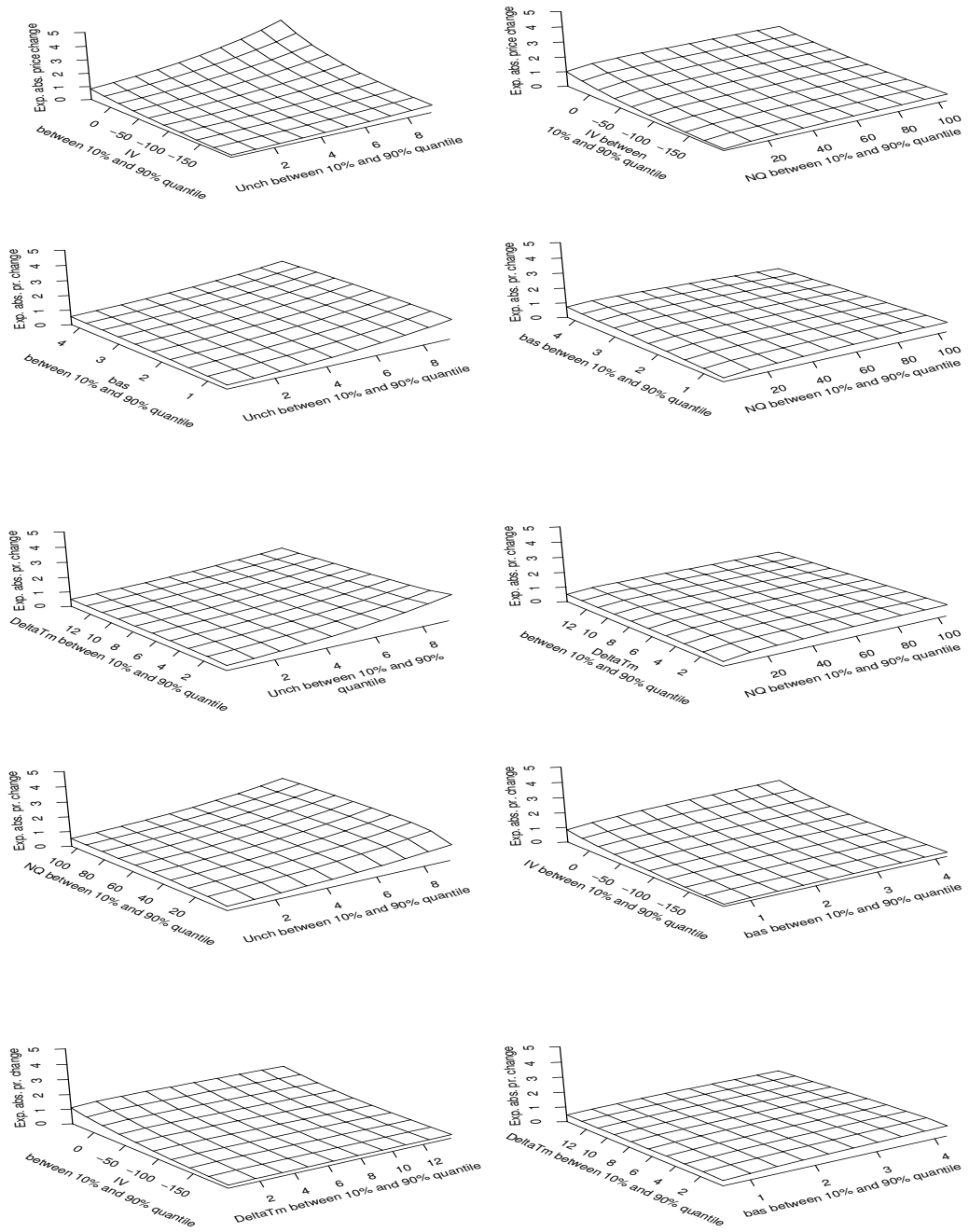

Figure 6.3: Fitted regression surfaces of the Poisson-GLM for the absolute price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003 when two regressors vary and the remaining regressors are set to their median values.

6.2 A Negative Binomial model

6.2.1 Implementation and estimation in S-Plus

As already elaborated in Chapter 3.1 Negative Binomial models can be considered as an extension of ordinary Poisson-GLMs since they allow for overdispersion of the data through the additional parameter α. We use the freely available MASS-library with the function *glm.nb* to fit a NegBin-GLM of the form (3.15) in S-Plus. The software can be downloaded from the website:

www.stats.ox.ac.uk/pub/MASS3/Software.html

Estimation of the parameters is carried out with maximum-likelihood-methods. We use exactly the same regressors as for the Poisson-GLM in order to make results comparable. The estimation results are shown in Table 6.2.

	Parameter	Est.	Std. error	t-value
main effects	β_1 (Intercept)	-0.53	0.03	-15.79
	β_2 (stand. $log(UnCh + c_1)$)	1.08	0.05	21.51
	β_3 (stand. IV)	-4.10	1.68	-2.44
	β_4 (stand. $log(NQ + c_2)$)	0.31	0.07	4.18
	β_5 (stand. $log(bas)$)	10.96	1.27	8.65
	β_6 (stand. $log(DeltaTm)$)	9.84	1.75	5.64
inter-action effects	β_7 (stand. $IV : log(NQ + c_2)$)	-0.25	0.07	-3.58
	β_8 (stand. $log(bas) : log(UnCh + c_1)$)	-10.59	1.27	-8.32
	β_9 (stand. $IV : log(UnCh + c_1)$)	5.03	1.70	2.97
	β_{10} (stand. $log(bas) : log(NQ + c_2)$)	-0.16	0.05	-2.97
	β_{11} (stand. $log(DeltaTm) : log(UnCh + c_1)$)	-9.69	1.76	-5.51
	β_{12} (stand. $log(NQ + c_2) : log(DeltaTm)$)	-0.17	0.05	-3.06
	α	9.55	1.99	

Table 6.2: Estimation results of the Negative Binomial model for the absolute price changes of the Call option on the XETRA DAX index with strike price 2600 and expiration month March 2003 ($c_1 = 80$, $c_2 = 1$).

Comparing the estimation results of the Negative Binomial-GLM (Table 6.2) to those of the Poisson-GLM (Table 6.1) it can be observed that the estimates for the parameter vector $\boldsymbol{\beta}$ are fairly similar in both models.

The estimate of the parameter α in the Negative Binomial model and its estimated standard error indicate, however, that there is statistical evidence for a slight overdispersion in the data. According to (3.16) the variance of the absolute option price changes given

the regressors is estimated by:

$$\begin{aligned} \widehat{Var}(Y_t|\mathbf{x}_t) &= \hat{\lambda}_t + \frac{1}{\hat{\alpha}}\hat{\lambda}_t^2 \\ &= exp(\mathbf{x}_t^t\hat{\boldsymbol{\beta}}) + \frac{1}{9.55}\left(exp(\mathbf{x}_t^t\hat{\boldsymbol{\beta}})\right)^2. \end{aligned} \tag{6.1}$$

6.2.2 Implementation in WinBUGS with the Ones-and-Bernoulli-trick

An alternative to fitting the Negative Binomial model (3.15) with the *glm.nb* function in S-Plus is MCMC estimation in WinBUGS. Yet, the general form of the Negative Binomial distribution with the density

$$P(Y_t = y_t) = \frac{\Gamma(y_t+\alpha)}{\Gamma(\alpha)y_t!}\left(\frac{\lambda_t}{\alpha+\lambda_t}\right)^{y_t}\left(\frac{\alpha}{\alpha+\lambda_t}\right)^{\alpha}$$

for $y = 0, 1, 2, 3, ...$ and $\alpha, \lambda_t > 0$, is not pre-specified in WinBUGS. However, a new sampling distribution can be specified with the so-called "Ones-and-Bernoulli-trick". To see how the trick for the Negative Binomial distribution works, suppose that we have for $t = 1, 2, ..., T$ observations of a random variable $Y_t \sim NegBin(\lambda_t, \alpha)$. Recall from Chapter 3.3 that

$$p(\boldsymbol{\lambda}, \alpha|\mathbf{y}) \propto f(\mathbf{y}; \boldsymbol{\lambda}, \alpha) \cdot p(\boldsymbol{\lambda}, \alpha)$$

where $p(\boldsymbol{\lambda}, \alpha|y_t)$ denotes the posterior density, $f(\mathbf{y}; \boldsymbol{\lambda}, \alpha)$ the likelihood of the model and $p(\boldsymbol{\lambda}, \alpha)$ the prior density. For $t = 1, ..., T$ we introduce new random variables $ones_t$, set $ones_t = 1$ for all t and assume that for each t $ones_t$ is a realization of a $Bern(p_t)$ distribution. The vector **ones** can thus be considered as an "extreme" realization of T independent Bernoulli random variables. We then set

$$p_t = c \cdot \frac{\Gamma(y_t+\alpha)}{\Gamma(\alpha)y_t!}\left(\frac{\lambda_t}{\alpha+\lambda_t}\right)^{y_t}\left(\frac{\alpha}{\alpha+\lambda_t}\right)^{\alpha} \quad t = 1, ..., T,$$

where c is a constant small enough to ensure that $0 < p_t < 1$ for all t. In the Bernoulli model with the vector **ones** we therefore obtain a posterior distribution for the parameters $\boldsymbol{\lambda}, \alpha$ of the following form:

$$p(\boldsymbol{\lambda}, \alpha|\mathbf{ones}, \mathbf{y}) \propto \left(\prod_{t=1}^{T} p_t^1 \cdot (1-p_t)^0\right) \cdot p(\boldsymbol{\lambda}, \alpha).$$

Due to the definition of p_t this is equal to:

$$p(\boldsymbol{\lambda}, \alpha|\mathbf{ones}, \mathbf{y}) \propto c^T \cdot \left[\prod_{t=1}^{T} \frac{\Gamma(y_t+\alpha)}{\Gamma(\alpha)y_t!}\left(\frac{\lambda_t}{\alpha+\lambda_t}\right)^{y_t}\left(\frac{\alpha}{\alpha+\lambda_t}\right)^{\alpha}\right] \cdot p(\boldsymbol{\lambda}, \alpha).$$

Thus, by sampling from the posterior distribution of $\boldsymbol{\lambda}, \alpha$ in the Bernoulli model with **ones** we get a posterior sample from a density which is proportional to the product of the $NegBin$-likelihood

$$f_{\text{NegBin}}(\mathbf{y}; \boldsymbol{\lambda}, \alpha) = \prod_{t=1}^{T} \frac{\Gamma(y_t + \alpha)}{\Gamma(\alpha) y_t!} \left(\frac{\lambda_t}{\alpha + \lambda_t} \right)^{y_t} \left(\frac{\alpha}{\alpha + \lambda_t} \right)^{\alpha}$$

and the prior

$$p(\boldsymbol{\lambda}, \alpha) = \prod_{t=1}^{T} p(\lambda_t, \alpha),$$

i.e. a sample from the posterior density of the parameters $\boldsymbol{\lambda}, \alpha$ in the Negative Binomial model. The examples in the WinBUGS user manual point out, however, that this method can be quite inefficient and lead to high standard errors of the estimates.

For our simulation we choose the rather uninformative priors

$$\begin{aligned} \beta_i &\sim N(0, 1000) \quad \text{for } i = 1, ..., 12 \\ \alpha &\sim \Gamma(0.1, 0.1) \end{aligned}$$

and run 3 chains for each parameter with different starting values and 100,000 iterations. It can be observed that convergence of the chains is very slow with very high autocorrelations. We therefore only store every 100th iteration and choose a burn-in of 500 stored iterations. The trajectories of the recorded iterations as well as the Gelman-Rubin convergence statistic plots (Figures 6.4 to 6.10) indicate that this is an acceptable approach for most of the parameters. The results of the MCMC estimation are shown in Table 6.3. They are fairly similar to those obtained by maximum-likelihood estimation in S-Plus (Table 6.2).

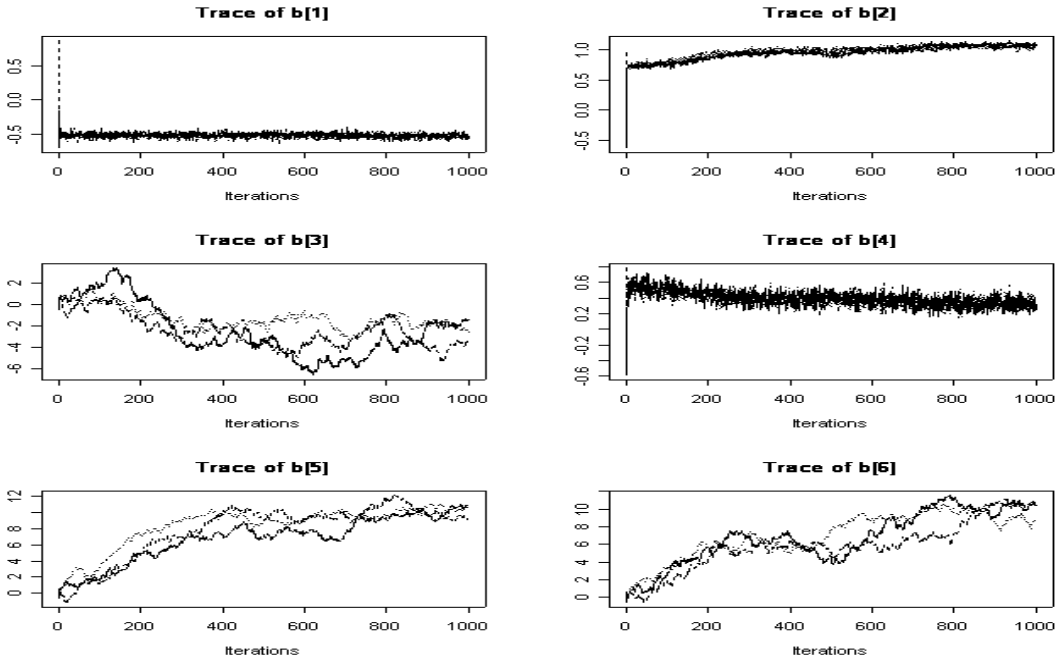

Figure 6.4: MCMC iterations for $\beta_1, ..., \beta_6$ (b[1],...,b[6]) for the Negative Binomial model.

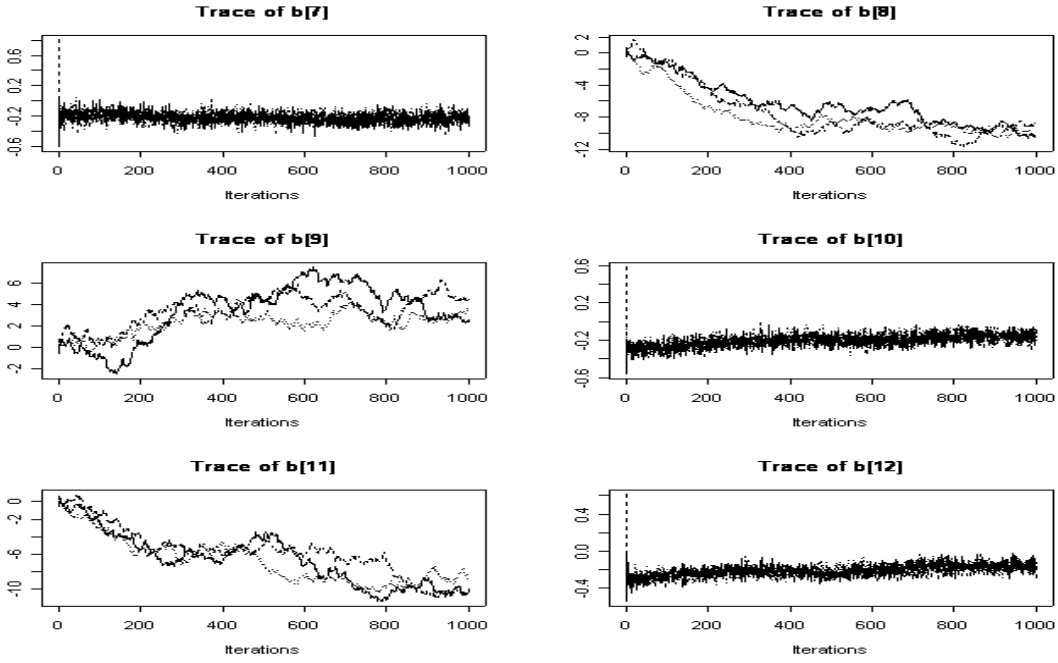

Figure 6.5: MCMC iterations for $\beta_7, ..., \beta_{12}$ (b[7],...,b[12]) for the Negative Binomial model.

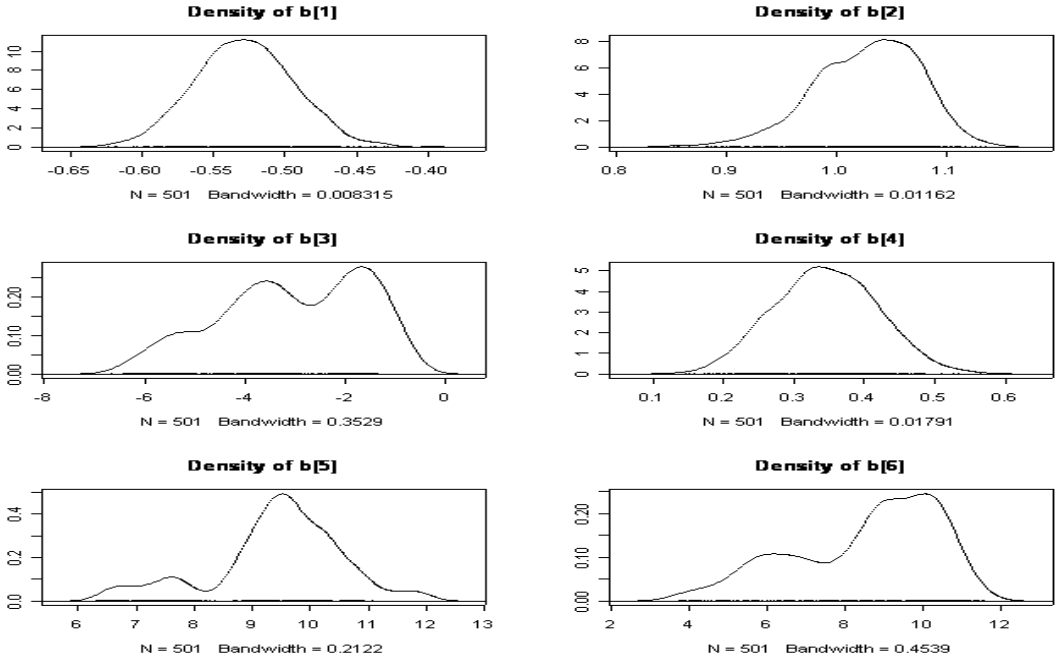

Figure 6.6: Posterior density estimates for $\beta_1, ..., \beta_6$ (b[1],...,b[6]) in the Negative Binomial model based on the last 500 recorded iterations in each chain.

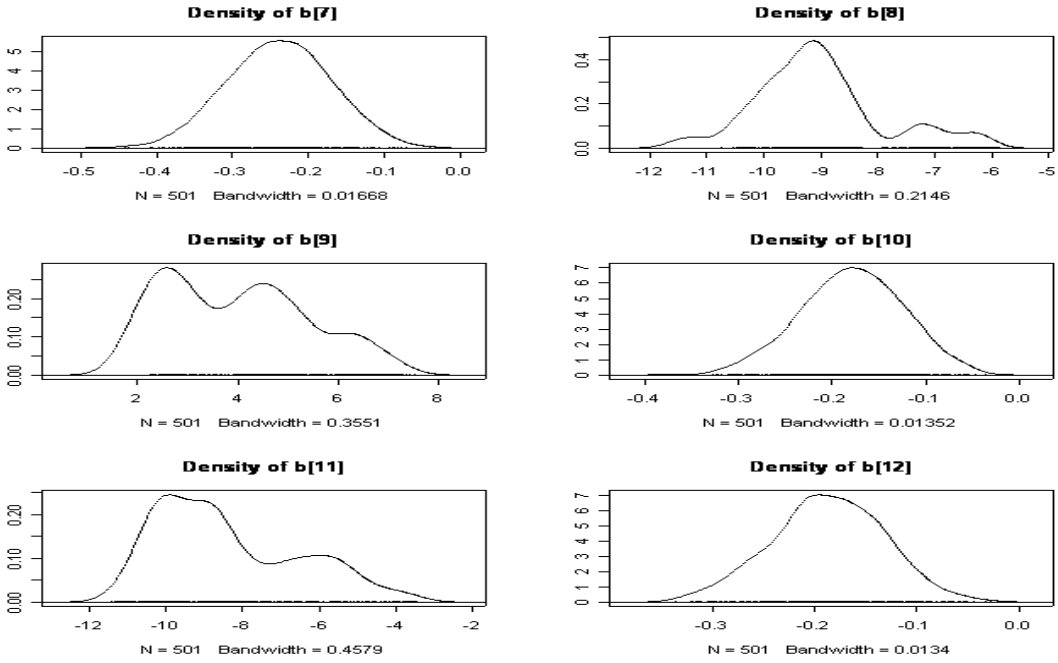

Figure 6.7: Posterior density estimates for $\beta_7, ..., \beta_{12}$ (b[7],...,b[12]) in the Negative Binomial model based on the last 500 recorded iterations in each chain.

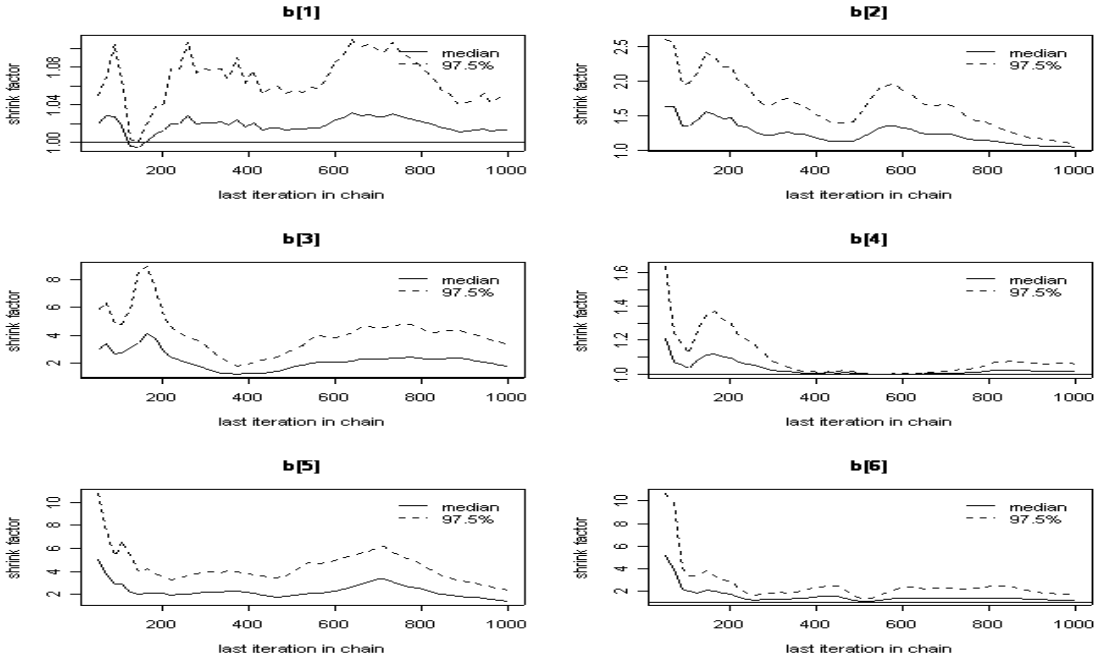

Figure 6.8: Gelman-Rubin convergence diagnostics for $\beta_1, ..., \beta_6$ (b[1],...,b[6]).

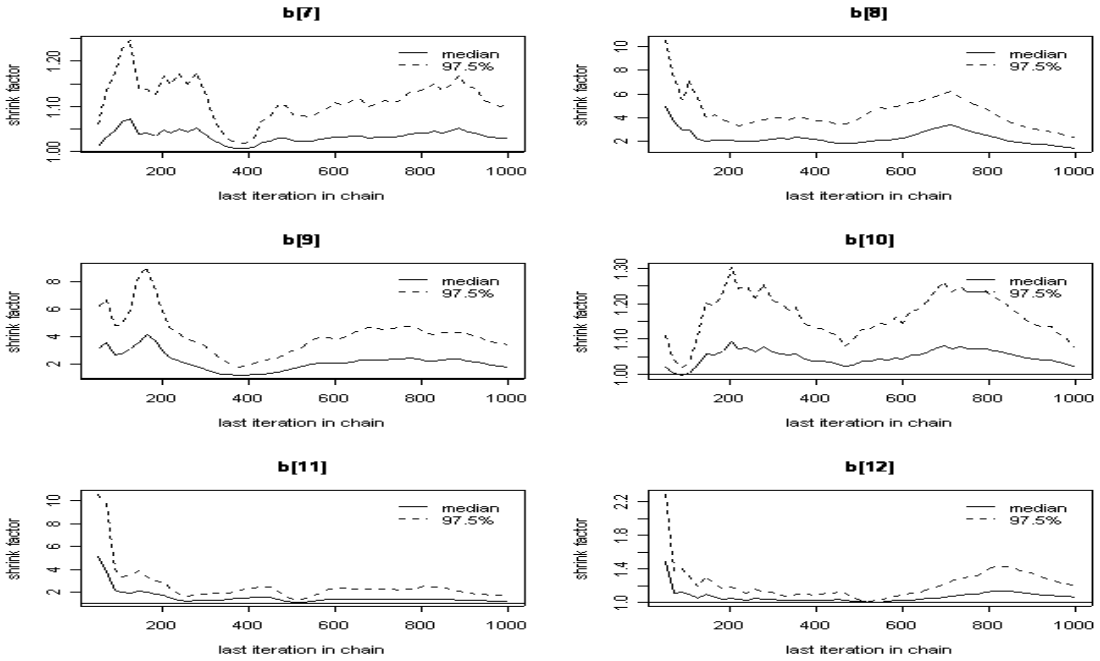

Figure 6.9: Gelman-Rubin convergence diagnostics for $\beta_7, ..., \beta_{12}$ (b[7],...,b[12]).

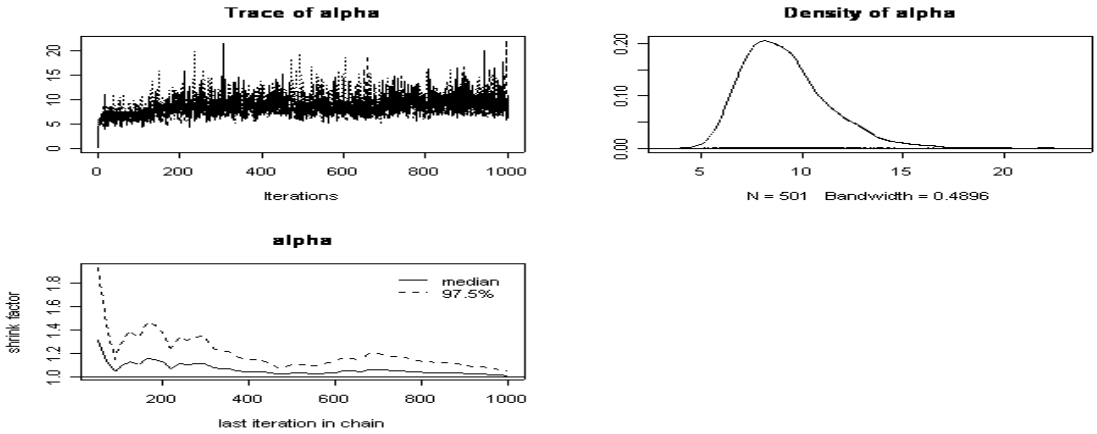

Figure 6.10: MCMC iterations, posterior density estimate and Gelman-Rubin convergence diagnostics for α in the Negative Binomial model.

	Parameter	mean	Std. err.	2.5%	med.	97.5%
main effects	β_1 (Intercept)	-0.53	0.03	-0.59	-0.53	-0.46
	β_2 (stand. $log(UnCh + c_1)$)	1.03	0.05	0.93	1.03	1.11
	β_3 (stand. IV)	-3.05	1.44	-5.92	-3.02	-0.92
	β_4 (stand. $log(NQ + c_2)$)	0.35	0.07	0.20	0.34	0.49
	β_5 (stand. $log(bas)$)	9.43	1.13	6.71	9.55	11.59
	β_6 (stand. $log(DeltaTm)$)	8.53	1.85	4.55	8.97	11.06
inter-action effects	β_7 (stand. $IV : log(NQ + c_2)$)	-0.24	0.07	0.20	0.34	0.49
	β_8 (stand. $log(bas) : log(UnCh + c_1)$)	-9.04	1.14	-11.19	-9.16	-6.29
	β_9 (stand. $IV : log(UnCh + c_1)$)	3.97	1.45	1.85	3.97	6.90
	β_{10} (stand. $log(bas) : log(NQ + c_2)$)	-0.18	0.06	-0.29	-0.18	-0.07
	β_{11} (stand. $log(DeltaTm) : log(UnCh + c_1)$)	-8.38	1.87	-10.94	-8.83	-4.39
	β_{12} (stand. $log(NQ + c_2) : log(DeltaTm)$)	-0.18	0.05	-0.30	-0.19	-0.08
	α	9.18	2.12	6.13	8.84	14.20

Table 6.3: MCMC sampling results for the Negative Binomial model for the absolute price changes of the Call option on the XETRA DAX index with strike price 2600 and expiration month March 2003 based on the last 500 recorded iterations in each chain ($c_1 = 80$, $c_2 = 1$).

6.3 A Poisson-GLM with an AR(1) latent process in the mean

We now take into account the time series structure of the data and fit a Poisson-GLM with an AR(1) latent process in the mean to our option data. The model as discussed in Chapter 4 is given by

$$\begin{aligned} Y_t &\sim Poi(\lambda_t) && (6.2)\\ \lambda_t &= exp(Z_t) && (6.3)\\ Z_t &= \mathbf{x}_t^t\boldsymbol{\beta} + u_t && (6.4)\\ u_t &= \rho \cdot u_{t-1} + \varepsilon_t && (6.5)\\ \varepsilon_t &\overset{iid}{\sim} N(0, \sigma_\varepsilon^2) \end{aligned}$$

In order to implement the model in WinBUGS the following transformations have to be made: From Equations (6.4) and (6.5) it follows that for $t \geq 2$

$$\begin{aligned} Z_t &= \mathbf{x}_t^t\boldsymbol{\beta} + \rho u_{t-1} + \varepsilon_t \\ &= \mathbf{x}_t^t\boldsymbol{\beta} + \rho \cdot (Z_{t-1} - \mathbf{x}_{t-1}^t\boldsymbol{\beta}) + \varepsilon_t \\ &= (\mathbf{x}_t^t - \rho \cdot \mathbf{x}_{t-1}^t)\boldsymbol{\beta} + \rho \cdot Z_{t-1} + \varepsilon_t. \end{aligned}$$

Thus,

$$Z_t | Z_{t-1} \sim N\left((\mathbf{x}_t^t - \rho \cdot \mathbf{x}_{t-1}^t)\boldsymbol{\beta} + \rho \cdot Z_{t-1}, \sigma_\varepsilon^2\right).$$

Since, as shown in (3.28) and (3.29)

$$\begin{aligned} E(u_t) &= 0 \quad , \\ Var(u_t) &= \frac{\sigma_\varepsilon^2}{1-\rho^2} \qquad \text{for all } t, \end{aligned}$$

we get:

$$Z_1 \sim N\left(\mathbf{x}_1^t\boldsymbol{\beta}, \frac{\sigma_\varepsilon^2}{1-\rho^2}\right).$$

These considerations at hand, the model can be implemented in WinBUGS.
As priors we choose again uninformative normal priors for the components of $\boldsymbol{\beta}$, an uninformative Gamma-prior for the precision $\tau_\varepsilon := 1/\sigma_\varepsilon^2$ and a uniform prior on the interval $[-1, 1]$ for ρ:

$$\begin{aligned} \rho &\sim Unif[-1, 1] \\ \beta_i &\sim N(0, 1000) \quad \text{for } i = 1, ..., 12 \\ \tau_\varepsilon &\sim \Gamma(0.05, 0.05). \end{aligned}$$

We run 3 independent chains for each parameter and reduce autocorrelations between the iterations by only storing every 50th value. We record a total of 500 observations of each chain and consider the first 100 iterations as burn-in. The trajectories of the chains and the Gelman-Rubin statistic plots (see Figures 6.11 to 6.17) justify these choices. In a second simulation we investigate prior sensitivity for the parameter ρ and use an informative $N(0, 0.3^2)$ prior truncated to $[-1, 1]$ for ρ. The MCMC simulation leads to similar results (with slower convergence of the simulation), so that the prior sensitivity of ρ can assumed to be negligible.
Table 6.4 shows the results of the MCMC simulation for the Poisson-GLM with an AR(1) latent process in the mean fit to our option data (with a uniform prior for ρ).

	Parameter	mean	Std. err.	2.5%	med.	97.5%
main effects	β_1 (Intercept)	-0.59	0.09	-0.76	-0.59	-0.42
	β_2 (stand. $log(UnCh + c_1)$)	0.99	0.05	0.90	0.99	1.08
	β_3 (stand. IV)	-0.85	1.59	-3.95	-0.81	2.13
	β_4 (stand. $log(NQ + c_2)$)	0.34	0.07	0.21	0.34	0.48
	β_5 (stand. $log(bas)$)	8.89	1.13	6.65	8.92	11.06
	β_6 (stand. $log(DeltaTm)$)	8.52	1.50	5.50	8.55	11.35
interaction effects	β_7 (stand. $IV : log(NQ + c_2)$)	0.12	0.19	-0.23	0.11	0.50
	β_8 (stand. $log(bas) : log(UnCh + c_1)$)	-8.56	1.13	-10.73	-8.56	-6.35
	β_9 (stand. $IV : log(UnCh + c_1)$)	1.61	1.58	-1.35	1.57	4.66
	β_{10} (stand. $log(bas) : log(NQ + c_2)$)	-0.14	0.05	-0.24	-0.14	-0.05
	β_{11} (stand. $log(DeltaTm) : log(UnCh + c_1)$)	-8.26	1.52	-11.13	-8.30	-5.17
	β_{12} (stand. $log(NQ + c_2) : log(DeltaTm)$)	-0.22	0.05	-0.33	-0.22	-0.13
	ρ	0.97	0.02	0.93	0.97	0.99
	σ_ε^2	0.009	0.002	0.006	0.009	0.014

Table 6.4: MCMC sampling results for the Poisson-GLM with an AR(1) latent process for the absolute price changes of the Call option on the XETRA DAX index with strike price 2600 and expiration month March 2003 based on the last 400 recorded iterations in each chain ($c_1 = 80$, $c_2 = 1$).

Here, the parameters β_7 and β_9 are not $\neq 0$ on a 5%-credible level. This implies that in this model the interaction terms $IV : log(NQ + c_2)$ and $IV : log(UnCh + c_1)$ are not significant. Since β_3 is not $\neq 0$ on a 5%-credible level either, it has to be investigated whether the explanatory IV is significant at all in this model setting. For this purpose we repeat the MCMC simulation without the interaction terms $IV : log(NQ + c_2)$ and $IV : log(UnCh + c_1)$. This yields very similar estimates for all other parameters of the

model and a value of 0.80 for β_3 with a 95%-credible interval of $[0.72, 0.88]$. So, it can be concluded that there is a significant positive relationship between the intrinsic value of the option and the absolute option price changes in this model setting.
The estimate for ρ implies that there is indeed a very high autocorrelation in the latent process. However, the posterior mean estimate of the variance of the latent variables $\hat{\sigma}_u^2$ is equal to 0.19, so that, in general, the values of $\hat{u}_t \sim N(0, \hat{\sigma}_u^2)$ and as a consequence the contribution of the latent process to the estimated expectation of $\{Y_t\}$ (conditional on the regressors) can expected to be rather small. This can be confirmed by considering in our MCMC simulation the 2419 posterior mean estimates of u_t, $t = 1, ..., 2419$ which lie within the interval $[-0.84, 0.68]$, while those for $\mathbf{x}_t^t\boldsymbol{\beta}$ lie within the significantly larger interval $[-4.92, 3.68]$. The lowest lower bound of all 2419 95%-credible intervals of $\hat{u}_t$, $t = 1, ..., 2419$, is -1.59 while the largest upper bound of these credible intervals has the value 1.33. For the quantity $\mathbf{x}_t^t\hat{\boldsymbol{\beta}}$, which measures the influence of the explanatory variables, these values are -5.63 and 4.38 respectively.
Defining $\mu_t := exp(\mathbf{x}_t^t\boldsymbol{\beta})$ and denoting the posterior mean estimate of this quantity by $\hat{\mu}_t$, the fitted mean, variance and covariances of $\{Y_t\}$ (conditional on the regressors) can be estimated by using the posterior mean estimates $\hat{\sigma}_u^2$ and $\hat{\mu}_t$ for σ_u^2 and μ_t in (4.8), (4.12) and (4.13):

$$\begin{aligned}\widehat{E(Y_t|\mathbf{x}_t)} &= e^{0.5\hat{\sigma}_u^2} \cdot \hat{\mu}_t \\ &= 1.10 \cdot \hat{\mu}_t\end{aligned}$$

$$\begin{aligned}\widehat{Var}(Y_t|\mathbf{x}_t) &= e^{0.5\hat{\sigma}_u^2} \cdot \hat{\mu}_t + e^{\hat{\sigma}_u^2} \cdot \left(e^{\hat{\sigma}_u^2} - 1\right) \cdot \hat{\mu}_t^2 \\ &= 1.10 \cdot \hat{\mu}_t + 0.25 \cdot \hat{\mu}_t^2\end{aligned}$$

$$\begin{aligned}\widehat{Cov}(Y_t, Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}) &= \hat{\mu}_t \cdot \hat{\mu}_{t+k} \cdot e^{\hat{\sigma}_u^2} \cdot \left(e^{\hat{\rho}^k \cdot \hat{\sigma}_u^2} - 1\right) \\ &= \hat{\mu}_t \cdot \hat{\mu}_{t+k} \cdot e^{0.19} \cdot \left(e^{0.97^k \cdot 0.19} - 1\right),\end{aligned}$$

which yields

$$\widehat{Cov}(Y_t, Y_{t+1}|\mathbf{x}_t, \mathbf{x}_{t+1}) = \hat{\mu}_t \cdot \hat{\mu}_{t+k} \cdot 0.24$$

for $k = 1$.

In order to be able to interpret the estimation results we draw plots similar to those for the ordinary Poisson-GLM in Figure 6.3 where the expected absolute option price

changes are again shown as a functions of two of the explanatory variables on a linear, untransformed scale in each plot with all other explanatory variables set to their median values (Figure 6.19). These plots look very similar to those in Figure 6.3 and thus, the general conclusions about the influence of the explanatory variables on the absolute option price changes remain valid for the Poisson-GLM with an AR(1) latent process.

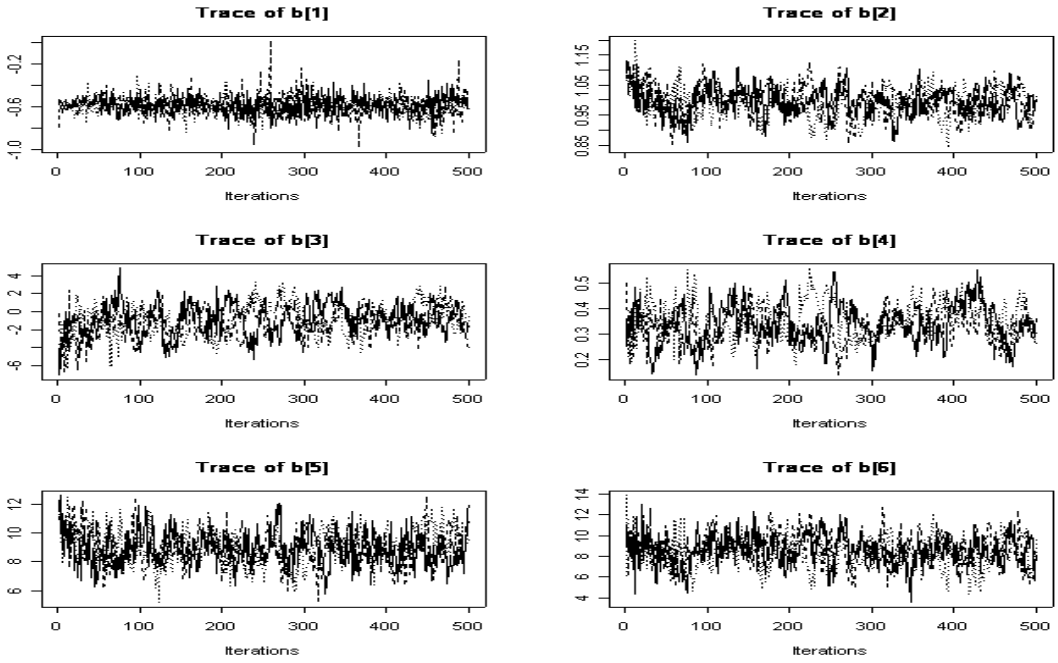

Figure 6.11: MCMC iterations for $\beta_1, ..., \beta_6$ (b[1],...,b[6]) for the Poisson-GLM with an AR(1) latent process.

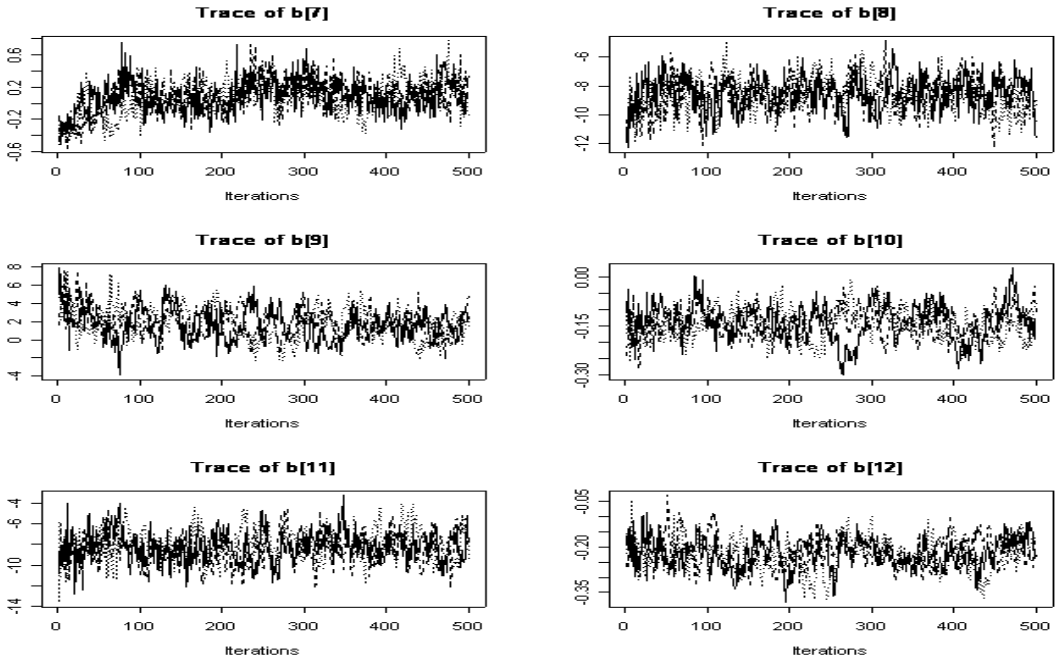

Figure 6.12: MCMC iterations for $\beta_7, ..., \beta_{12}$ (b[7],...,b[12]) for the Poisson-GLM with an AR(1) latent process.

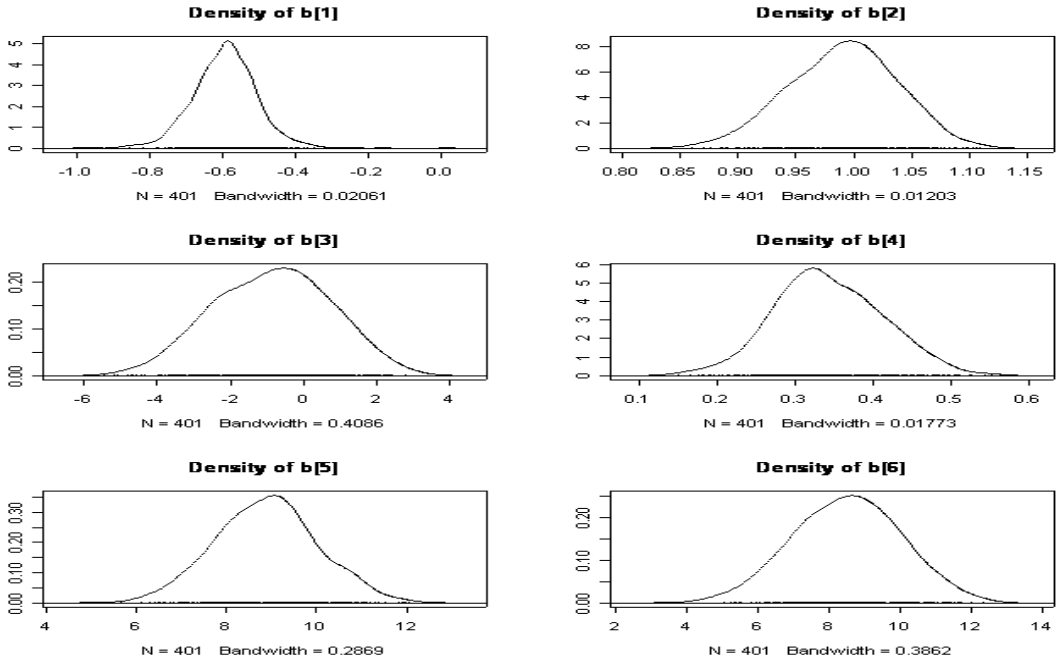

Figure 6.13: Posterior density estimates for $\beta_1, ..., \beta_6$ (b[1],...,b[6]) in the Poisson-GLM with an AR(1) latent process based on the last 400 recorded iterations in each chain .

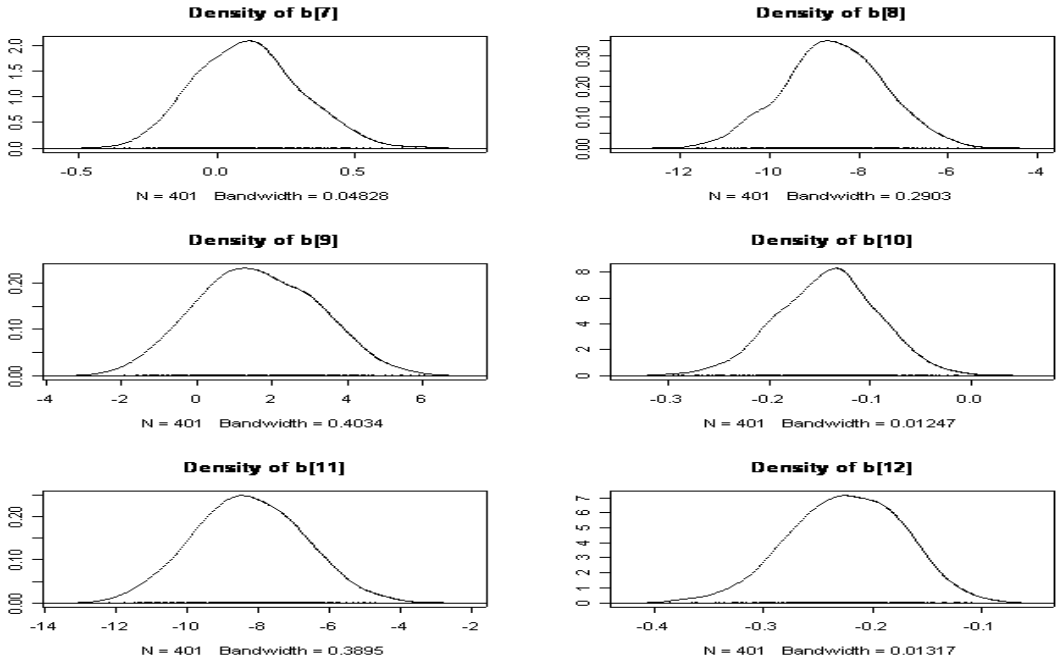

Figure 6.14: Posterior density estimates for $\beta_7, ..., \beta_{12}$ (b[7],...,b[12]) in the Poisson-GLM with an AR(1) latent process based on the last 400 recorded iterations in each chain .

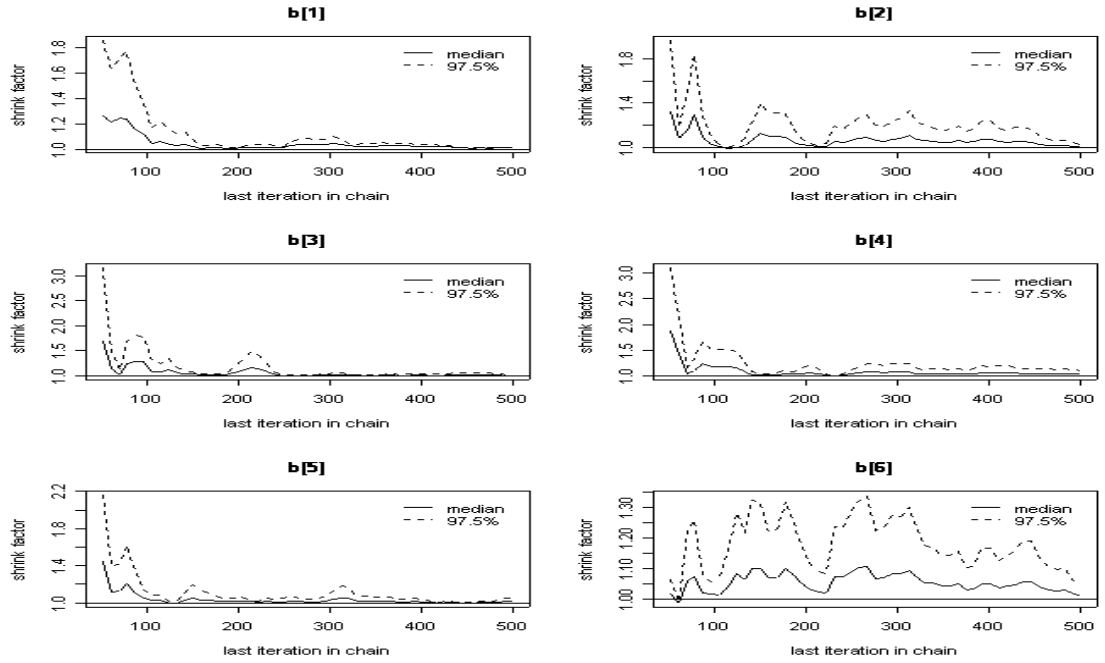

Figure 6.15: Gelman-Rubin convergence diagnostics for $\beta_1, ..., \beta_6$ (b[1],...,b[6]).

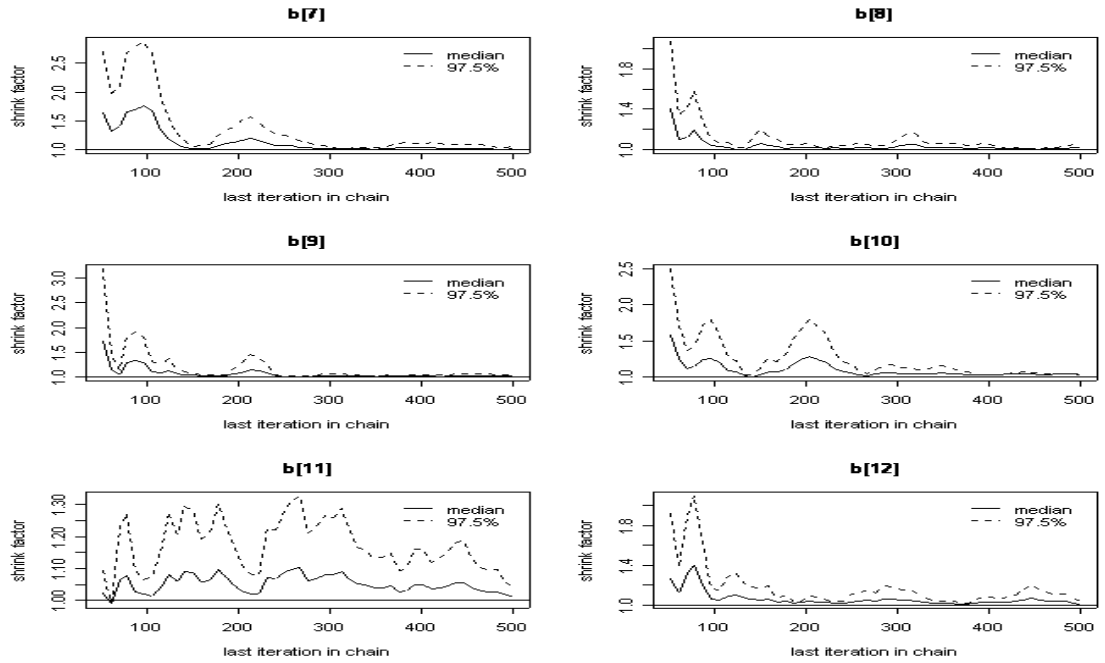

Figure 6.16: Gelman-Rubin convergence diagnostics for $\beta_7, ..., \beta_{12}$ (b[7],...,b[12]).

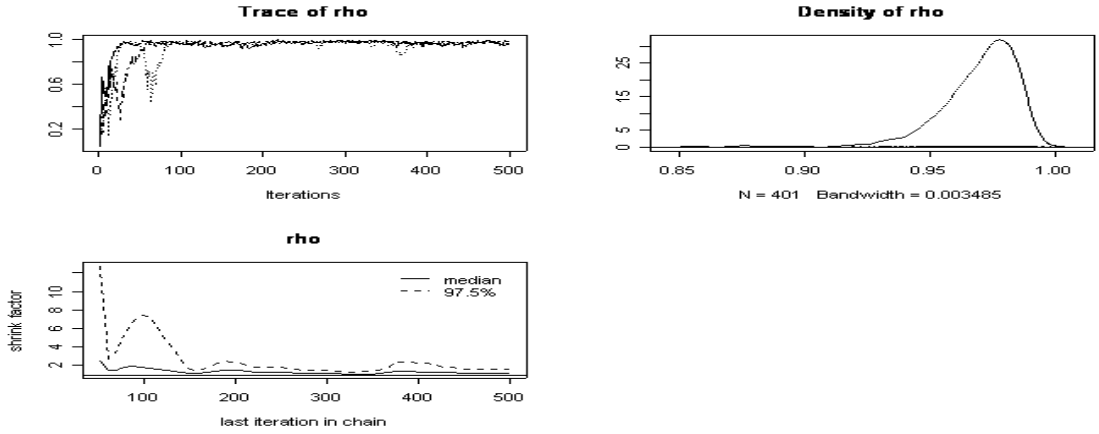

Figure 6.17: MCMC iterations, posterior density estimate and Gelman-Rubin convergence diagnostics for ρ in the Poisson-GLM with an AR(1) latent process.

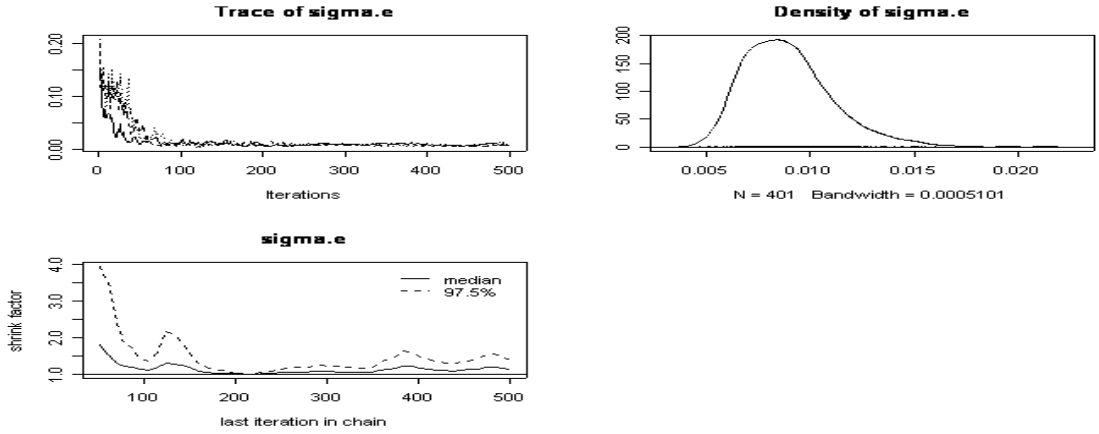

Figure 6.18: MCMC estimation and Gelman-Rubin convergence diagnostics for σ_ε^2 (sigma.e) in the Poisson-GLM with an AR(1) latent process.

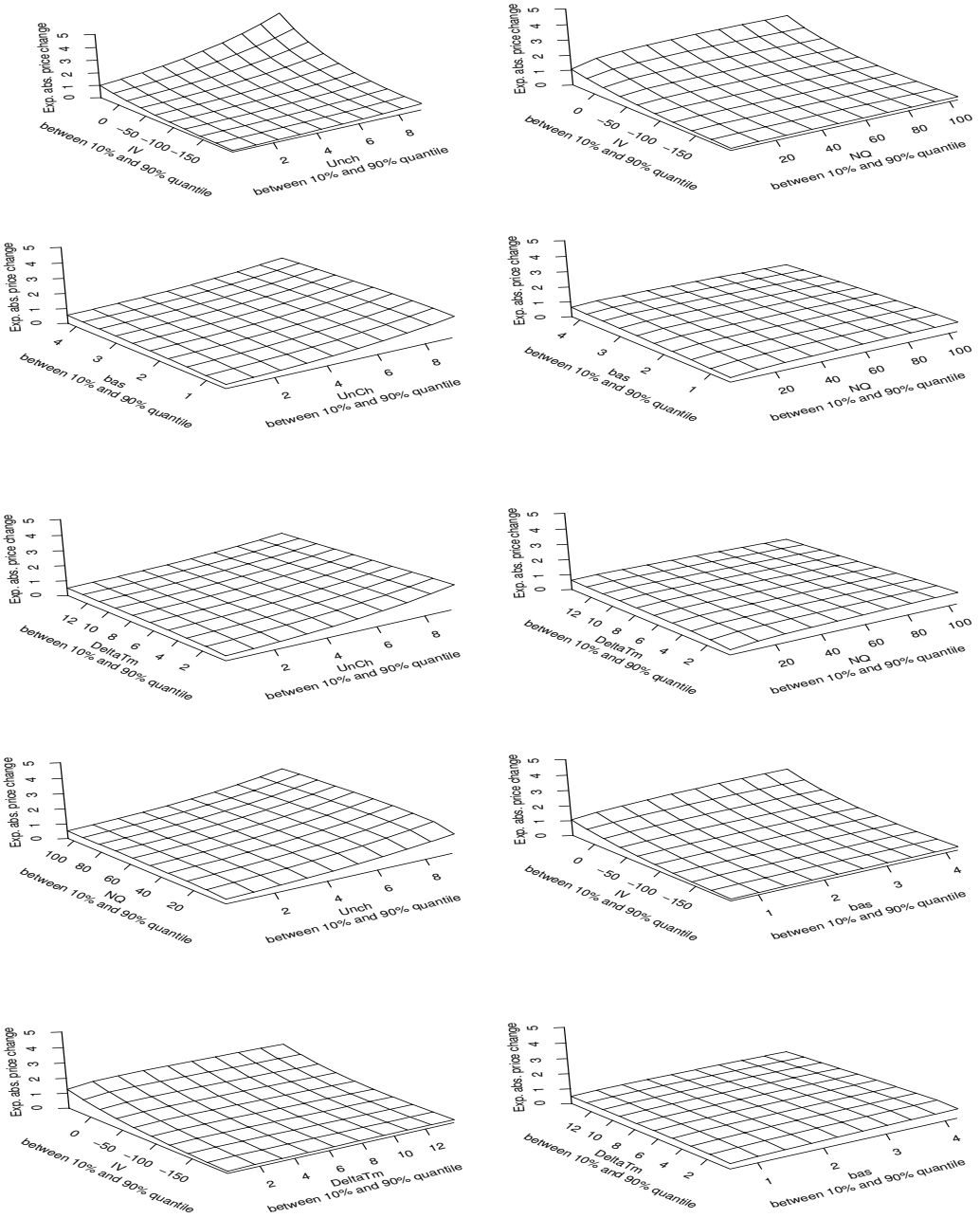

Figure 6.19: Fitted regression surfaces of the Poisson-GLM with an AR(1) latent process for the absolute price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003 when two regressors vary and the remaining regressors are set to their median values.

6.4 A Poisson-GLM with heterogeneous error term variance

In both the Negative Binomial model and in the Poisson-GLM with an AR(1) latent process the variance structure of $\{Y_t\}$ conditional on the regressors was given by

$$Var(Y_t|\mathbf{x}_t) = \lambda_t + c \cdot \lambda_t^2$$

where $\lambda_t := E(Y_t|\mathbf{x}_t)$ and c is a constant. The assumption of a homogeneous variance structure of this kind might be a too restrictive approach for our option data. We therefore fit another model which allows for a heterogeneous variance structure.
The model is given by:

$$\begin{aligned} Y_t &\sim Poi(\lambda_t) \\ \lambda_t &= exp(\mathbf{x}_t^t\boldsymbol{\beta} + u_t) \\ u_t &= exp\left(\frac{\tilde{\mathbf{x}}_t^t\boldsymbol{\alpha}}{2}\right) \cdot \varepsilon_t \\ \varepsilon_t &\overset{iid}{\sim} N(0,1). \end{aligned} \tag{6.6}$$

Since

$$u_t \sim N\left(0, exp(\tilde{\mathbf{x}}_t^t\boldsymbol{\alpha})\right)$$

it follows that $exp(u_t)$ is log-normally distributed and, using the common formulas for the expectation and variance of the log-normal distribution (already given in (4.5) and (4.6)), the expectation and the variance of Y_t conditional on the regressors in the model (6.6) can be calculated similarly to the calculation of these quantities in the Poisson-GLM with an AR(1) latent process in (4.8) and (4.12):

$$\begin{aligned} E(Y_t|\mathbf{x}_t) &= E(E(Y_t|u_t, \mathbf{x}_t)) \\ &= E(exp(u_t)) \cdot exp(\mathbf{x}_t^t\boldsymbol{\beta}) \\ &= exp\left(0.5e^{\tilde{\mathbf{x}}_t^t\boldsymbol{\alpha}} + \mathbf{x}_t^t\boldsymbol{\beta}\right) \end{aligned} \tag{6.7}$$

$$\begin{aligned} Var(Y_t|\mathbf{x}_t) &= E(Var(Y_t|\mathbf{x}_t, u_t)) + Var(E(Y_t|\mathbf{x}_t, u_t)) \\ &= exp\left(0.5e^{\tilde{\mathbf{x}}_t^t\boldsymbol{\alpha}} + \mathbf{x}_t^t\boldsymbol{\beta}\right) + Var(exp(u_t)) \cdot \left(exp(\mathbf{x}_t^t\boldsymbol{\beta})\right)^2 \\ &= exp\left(0.5e^{\tilde{\mathbf{x}}_t^t\boldsymbol{\alpha}} + \mathbf{x}_t^t\boldsymbol{\beta}\right) + \\ &\quad + exp\left(e^{\tilde{\mathbf{x}}_t^t\boldsymbol{\alpha}}\right) \cdot \left(exp\left(e^{\tilde{\mathbf{x}}_t^t\boldsymbol{\alpha}}\right) - 1\right) \cdot \left(exp(\mathbf{x}_t^t\boldsymbol{\beta})\right)^2 \end{aligned} \tag{6.8}$$

As regressors $\tilde{\mathbf{x}}_t$ we choose the same (transformed and standardized) explanatory variables that already make up the vector $\mathbf{x}_t$, but without taking into account any interaction terms. A first MCMC simulation yields that only the Intercept term and the Bid-Ask-spread are significant on a 5%-credible level, so that we reduce the vector $\tilde{\mathbf{x}}_t$ to $(1, \text{stand. } log(bas_t))^t$. So far, the model does not account for the time series structure of the data, so that we incorporate an autoregressive structure into the process $\{u_t\}$ in a further MCMC simulation. We change the model (6.6) to

$$\begin{aligned} Y_t &\sim Poi(\lambda_t) \\ \lambda_t &= exp(\mathbf{x}_t^t \boldsymbol{\beta} + u_t) \\ u_t &= \rho \cdot u_{t-1} + \varepsilon_t \\ \varepsilon_t &= exp\left(\frac{\tilde{\mathbf{x}}_t^t \boldsymbol{\alpha}}{2}\right) \cdot \eta_t \\ \eta_t &\overset{iid}{\sim} N(0,1). \end{aligned}$$

In the simulation the parameter ρ is not $\neq 0$ on a 5%-credible level and we therefore discard this model in favour of the less complex Model (6.6).
We finally run 3 independent chains in order to estimate the parameters of the model (6.6) and again, reduce autocorrelations within the chains by only storing every 100th value. The trajectories of the chains and the Gelman-Rubin statistic plots (see Figures 6.20 to 6.26) indicate that convergence is then reached quite quickly. We choose a burn-in of 100 iterations and record a further 400 iterations for each chain. Again, we use Normal priors for the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$:

$$\begin{aligned} \beta_i &\sim N(0,10) \quad \text{for } i = 1, ..., 12 \\ \alpha_i &\sim N(0,10) \quad \text{for } i = 1, 2 \end{aligned}$$

We had to choose rather informative priors for this simulation due to problems of exponential overflow in WinBUGS with the less informative priors used for the simulations of the models discussed in the previous sections. Table 6.5 shows the results of the final MCMC simulation.

We also draw for this model plots similar to those shown in Figures 6.3 and 6.19 in order to make interpretations of the estimation results possible. These plots, shown in Figure 6.27, do not lead to any different conclusions about the influence of the covariates on the absolute option price changes compared to those already given in the previous sections for the ordinary Poisson-GLM and the Poisson-GLM with an AR(1) latent process.

We finally illustrate comparatively the way the previously presented models estimate

the influence of each explanatory variable on the absolute option price changes in Figures 6.28 and 6.29. In each plot, one of the explanatory variables is again shown on a linear, untransformed scale between the 10%- and 90%-quantiles. With all other explanatory variables set to their median values, the expected absolute option price change according to each model is calculated at ten equidistant grid points. These plots once again confirm that there are no substantial differences in the modelling of the influence of the explanatory variables on our variable of interest. It can be observed, however, that in the Poisson-GLM with an AR(1) latent process high values of the explanatory variable IV lead to larger expected absolute option price changes than in the other models, whereas for high values of the explanatory variable *bas* the expected absolute option price changes in the Poisson-GLM with an AR(1) latent process are smaller than in the other models.

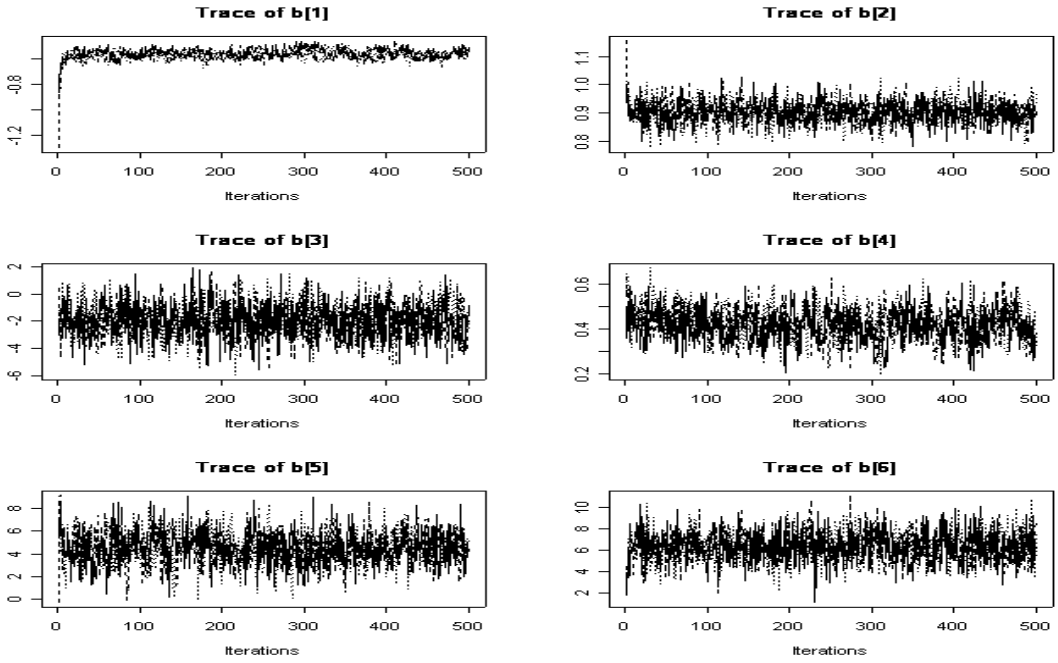

Figure 6.20: MCMC iterations for $\beta_1, ..., \beta_6$ (b[1],...,b[6]) for the model (6.6).

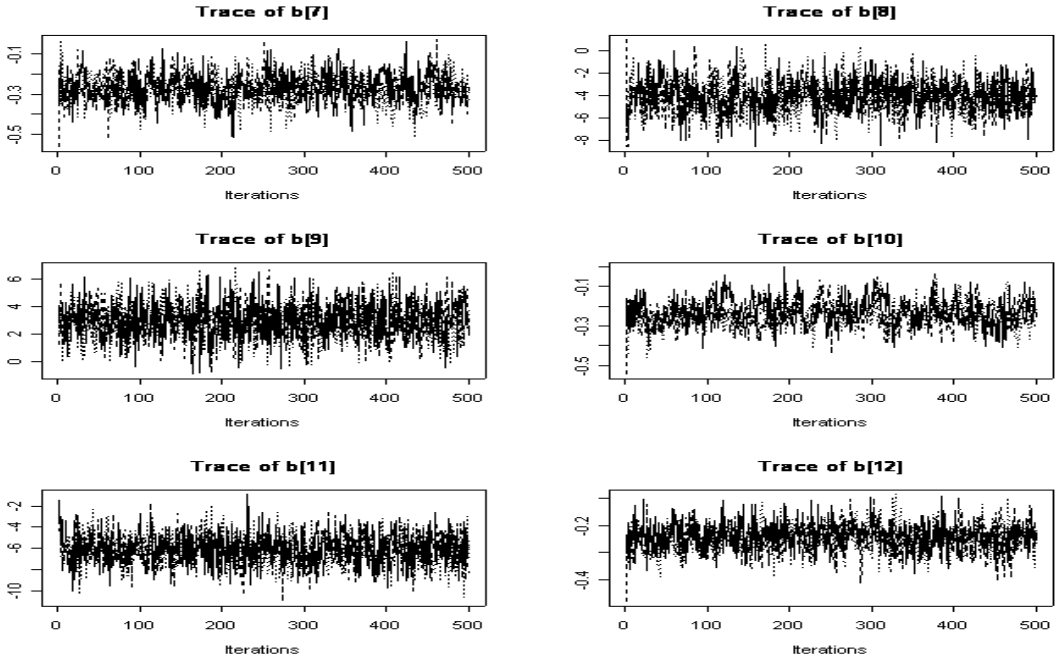

Figure 6.21: MCMC iterations for $\beta_7, ..., \beta_{12}$ (b[7],...,b[12]) for the model (6.6).

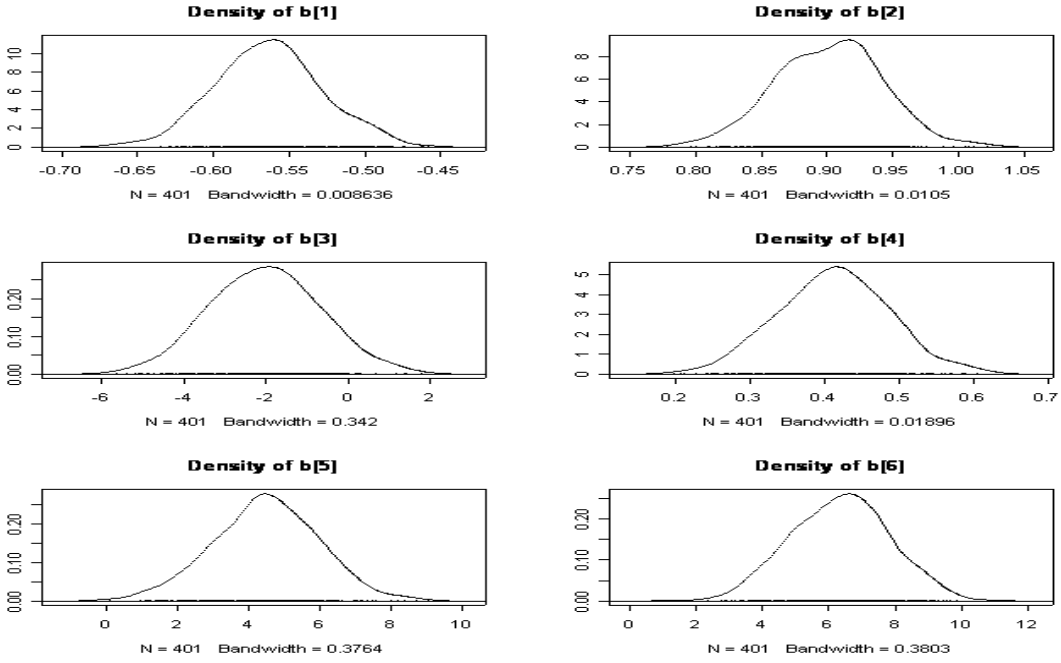

Figure 6.22: Posterior density estimates for $\beta_1, ..., \beta_6$ (b[1],...,b[6]) in the model (6.6) based on the last 400 recorded iterations in each chain .

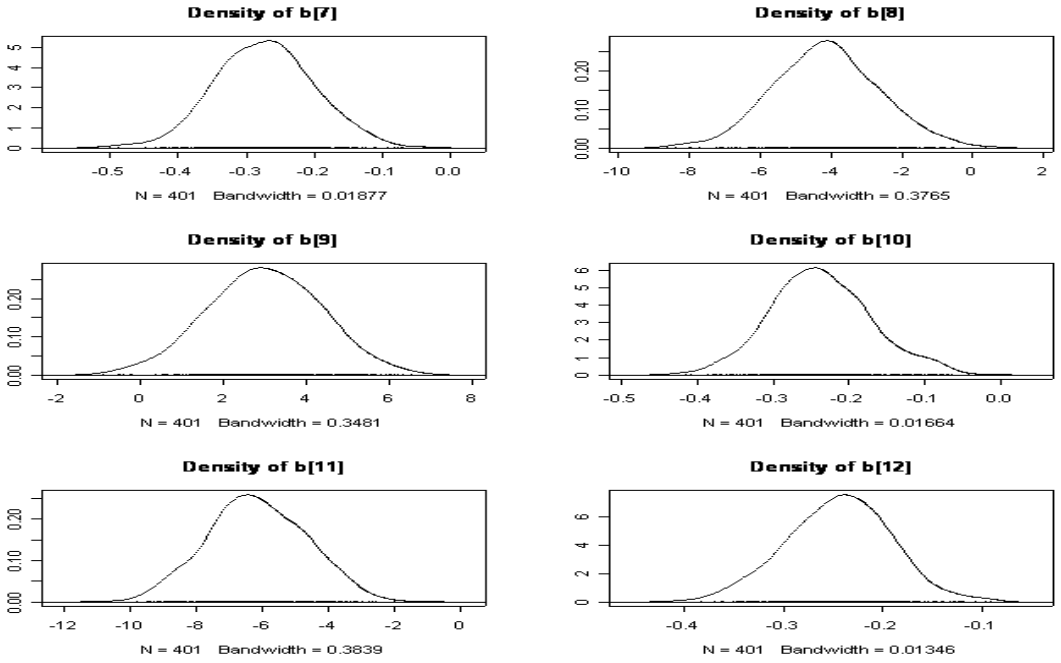

Figure 6.23: Posterior density estimates for $\beta_7, ..., \beta_{12}$ (b[7],...,b[12]) in the model (6.6) based on the last 400 recorded iterations in each chain .

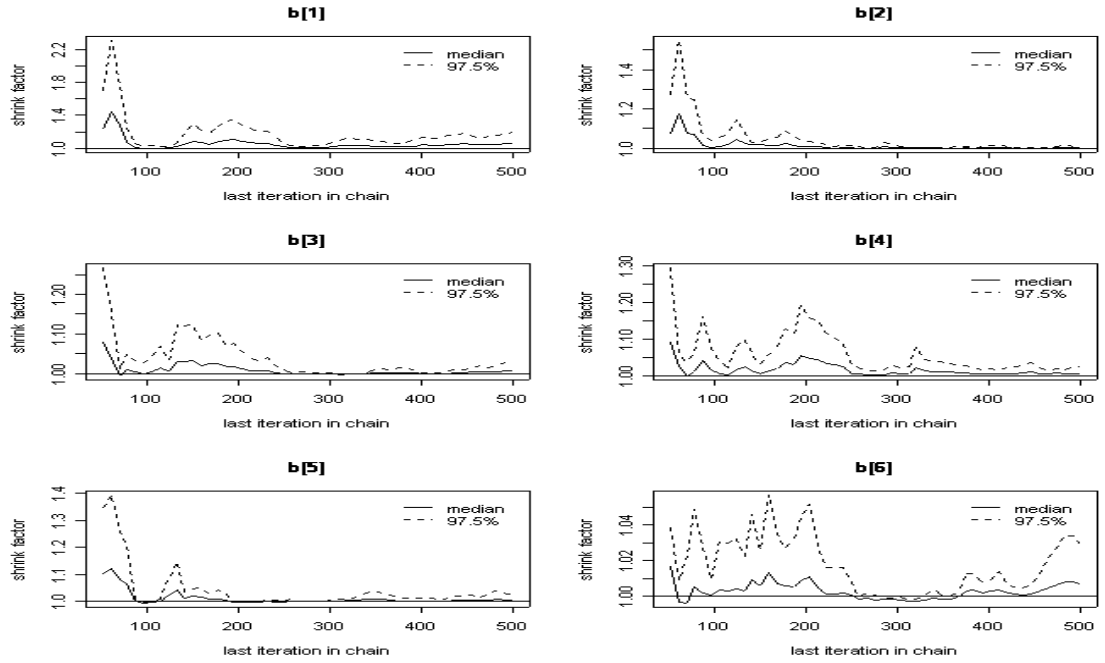

Figure 6.24: Gelman-Rubin convergence diagnostics for $\beta_1, ..., \beta_6$ (b[1],...,b[6]).

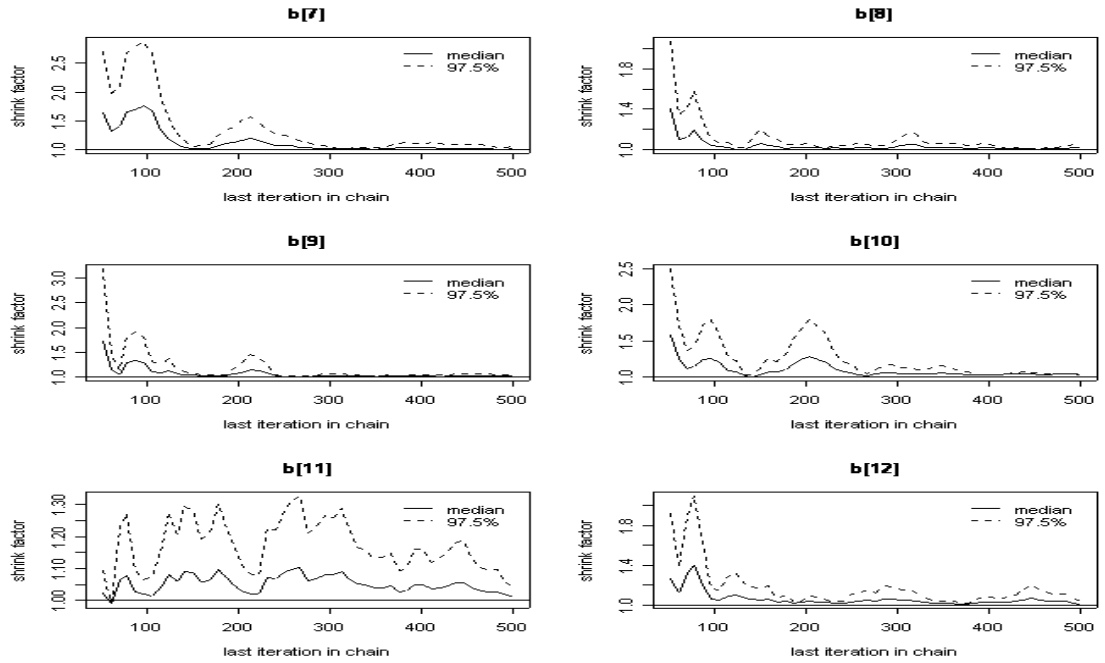

Figure 6.25: Gelman-Rubin convergence diagnostics for $\beta_7, ..., \beta_{12}$ (b[7],...,b[12]).

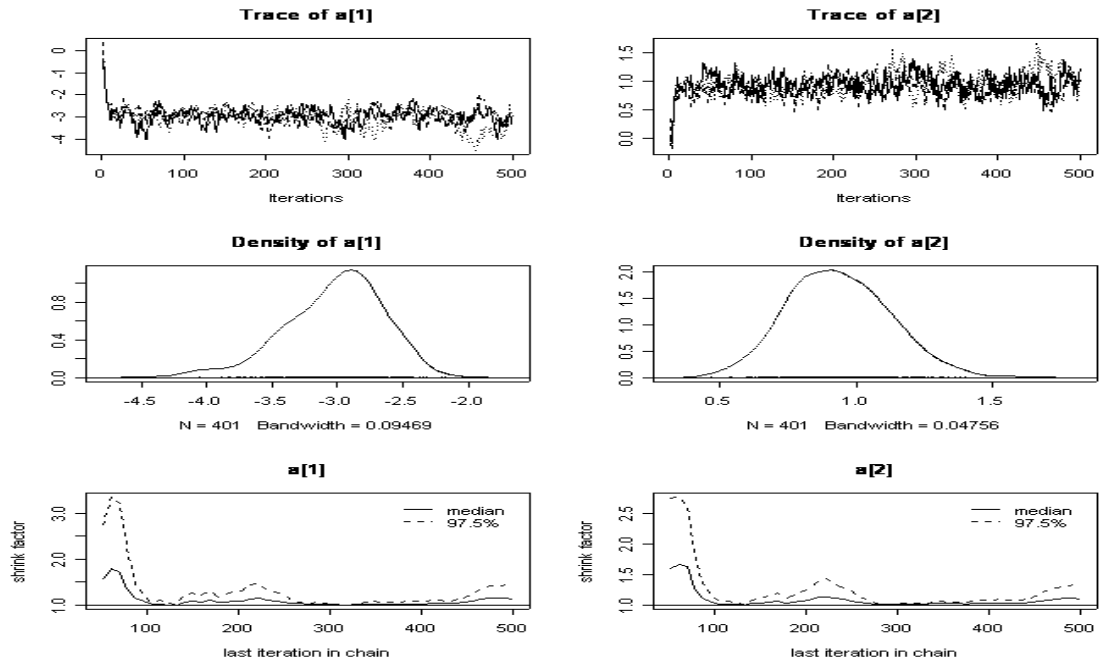

Figure 6.26: MCMC iterations, posterior density estimates and Gelman-Rubin convergence diagnostics for α_1 and α_2 in the model (6.6).

	Parameter	mean	Std. err.	2.5%	med.	97.5%
main effects	β_1 (Intercept)	-0.56	0.04	-0.63	-0.56	-0.49
	β_2 (stand. $log(UnCh + c_1)$)	0.90	0.04	0.82	0.90	0.98
	β_3 (stand. IV)	-2.01	1.34	-4.62	-2.01	0.69
	β_4 (stand. $log(NQ + c_2)$)	0.41	0.07	0.27	0.41	0.56
	β_5 (stand. $log(bas)$)	4.51	1.49	1.44	4.51	7.38
	β_6 (stand. $log(DeltaTm)$)	6.35	1.48	3.52	6.41	9.17
interaction effects	β_7 (stand. $IV : log(NQ + c_2)$)	-0.27	0.07	-0.41	-0.27	-0.13
	β_8 (stand. $log(bas) : log(UnCh + c_1)$)	-4.12	1.49	-7.01	-4.15	-1.07
	β_9 (stand. $IV : log(UnCh + c_1)$)	2.99	1.36	0.24	2.99	5.65
	β_{10} (stand. $log(bas) : log(NQ + c_2)$)	-0.24	0.07	-0.37	-0.24	-0.09
	β_{11} (stand. $log(DeltaTm) : log(UnCh + c_1)$)	-6.16	1.50	-9.04	-6.22	-3.32
	β_{12} (stand. $log(NQ + c_2) : log(DeltaTm)$)	-0.24	0.05	-0.35	-0.24	-0.14
	α_1 (Intercept)	-3.02	0.38	-3.96	-2.97	-2.41
	α_2 (stand. $log(bas)$)	0.94	0.19	0.60	0.93	1.32

Table 6.5: MCMC sampling results for the Poisson-GLM with heterogeneous error term variance (6.6) for the absolute price changes of the Call option on the XETRA DAX index with strike price 2600 and expiration month March 2003 ($c_1 = 80$, $c_2 = 1$) based on the last 400 recorded iterations in each chain .

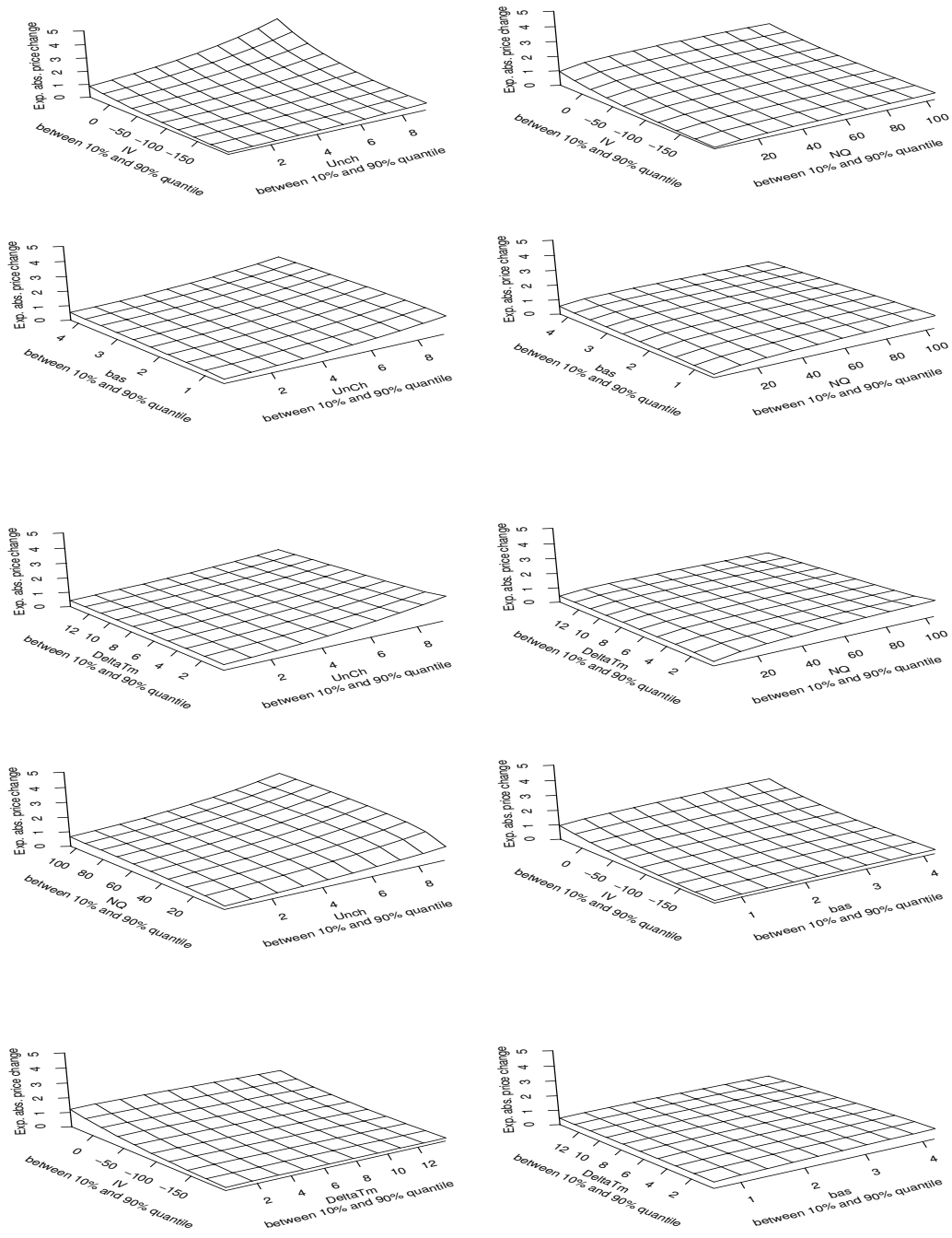

Figure 6.27: Fitted regression surfaces of the Poisson-GLM with heterogeneous variance structure for the absolute price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003 when two regressors vary and the remaining regressors are set to their median values.

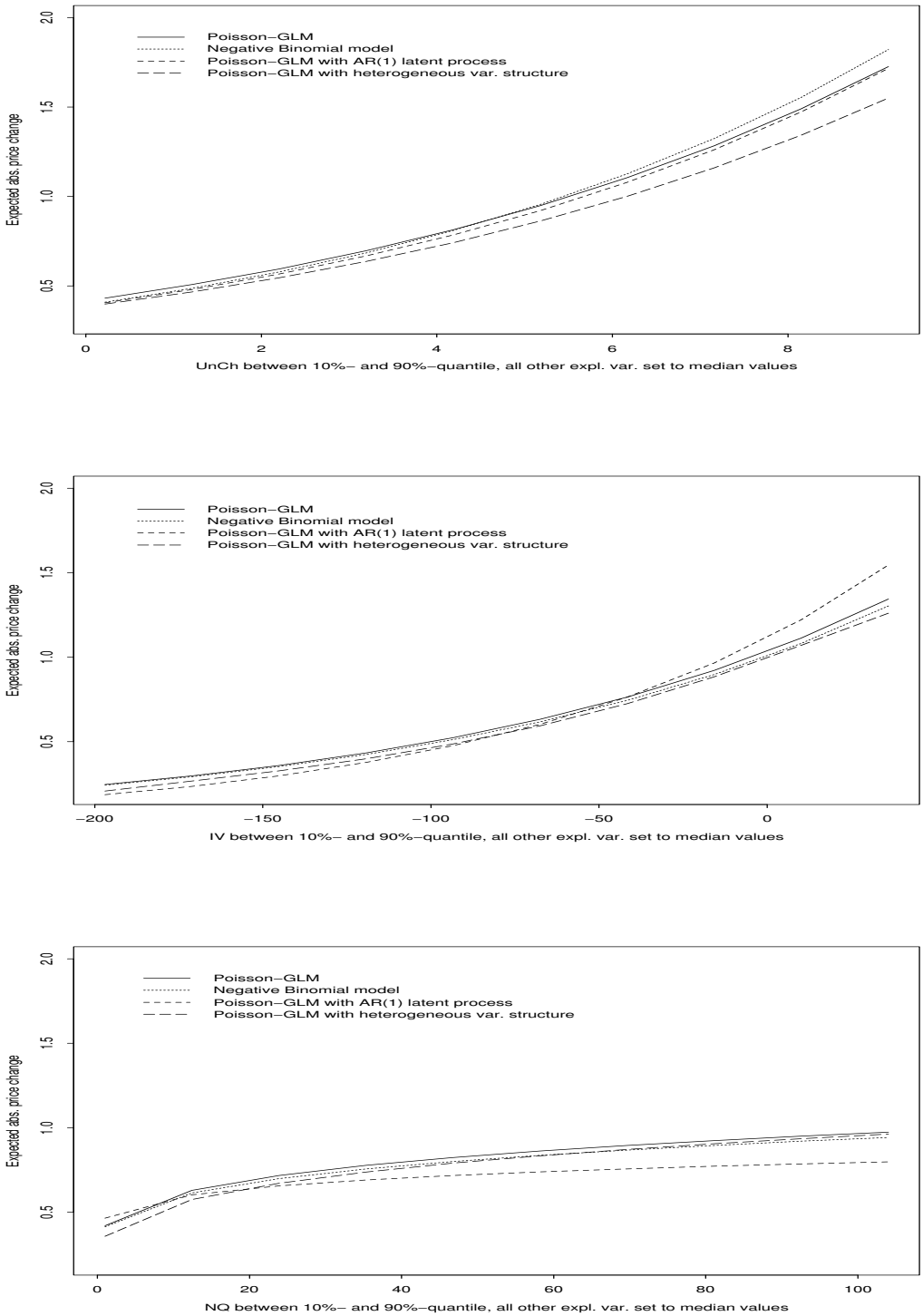

Figure 6.28: Comparative illustration of the estimated influence of the explanatory variables $UnCh$, IV and NQ on the absolute price changes of the Call option on the XETRA DAX with strike pr. 2600 and exp. month March 2003 in each of the discussed models.

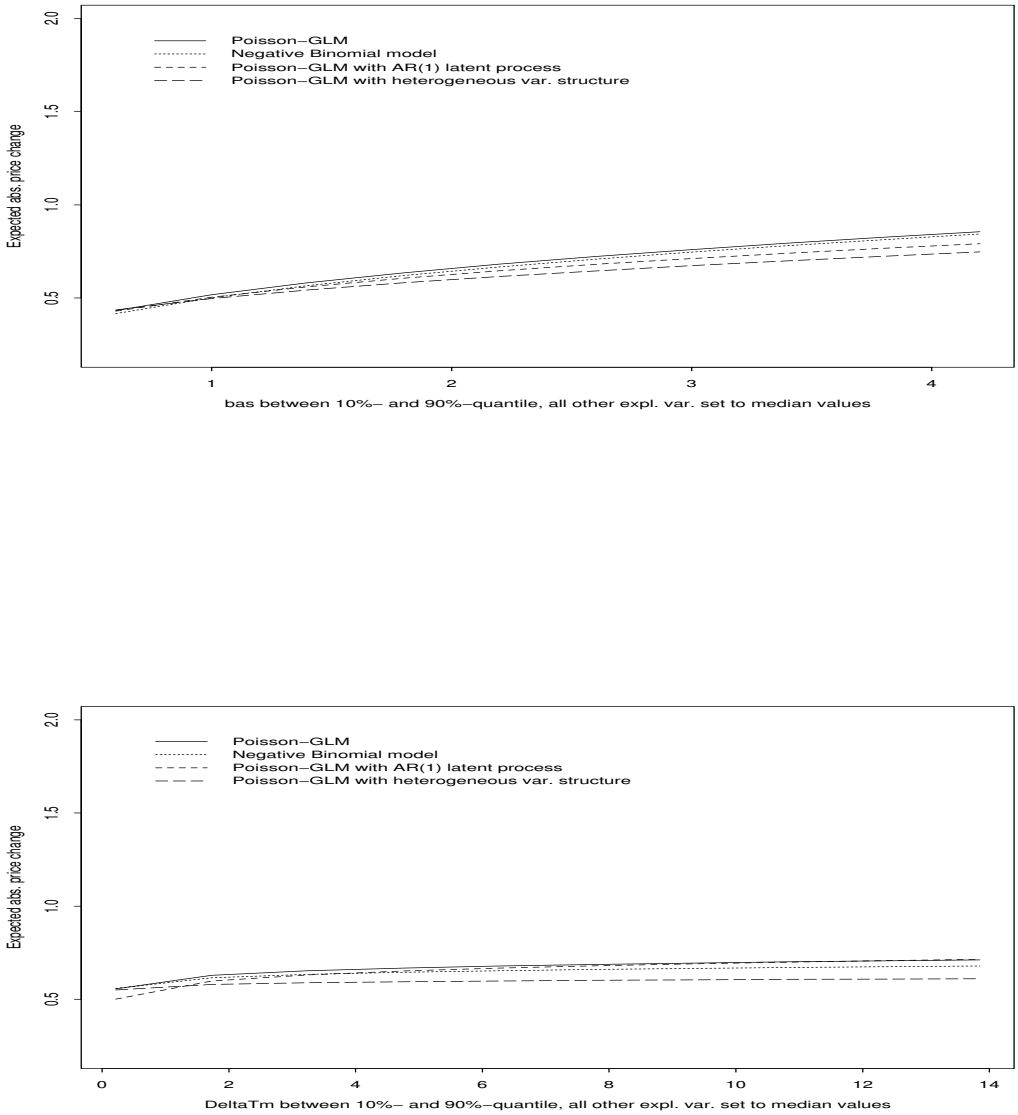

Figure 6.29: Comparative illustration of the estimated influence of the explanatory variables *bas* and $DeltaTm$ on the absolute price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003 in each of the discussed models.

Chapter 7

Assessment of model adequacy

In order to assess the adequacy of a statistical model, measures of fit, such as the residual deviance already introduced for GLMs in Chapter 3.1, have to be defined. In Section 7.1 we first discuss an informal approach and then extend the concept of residual deviance to Bayesian models.
Increasing model complexity (e.g. additional model parameters) is normally accompanied by better fit and comparison of models is usually carried out by trading off these quantities. For nested GLMs there exist well established criteria for model comparison such as the *Akaike Information Criterion* (AIC) which we have already used within the *step*-function in S-Plus in Chapter 6.1. The AIC is based on the idea of trading off the improvement of fit and the increasing complexity of the model when incorporating new parameters. For details on the AIC see for example McCullagh and Nelder (1989). For Bayesian models it is not straightforward to define a measure of complexity of the model. In Section 7.2 we introduce a measure proposed by Spiegelhalter et al. (2002) which can be used to define the *Deviance Information Criterion* for model comparison. We also calculate these quantities for the different models fit to our option data in Chapter 6 to finally assess and compare the adequacy of these models.

7.1 Measures of fit

7.1.1 An informal approach

Measures of fit are, in general, obtained by comparing the observations y_t of the variable of interest to the values $\hat{y}_t$ fitted by the model. For count data distributions it is, however, not perfectly clear how to determine $\hat{y}_t$. For all the count data models considered in this thesis we obtain an estimate $\hat{\lambda}_t$ not necessarily $\in \mathbb{N}_0$ for the mean of a Poisson- or Negative Binomial distribution. One possibility is to take the mode of this particular $Poi(\hat{\lambda}_t)$- or

$NegBin(\hat{\lambda}_t, \hat{\alpha})$-distribution as estimate $\hat{y}_t$.
Since for a $Poi(\lambda)$-distributed random variable Y and for all $n \in \mathbb{N}_0$

$$\begin{aligned} P(Y=n) &> P(Y=n+1) \\ \iff e^{-\lambda} \cdot \frac{\lambda^n}{n!} &> e^{-\lambda} \frac{\lambda^{n+1}}{(n+1)!} \\ \iff 1 &> \frac{\lambda}{n+1} \\ \iff n &> \lambda - 1 \end{aligned}$$

it follows that for $\lambda \notin \mathbb{N}$ the mode of the $Poi(\lambda)$ distribution is equal to $\lfloor \lambda \rfloor$. For the case that $\lambda \in \mathbb{N}$, both $\lambda - 1$ and λ have the same (maximum) probability.
Similarly, since for a $NegBin(\lambda, \alpha)$-distributed random variable Y

$$\begin{aligned} P(Y=n) &> P(Y=n+1) \\ \iff \frac{\Gamma(n+\alpha)}{\Gamma(\alpha) n!} \left(\frac{\lambda}{\alpha+\lambda}\right)^n \left(\frac{\alpha}{\alpha+\lambda}\right)^\alpha &> \frac{\Gamma(n+1+\alpha)}{\Gamma(\alpha)(n+1)!} \left(\frac{\lambda}{\alpha+\lambda}\right)^{n+1} \left(\frac{\alpha}{\alpha+\lambda}\right)^\alpha \\ \iff (n+1) \cdot \Gamma(n+\alpha) &> \Gamma(n+\alpha+1) \cdot \left(\frac{\lambda}{\alpha+\lambda}\right) \\ \overset{(*)}{\iff} (n+1) \cdot \Gamma(n+\alpha) &> (n+\alpha)\Gamma(n+\alpha) \cdot \left(\frac{\lambda}{\alpha+\lambda}\right) \\ \iff (n+1) \cdot (\alpha+\lambda) &> (n+\alpha) \cdot \lambda \\ \iff n &> \lambda - \frac{\lambda}{\alpha} - 1 \end{aligned}$$

it follows that the mode of a $NegBin(\lambda, \alpha)$-distribution is equal to $\lfloor \lambda - \frac{\lambda}{\alpha} \rfloor$. Here we used

$$(*) \quad \Gamma(z+1) = z \cdot \Gamma(z), \ \ \forall z \neq 0, -1, -2, ...$$

(see Rade et al. (1997), page 283).

The observations of the variable of interest can then be compared to the fitted values $\hat{y}_t$, equal to the mode of the corresponding distribution with posterior mean estimates for the parameters λ_t (and for α in the Negative Binomial case), by creating a contingency table with an equal number of rows and columns where the columns $0, 1, 2, ..., n$ correspond to the observations y_t and the rows $0, 1, 2, ..., n$ correspond to the fitted values $\hat{y}_t$. In order to keep the number of cells small we set $n = 10$ and aggregate all values ≥ 10 in the category n. As a rough measure of fit of the model the diagonal of the created contingency table can be considered. As a general rule, the better the fit the higher the values on the diagonal of the table. Of course, the table can also be considered as a matrix $\mathbf{M}$ and the trace of the matrix (or functions of it) as a rough measure of fit. In the following we will proceed

in the previously described way for all of the models fit to our data in Chapter 6 and calculate the ratio of the "correctly" fitted values to the total number of observations $T = 2419$. We denote this ratio by

$$RoF := \frac{tr(\mathbf{M})}{T}$$

and consider it as a first, informal measure of fit. The closer RoF to 1 the better the fit of the model. Of course, when considering a measure of fit of a statistical model one might also want to take into account not only the number of "correctly" fitted values, but also how much the "incorrectly" fitted values differ from the observed values. We therefore also calculate

$$SEM := \sum_{t=1}^{T} |\hat{y}_t - y_t|^2$$

as a second informal measure of fit where residuals are squared and added for $t = 1, ..., T$. Here, the better the fit, the smaller the value of SEM.
Table 7.1 shows the results for the ordinary Poisson-GLM (as of Table 6.1), Table 7.2 for the Negative Binomial model (as of Table 6.2), Table 7.3 for the Poisson-GLM with an AR(1) latent process in the mean (Table 6.4) and, finally, Table 7.4 shows the results for the Poisson-GLM with heterogeneous error term variance (Table 6.5).

It can be observed that, in general, the fitted values $\hat{y}_t$ seem to underestimate the real values of the absolute option price changes in all models. To some extent this could be expected since we consider the modes of the estimated response distributions as fitted values $\hat{y}_t$ and, as previously developed, we know that in the Poisson-models

$$\hat{y}_t = \lfloor \hat{\lambda} \rfloor \leq \hat{\lambda}_t = \widehat{E(Y_t)}$$

and that in the Negative Binomial model

$$\hat{y}_t = \lfloor \hat{\lambda} - \frac{\hat{\lambda}}{\hat{\alpha}} \rfloor \leq \hat{\lambda}_t = \widehat{E(Y_t)}.$$

The values of our ratios RoF lead to the assumption that there is not a big difference between the ordinary Poisson-GLM and the Negative Binomial model with respect to the goodness of fit. The incorporation of a latent AR(1)-process into the Poisson-GLM seems to lead to an improvement in the overall fit of the model. The best of all of the considered models, according to our RoF ratio, is the Poisson-GLM with the heterogeneous error term variance. The values of the SEM measure confirm these results with the exception of the Negative Binomial model which, according to this measure, leads to the worst fit of all of the considered models.
Once again, we point out that these results are rough and informal and should not be entirely relied on without further investigation.

observed values y_t / fitted values $\hat{y}_t$	0	1	2	3	4	5	6	7	8	9	≥10
0	1159	351	90	18	3	1	0	0	1	0	0
1	148	151	127	50	20	7	5	2	3	1	0
2	11	19	30	38	26	7	2	1	0	2	1
3	1	2	13	8	15	7	2	2	1	1	1
4	1	1	0	2	4	14	6	2	5	3	0
5	0	0	2	1	5	2	3	2	5	0	0
6	1	0	0	0	0	2	3	3	1	1	2
7	0	0	0	0	0	0	0	0	0	1	2
8	0	0	0	1	0	0	0	1	0	0	1
9	0	0	0	0	0	0	0	0	0	0	0
≥10	0	0	0	0	0	0	1	1	0	3	10

Table 7.1: Comparison of observed response values and fitted response values (mode of the estimated $Poi(\hat{\lambda}_t)$ distribution) for the ordinary Poisson-GLM. Cell entries are observed frequencies, $RoF = 56.5\%$, $SEM = 3392$.

observed values y_t / fitted values $\hat{y}_t$	0	1	2	3	4	5	6	7	8	9	≥10
0	1206	397	113	22	10	1	0	0	1	0	0
1	106	113	114	57	16	7	5	3	3	1	0
2	6	11	26	28	26	11	2	1	0	2	1
3	1	3	7	7	14	4	4	2	2	2	1
4	1	0	0	2	3	13	4	1	5	2	0
5	0	0	2	1	4	2	3	2	5	0	0
6	1	0	0	0	0	2	3	2	0	1	2
7	0	0	0	0	0	0	0	1	0	1	2
8	0	0	0	0	0	0	0	0	0	0	1
9	0	0	0	1	0	0	0	1	0	0	0
≥10	0	0	0	0	0	0	1	1	0	3	9

Table 7.2: Comparison of observed response values and fitted response values (mode of the estimated $NegBin(\hat{\lambda}_t, \hat{\alpha})$ distribution) for the Negative Binomial model. Cell entries are observed frequencies, $RoF = 56.6\%$, $SEM = 3669$.

observed values y_t / fitted values $\hat{y}_t$	0	1	2	3	4	5	6	7	8	9	≥10
0	1200	364	90	12	2	2	0	0	1	0	0
1	109	139	126	58	15	4	2	1	1	1	0
2	10	16	32	33	25	6	0	1	2	2	0
3	0	5	12	11	18	8	5	3	0	0	0
4	0	0	2	2	8	10	5	2	2	1	1
5	1	0	0	1	4	8	6	2	5	2	1
6	0	0	0	0	0	2	3	3	2	3	0
7	1	0	0	0	0	0	0	1	3	0	1
8	0	0	0	0	1	0	0	0	0	0	1
9	0	0	0	1	0	0	1	1	0	0	0
≥10	0	0	0	0	0	0	0	0	0	3	13

Table 7.3: Comparison of observed response values and fitted response values (mode of the estimated $Poi(\hat{\lambda}_t)$ distribution) for the Poisson-GLM with an AR(1) latent process. Cell entries are observed frequencies, $RoF = 58.5\%$, $SEM = 2929$.

observed values y_t / fitted values $\hat{y}_t$	0	1	2	3	4	5	6	7	8	9	≥10
0	1202	354	80	14	2	0	0	0	0	0	0
1	115	161	141	45	18	3	1	1	1	0	0
2	3	9	33	47	25	9	3	0	2	0	0
3	0	0	5	10	19	8	2	2	0	0	0
4	1	0	2	0	6	15	5	2	0	1	0
5	0	0	1	1	3	5	7	2	5	2	0
6	0	0	0	1	0	0	3	4	7	2	1
7	0	0	0	0	0	0	0	1	1	2	1
8	0	0	0	0	0	0	1	1	0	2	2
9	0	0	0	0	0	0	0	0	0	0	1
≥10	0	0	0	0	0	0	0	1	0	3	11

Table 7.4: Comparison of observed response values and fitted response values (mode of the estimated $Poi(\hat{\lambda}_t)$ distribution) for the Poisson-GLM with heterogeneous error term variance. Cell entries are observed frequencies, $RoF = 59.2\%$, $SEM = 2394$.

7.1.2 The Deviance

We have already discussed the residual deviance as a measure of fit for GLMs in Chapter 3.1. For Poisson-GLMs the deviance is given by:

$$D(\mathbf{y}, \hat{\boldsymbol{\lambda}}) := -2 \cdot [l(\mathbf{y}, \hat{\boldsymbol{\lambda}}) - l(\mathbf{y}, \mathbf{y})]$$

where $Y_t \sim Poi(\lambda_t)$ for $t = 1, 2, ..., T$ and l denotes the log-likelihood of the model. Basically, the deviance can be considered as a function of the normalized log-likelihood of the model with the log-likelihood of the saturated model as the normalizing constant. Of course, the question arises whether the concept of the deviance can be extended to Negative Binomial models. Note that in a Negative Binomial model it is not perfectly clear how the normalizing constant looks like since, due to the additional parameter α, the parameters in the saturated model are not identified. Thus, the choice of a normalizing constant becomes arbitrary and the possibility of comparing deviances of these different models gets questionable. Yet, since both the Poisson- and the Negative Binomial distribution are count data distributions one can simply add up the log-probabilities of the observed data after estimating the parameters in both models. This is, of course, equal to the calculation of the log-likelihood of the models. This approach is not arbitrary and thus, the results can be compared. The model with the higher value of the added log-probabilities can be considered as the "better" model with respect to the overall fit. For our Poisson-GLM of Chapter 6.1 the added log-probabilities of the observed data can be calculated by

$$\begin{aligned} l_{Poi}(\mathbf{y}; \hat{\boldsymbol{\lambda}}) &= \sum_{t=1}^{2419} \left(-\hat{\lambda}_t + y_t \cdot log(\hat{\lambda}_t) \right) - \sum_{t=1}^{2419} log(y_t!) \\ &= -827 - \sum_{t=1}^{2419} log(y_t!). \end{aligned}$$

The term $\sum_{t=1}^{2419} log(y_t!)$ does not need to be calculated since it appears in the same form in the Negative Binomial log-likelihood. The log-probability of the observed data in the Negative Binomial model can be calculated by:

$$\begin{aligned} l_{NegBin}(\mathbf{y}; \hat{\boldsymbol{\lambda}}, \hat{\alpha}) &= \sum_{t=1}^{2419} \left(log(\Gamma(y_t + \hat{\alpha})) + y_t \cdot log\left(\frac{\hat{\lambda}_t}{\hat{\alpha} + \hat{\lambda}_t} \right) + \hat{\alpha} \cdot log\left(\frac{\hat{\alpha}}{\hat{\alpha} + \hat{\lambda}_t} \right) \right) - \\ &\quad - 2419 \cdot log\left(\Gamma(\hat{\alpha})\right) - \sum_{t=1}^{2419} log(y_t!) \\ &= -811 - \sum_{t=1}^{2419} log(y_t!). \end{aligned}$$

Thus, the Negative Binomial model seems to have a slightly better overall fit than the ordinary Poisson-GLM.

Spiegelhalter et al. (2002) extend the concept of deviances to Bayesian models. For a Bayesian model, the probability function or density of the data $\mathbf{y}$ given the parameters $\boldsymbol{\theta}$ as well as a prior density for $\boldsymbol{\theta}$ have to be specified so that the marginal density of $\mathbf{y}$ is given by

$$p(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}) \quad d\boldsymbol{\theta}. \tag{7.1}$$

Particular choices for $p(\mathbf{y}|\boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$ define a model *focused* on $\boldsymbol{\theta}$. Following this denotation, the models fit to our option data in Chapters 6.3 - 6.4 were focused on the mean vector $\boldsymbol{\lambda}$. Of course, the priors can be further parameterized with unknown hyper-parameters which will lead to different likelihoods and to different numbers of model parameters. The marginal distribution (7.1) will, however, remain the same. In fact, by focusing models on a particular set of parameters, all models are reduced to non-hierarchical structures.
Using this denotation, Spiegelhalter et al. (2002) define a term

$$D(\boldsymbol{\theta}) := -2 \cdot log[p(\mathbf{y}|\boldsymbol{\theta})] + 2 \cdot log[f(\mathbf{y})] \tag{7.2}$$

as the *Bayesian deviance* of a Bayesian model, where $p(\mathbf{y}|\boldsymbol{\theta})$ is the probability function or density of the data given the parameter vector $\boldsymbol{\theta}$ that the model is focused on. $f(\mathbf{y})$ is again some fully specified standardizing term. If $p(\mathbf{y}|\boldsymbol{\theta})$ is a member of the exponential family (3.1) with $E(\mathbf{Y}) = \boldsymbol{\theta}$ they suggest using $f(\mathbf{y}) = p(\mathbf{y}|\mathbf{y})$ as normalizing constant.
Provided that $D(\boldsymbol{\theta})$ is available in closed form, it can easily be computed in an MCMC simulation and the posterior mean of the deviance $\overline{D(\boldsymbol{\theta})}$ can be considered as a Bayesian measure of fit. Of course, the quantity $D(\overline{\boldsymbol{\theta}})$, where $\overline{\boldsymbol{\theta}}$ denotes the posterior mean of the parameter vector $\boldsymbol{\theta}$ can also be calculated and considered as a classical plug-in measure of fit. We develop the relationship between these two quantities in the following section.

7.2 The Deviance Information Criterion

As previously mentioned, measures of fit have to be traded off with model complexity when comparing and selecting models.
Suppose that $p(\mathbf{Y}|\boldsymbol{\theta}^{true})$ is the "true" density or probability function of the observed data $\mathbf{y}$ with the "true" parameter value $\boldsymbol{\theta}^{true}$ and suppose that we have an estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}^{true}$. Spiegelhalter et al. (2002) then define

$$d(\mathbf{y}, \boldsymbol{\theta}^{true}, \hat{\boldsymbol{\theta}}) := -2 \cdot log[p(\mathbf{y}|\boldsymbol{\theta}^{true})] + 2 \cdot log[p(\mathbf{y}|\hat{\boldsymbol{\theta}})]$$

and interpret it as a measure of reduction in surprise or uncertainty due to estimation or, alternatively, as the degree of "over-fitting" due to adapting $\boldsymbol{\theta}$ to the data. The residual

information $-2log[p(\mathbf{y}|\boldsymbol{\theta}^{true})]$ is considered as a general measure of surprise or uncertainty. They suggest the posterior expectation of $d(\mathbf{y}, \boldsymbol{\theta}^{true}, \hat{\boldsymbol{\theta}})$ as the "effective number of parameters" in a Bayesian model focused on $\boldsymbol{\theta}$ and thus as a measure of complexity of the model.
If we take the posterior mean $\overline{\boldsymbol{\theta}}$ as estimator for $\boldsymbol{\theta}$, we get:

$$\begin{aligned} E_{\boldsymbol{\theta}|\mathbf{y}}\left[d(\mathbf{y}, \boldsymbol{\theta}^{true}, \hat{\boldsymbol{\theta}})\right] &= E_{\boldsymbol{\theta}|\mathbf{y}}\left[-2 \cdot log[p(\mathbf{y}|\boldsymbol{\theta})]\right] + 2 \cdot log[p(\mathbf{y}|\overline{\boldsymbol{\theta}})] \\ &= \overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}}) \quad := p_D \end{aligned} \tag{7.3}$$

with $D(\boldsymbol{\theta})$ defined as in (7.2). The complexity measure p_D can therefore be considered as a mean deviance minus the deviance of the means. Note that it depends on the observed data, on the choice of the focus $\boldsymbol{\theta}$, on the prior information and on the choice of the estimator $\hat{\boldsymbol{\theta}}$.

We have already discussed the mean deviance $\overline{D(\boldsymbol{\theta})}$ as a measure of fit in the previous section. Due to (7.3) we get:

$$\overline{D(\boldsymbol{\theta})} = D(\overline{\boldsymbol{\theta}}) + p_D.$$

I.e. the Bayesian measure of fit $\overline{D(\boldsymbol{\theta})}$ is equal to the classical plug-in measure of fit plus a measure of complexity. As a criteria for model comparison Spiegelhalter et al. (2002) finally suggest the *Deviance Information Criterion* defined by:

$$DIC := D(\overline{\boldsymbol{\theta}}) + 2p_D = \overline{D(\boldsymbol{\theta})} + p_D. \tag{7.4}$$

The DIC is therefore a Bayesian measure of fit $\overline{D(\boldsymbol{\theta})}$, penalized by an additional complexity term p_D. A certain model is, as a general rule, to be preferred to another model, if its DIC value is lower.
Of course, for non-Bayesian models there is no DIC value. In order to be able to compare the ordinary Poisson-GLM to the Bayesian models we reconsider it in a Bayesian setting, choose an $N(0, 1000)$ prior for each component of the parameter vector $\boldsymbol{\beta}$ and estimate the model by a MCMC simulation in WinBUGS. Similarly to the simulations carried out for the other models we run again 3 independent chains, reduce autocorrelation by only storing every 50th value and record 500 iterations after an initial burn-in of 500 iterations. This leads to estimation results for $\boldsymbol{\beta}$ which are very similar to those obtained in the classical approach with maximum-likelihood estimation. Thus, it is no surprise that the "plug-in" deviance $D(\overline{\boldsymbol{\theta}})$ of the Bayesian Poisson-GLM is equal to the deviance of the classical model setting already stated in the previous section (see Table 7.5).
In Table 7.5 we show the deviances (with the log-likelihood of the saturated model as normalizing constant) and the DIC values of the three Poisson-models considered in Chapter

6. The deviances in Column 2 and Column 3 of Table 7.5 confirm the assumptions obtained after the informal goodness of fit considerations in Chapter 7.1. The fit of the Poisson-GLM can indeed be improved by incorporating a latent AR(1)-process into the mean of the Poisson-GLM and an even better fit can be obtained when the error term variance of the ordinary Poisson-GLM is modelled as heterogeneous. However, since the "effective number of parameters" p_D, as defined by Spiegelhalter et al. (2002) as a measure of complexity, is significantly larger in the model with heterogeneous variance structure than in the model with the latent AR(1)-process, the DIC value of the latter is lower and thus, the Poisson-GLM with the latent AR(1)-process should be considered as "best" model within the group of Poisson-models.

Model	$D(\overline{\boldsymbol{\theta}})$	$\overline{D(\boldsymbol{\theta})}$	p_D	DIC
Ordinary Poisson-GLM (Bayesian setting)	2416	2428	12	2440
Poisson-GLM with latent AR(1)-process	2107	2217	110	2327
Poisson-GLM with heterogeneous var. structure	1974	2161	187	2348

Table 7.5: Deviances and DIC values for the Poisson option data models discussed in Chapter 6.

We have also considered a Negative Binomial model in a Bayesian setting in Chapter 6.2 and we have carried out a MCMC simulation in WinBUGS with the Ones-and-Bernoulli-trick. With regard to a possible comparison of this model with the (Bayesian) Poisson-models the question again arises how the standardizing term $f(\mathbf{y})$ for the calculation of the Bayesian deviance in (7.2) looks like, since, as already elaborated in Chapter 3.1, the Negative Binomial distribution is, if the parameter α is unknown, not a member of the exponential family. Thus, the choice of the standardizing term is again arbitrary and a calculation of a DIC value which can be compared to those of the Poisson-models in Table 7.5 is not possible. One can, however, define $f(\mathbf{y})$ by

$$f(\mathbf{y}) := \left(\prod_{t=1}^{T} y_t! \right)^{-1}, \tag{7.5}$$

so that

$$2 \cdot log[f(\mathbf{y})] = -2 \cdot \sum_{t=1}^{T} log(y_t!),$$

and use it for the calculation of the Bayesian deviances according to (7.2) for all models (Negative Binomial- and Poisson-models). This approach is similar to comparing log-

likelihoods and the Bayesian deviances as well as the DIC values of the different models can thus be compared. The results are shown in Table 7.6.

Model	$D(\overline{\boldsymbol{\theta}})$	$\overline{D(\boldsymbol{\theta})}$	p_D	DIC
Ordinary Poisson-GLM (Bayesian setting)	1656	1668	12	1680
Negative Binomial model (Bayesian setting)	1622	1640	18	1658
Poisson-GLM with latent AR(1)-process	1348	1458	110	1568
Poisson-GLM with heterogeneous var. structure	1214	1402	188	1590

Table 7.6: Deviances and DIC values for the option data models discussed in Chapter 6 with a standardizing term $f(\mathbf{y})$ as defined in (7.5).

The DIC values in Column 5 of Table 7.6 lead to the conclusion that the Negative Binomial model is "better" than the ordinary Poisson-GLM. The Poisson-GLM with the AR(1) latent process, however, can still be considered as the "best" of all of the discussed models.

Chapter 8

Discussion of results in their economic context

We finally summarize the results obtained in this thesis and relate our empirical results to some aspects of market microstructure theory.

In this thesis we have statistically analyzed the process of the absolute non-zero transaction price changes of an exchange-traded option on the XETRA DAX index and we have investigated the influence of a series of explanatory variables on this price process. For this purpose we have fitted regression models of different complexity to the data which account for the discreteness of the price changes and, depending on the particular model, allow for overdispersion, an autoregressive and heterogeneous variance structure of the data. Using a deviance information criterion we have selected a Poisson-GLM with an AR(1) latent process as the "best" model for our data (at least within the class of the Poisson-models).
We have seen that all of the discussed models lead to the same conclusions about the relationship between our selected explanatory variables and the absolute price changes of the option. In general the absolute price changes of the underlying could be determined as the variable with the strongest impact on the price changes of the option, which could be expected considering the construction of a Call option as a right to buy the underlying at a pre-specified price. If the price of the underlying changes significantly it is obvious that the price of the option must reflect this price movement of the underlying and it is likely that the option price also changes significantly. However, we have observed that this can only be expected for options that are in-the-money. In general, we could observe that the more the option is in-the-money, the larger the absolute price changes of the option. Particularly high absolute option price changes can, according to our modelling results, be expected when there is a large absolute price change of the underlying since

the previous price change of the option and, at the same time, the option is clearly in the money. Moreover, we have seen that, as a general rule, the higher the Bid-Ask spread at the time of the trade, the larger the absolute option price changes. This relationship is also quite obvious since the fair price of the option can be expected to lie somewhere between the Ask- and the Bid-price and the difference between these prices and the fair price can be considered as some kind of additional premium payment to the market maker. Higher premiums will then, evidently, lead to higher absolute transaction price changes.
We have furthermore observed that the number of new quotations between two price changes of the option also has quite a strong impact on the absolute option price changes. According to our modelling results, the more new quotations between the price changes the higher the absolute value of these price changes. A possible explanation for this may be the fact that the price process is not mainly driven by some exogenous forces of supply and demand, but strongly associated with some other publicly available information process. In this case, the market makers have to constantly adjust their quotes to the new information even if there is no current trading interest. When finally a trader wants to trade, the price must reflect all the accumulated information and a high absolute price change is likely. For our option data example the information process that strongly influences the price process of the option, is, of course, the price process of the underlying. We have observed that many new quotations between two price changes of the option associated with a high absolute price change of the underlying lead to particularly high absolute option price changes which gives empirical evidence for this explanation.
We have finally observed that longer times between consecutive price changes also lead to higher expected absolute option price changes. It is not perfectly clear if there is an economic explanation for this relationship and we will come back to this topic in the following paragraph where we relate our empirical results to some implications of market microstructure theory.

Market dynamics and especially the adjustment process of security prices to new information are, as already indicated in Chapter 1, a topic of major interest in market microstructure theory. In this context, traders are normally divided into two categories: Traders that have private, new information about the true value of the security, commonly denoted by *informed traders* and traders that do not have these information and simply trade for liquidity reasons which are not further specified. This second class of traders is usually called *liquidity traders* or *noise traders*. The trading behaviour of informed and liquidity traders then determines the price process of the security. For an extensive discussion of information-based trading models of this kind see O'Hara (1995). The question we finally want to investigate is whether the empirical results obtained in the previous

chapters for our option data example sustain some of the implications of the theoretical information-based trading models.

A first question is how security prices adjust to new private information. This question is closely connected to whether uninformed traders, and especially market makers, can learn about the existence of such private information of informed traders by observing the price process. In general, informed traders buy when the Ask-price set by the market maker is too low and they sell when the Bid-price is too high compared to the "true" value of the security according to the private information. Assume that a trader wants to buy the security at the currently quoted Ask-price. Of course, this trader may simply be a liquidity trader without any private information. But he may also have private, superior information about the "true" value of the security and may want to explore his informational advantage. Considering this possibility the market maker updates his beliefs about the "true" value of the security after the trade and he is likely to increase his Ask-price. If this new Ask-price is still considered as too low by the informed traders they continue to buy until the price has finally converged to a full-information level. These models are called *Bayesian learning models* (see O'Hara (1995), Chapter 3), since after each observed trade, the market maker and other uninformed market participants can update their knowledge about the "true" value of the security. This learning process can be considered as an application of the theory of Bayesian Inference as discussed in Chapter 3.3. Under such a scenario asymmetric private information in the market must be associated with a sequence of upward price movements if the private information is good news and with a sequence of downward price movements if the private information is bad news. The adjustment process continues until prices have reached full-information efficiency.

If the Bayesian learning model does indeed have practical relevance in real markets, observed price changes of a particular security can not assumed to be completely uncorrelated. For the process of the absolute price changes, considered in this thesis, the implications of the this model are less clear. It is however reasonable to expect that, given that there is autocorrelation in the process of the price changes, there is also some autocorrelation in the process of the absolute price changes.

For the absolute price changes of the Call option considered in this thesis, we have incorporated a time series structure in the modelling in Chapter 6.3 and we have come to the conclusion that there is indeed statistical evidence for autocorrelation in this process. We have also come to the conclusion that the model with the autoregressive structure is the "best" model, which can be interpreted as an indication of empirical evidence for the Bayesian learning trade model on the basis of our data.

Another question addressed in the market microstructure literature is the relationship

between the speed of the market, i.e. the time between trades and/or price changes, and the existence of new information. Provided that short-selling is not feasible, Diamond and Verecchia (1987) consider longer times between transactions as a possible signal for private bad news in the market. Since informed traders who do not already own the security are unable to trade and to explore their informational advantage in this case by selling the security, the probability that there are no trades at all for a certain time increases, given that the probability of a trade by a liquidity trader remains constant. When finally the bad news have been spread in the market the price of the security must reflect the bad news and a rather large downward price move (and thus a large absolute price move) is likely. For our option data we have observed that longer times between price changes are indeed associated with larger absolute price changes of the option. So, our results may indicate empirical evidence for this theory. Yet, it has to be pointed out that in order to provide better evidence, it is crucial to consider the direction of the price changes which we have not done in this thesis. Note also that in option markets the purchase of the corresponding Put option can, to some extent, replace the constraints on short-selling of a Call option, so that consideration of a bivariate price process for the Call and the corresponding Put option could certainly provide better evidence for the verification of this theory.

In the previous discussion of our modelling results in the context of market microstructure theory we have seen that the modelling of absolute price changes of a particular security may indicate empirical evidence for some theoretical hypotheses in this field of economic research. Yet, in order to provide better evidence, the incorporation of the direction of the price changes and maybe also the observation of price changes on a real trade-by-trade level (including trades that do not change the price of the security) seem indispensable. Obviously, this makes the statistical modelling much more complicated and will be the focus of further research.

Appendix A

WinBUGS codes

The following abbreviations for the observed variables are used throughout all programme codes:

Observed variable	Abbreviation in progr. code
absolute option price changes (response variable, y_t)	resp
stand. $log(UnCh + 80)$	Un
stand. IV	Iv
stand. $log(NQ + 1)$	Nq
stand. $log(bas)$	Ba
stand. $log(DeltaTm)$	De
stand. $(IV \cdot log(NQ + 1))$	IA1
stand. $(log(UnCh + 80) \cdot log(bas))$	IA2
stand. $(log(UnCh + 80) \cdot IV \longrightarrow)$	IA3
stand. $(log(bas) \cdot log(NQ + 1))$	IA4
stand. $(log(DeltaTm) \cdot log(UnCh + 80))$	IA5
stand. $(log(NQ + 1) \cdot log(DeltaTm))$	IA6

Ordinary Poisson-GLM

```
model {
for (t in 1:2419){

## Calculation of x_t^t β

mu[t]<- b[1]+b[2]*Un[t]+b[3]*Iv[t]+b[4]*Nq[t]+b[5]*Ba[t]+b[6]*De[t]
+b[7]*IA1[t]+b[8]*IA2[t]+b[9]*IA3[t]+b[10]*IA4[t]+b[11]*IA5[t]+b[12]*IA6[t];

lam[t]<-exp(mu[t]);
resp[t] ~ dpois(lam[t]);

## Calculation of components of the log-likelihood of the model

m[t]<- -lam[t];
n[t]<- resp[t]*log(lam[t]);
}

## Calculation of log-likelihood, omitting the additive constant − Σ_{t=1}^{2419} log(y_t!)

li <- sum(m[])+sum(n[])

## Vague normal priors for components of β

b[1] ~ dnorm(0, 0.001);
b[2] ~ dnorm(0, 0.001);
b[3] ~ dnorm(0,0.001);
b[4] ~ dnorm(0, 0.001);
b[5] ~ dnorm(0, 0.001);
b[6] ~ dnorm(0, 0.001);
b[7] ~ dnorm(0, 0.001);
b[8] ~ dnorm(0, 0.001);
b[9] ~ dnorm(0, 0.001);
b[10] ~ dnorm(0, 0.001);
b[11] ~ dnorm(0, 0.001);
b[12] ~ dnorm(0, 0.001);
}
```

Negative Binomial model

```
model {
magnitude.control<-100
for (t in 1:2419){

## Calculation of x_t^t β

mu[t]<- b[1]+b[2]*Un[t]+b[3]*Iv[t]+b[4]*Nq[t]+b[5]*Ba[t]+b[6]*De[t]
+b[7]*IA1[t]+b[8]*IA2[t]+b[9]*IA3[t]+b[10]*IA4[t]+b[11]*IA5[t]+b[12]*IA6[t];

lam[t]<-exp(mu[t]);
y[t] <-resp[t];

## Ones-and-Bernoulli-trick:

ones[t] <- 1
ones[t] ~ dbern(p[t])
a1[t] <- exp(loggam(alpha+y[t]))
a2[t] <- exp(loggam(alpha))* exp(loggam(y[t]+1))
t1[t] <- a1[t]/a2[t]
t3[t] <- pow(lam[t]/(lam[t]+alpha),y[t])
t2[t] <- pow(alpha/(lam[t]+alpha),alpha)
full[t] <- t1[t]* t2[t]* t3[t]
p[t] <- full[t]/magnitude.control

## Calculation of components of the log-likelihood of the model

n[t]<- loggam(resp[t]+alpha);
m[t]<-resp[t]* log(lam[t]/(alpha+lam[t]));
o[t]<-alpha* log(alpha/(alpha+lam[t]))-loggam(alpha);
}

## Calculation of log-likelihood, omitting the additive constant - sum_{t=1}^{2419} log(y_t!)

li<- sum(n[])+sum(m[])+sum(o[]);

## Vague normal priors for components of β, Gamma prior for α

b[1] ~ dnorm(0, 0.001);
```

```
b[2] ∼ dnorm(0, 0.001);
b[3] ∼ dnorm(0,0.001);
b[4] ∼ dnorm(0, 0.001);
b[5] ∼ dnorm(0, 0.001);
b[6] ∼ dnorm(0, 0.001);
b[7] ∼ dnorm(0, 0.001);
b[8] ∼ dnorm(0, 0.001);
b[9] ∼ dnorm(0, 0.001);
b[10] ∼ dnorm(0, 0.001);
b[11] ∼ dnorm(0, 0.001);
b[12] ∼ dnorm(0, 0.001);
alpha∼ dgamma(.1,.1)
}
```

Poisson-GLM with a latent AR(1)-process in the mean

```
model {

## Calculation of x_1^t β

mu[1] <- b[1]+b[2]* Un[1]+b[3]* Iv[1]+b[4]* Nq[1]+b[5]* Ba[1]+b[6]* De[1]
+b[7]* IA1[1]+b[8]* IA2[1]+b[9]* IA3[1] +b[10]* IA4[1]+b[11]* IA5[1]+b[12]* IA6[1];

z[1]∼ dnorm(mu[1],tau.u);
lam[1]<-exp(z[1]);
resp[1] ∼ dpois(lam[1]);

## Record of the quantities x_1^t β, u_1

eta[1]<-mu[1];
u[1]<-z[1]-eta[1];

## Components of the log-likelihood of the model for t = 1

m[1]<- -lam[1];
n[1]<- resp[1]* log(lam[1]);

for (t in 2:2419){
```

Calculation of $(\mathbf{x}_t^t - \rho\mathbf{x}_{t-1}^t)\boldsymbol{\beta} + \rho Z_{t-1}$ for $t = 2, ..., 2419$

```
temp[t]<-b[1]* (1-rho)+b[2]* (Un[t]-(rho* Un[t-1]))+b[3]* (Iv[t]-(rho* Iv[t-1]))
temp2[t]<-temp[t]+b[4]* (Nq[t]-(rho* Nq[t-1]))+b[5]* (Ba[t]-(rho* Ba[t-1]))+b[6]*
(De[t]-(rho* De[t-1]))
mu[t]<-temp2[t]+b[7]* (IA1[t]-(rho* IA1[t]))+b[8]* (IA2[t]-(rho* IA2[t-1]))
+b[9]* (IA3[t]-(rho* IA3[t-1]))+b[10]* (IA4[t]-(rho* IA4[t-1]))
+b[11]* (IA5[t]-(rho* IA5[t-1]))+b[12]* (IA6[t]-(rho* IA6[t-1]))+rho* z[t-1];

z[t]~ dnorm(mu[t],tau.e);
lam[t]<-exp(z[t]);
resp[t]~ dpois(lam[t]);
```

Record of the quantities $\mathbf{x}_t\boldsymbol{\beta}$ and of the values of the latent process u_t for $t = 2, ..., 2419$

```
eta[t]<- b[1]+b[2]* Un[t]+b[3]* Iv[t]+b[4]* Nq[t]+b[5]* Ba[t]+b[6]* De[t]
+b[7]* IA1[t]+b[8]* IA2[t]+b[9]* IA3[t] +b[10]* IA4[t]+b[11]* IA5[t]+b[12]* IA6[t];
u[t]<-z[t]-eta[t];
```

Calculation of components of the log-likelihood of the model

```
m[t]<- -lam[t];
n[t]<- resp[t]* log(lam[t]);
}
```

Calculation of variance of latent process and conversion of precisions to variances

```
sigma.e <- 1/tau.e ;
sigma.u <- sigma.e/(1-pow(rho,2)) ;
tau.u <- 1/sigma.u;
```

Calculation of log-likelihood, omitting the additive constant $-\sum_{t=1}^{2419} log(y_t!)$

```
li <- sum(m[])+sum(n[]);
```

Vague normal priors for components of $\boldsymbol{\beta}$, uninformative Gamma prior for τ_ε and uniform prior for ρ

```
rho ~ dunif(-1,1);
```

```
b[1] ~ dnorm(0, 0.001);
b[2] ~ dnorm(0, 0.001);
b[3] ~ dnorm(0,0.001);
b[4] ~ dnorm(0, 0.001);
b[5] ~ dnorm(0, 0.001);
b[6] ~ dnorm(0, 0.001);
b[7] ~ dnorm(0, 0.001);
b[8] ~ dnorm(0, 0.001);
b[9] ~ dnorm(0, 0.001);
b[10] ~ dnorm(0, 0.001);
b[11] ~ dnorm(0, 0.001);
b[12] ~ dnorm(0, 0.001);
tau.e ~ dgamma(.05, .05);
}
```

Poisson-GLM with heterogeneous variance structure

```
model {
for (t in 1:2419){

## Calculation of x_t^t β

mu[t] <- b[1]+b[2]* Un[t]+b[3]* Iv[t]+b[4]* Nq[t]+b[5]* Ba[t]+b[6]* De[t]
+b[7]* IA1[t]+b[8]* IA2[t]+b[9]* IA3[t] +b[10]* IA4[t]+b[11]* IA5[t]+b[12]* IA6[t];

## Calculation of x̃_t^t α

sigma.e[t]<- exp(a[1]+a[2]* Ba[t]);

## Conversion of variance to precision

tau.e[t] <- 1/sigma.e[t] ;

z[t]~ dnorm(mu[t],tau.e[t]);
lam[t]<-exp(z[t]);
resp[t] ~ dpois(lam[t]);

## Calculation of components of the log-likelihood of the model
```

```
m[t]<- -lam[t];
n[t]<- resp[t]* log(lam[t]);
}

## Calculation of log-likelihood, omitting the additive constant -∑_{t=1}^{2419} log(y_t!)

li <- sum(m[])+sum(n[]);

## Normal priors for components of β and α


b[1] ~ dnorm(0, 0.1);
b[2] ~ dnorm(0, 0.1);
b[3] ~ dnorm(0,0.1);
b[4] ~ dnorm(0, 0.1);
b[5] ~ dnorm(0, 0.1);
b[6] ~ dnorm(0, 0.1);
b[7] ~ dnorm(0, 0.1);
b[8] ~ dnorm(0, 0.1);
b[9] ~ dnorm(0, 0.1);
b[10] ~ dnorm(0, 0.1);
b[11] ~ dnorm(0, 0.1);
b[12] ~ dnorm(0, 0.1);
a[1] ~ dnorm(0, 0.1);
a[2] ~ dnorm(0, 0.1);
}
```

Bibliography

Bartram, B. and F. Fehle (2003). Competition among Alternative Option Market Structures. *Working Paper, University of South Carolina, Moore School of Business*.

Brockwell, P. and R. Davis (1991). *Time Series: Theory and Methods*. New York: Springer-Verlag.

Brooks, S. and G. Roberts (1998). Assessing Convergence of Markov Chain Monte Carlo Algorithms. *Statistics and Computing 8*, 319–335.

Cameron, A. and P. Trivedi (1998). *Regression analysis of count data*. Cambridge: Cambridge University Press.

Casella, G. and R. Berger (1990). *Statistical Inference*. Brooks Cole.

Chan, K. and J. Ledolter (1995). Monte Carlo EM Estimation for Time Series Models Involving Counts. *Journal of the American Statistical Association 90*, 242–251.

Cowles, M. and B. Carlin (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association 91*, 883–904.

Cox, D. (1993). Statistical analysis of time series, some recent developments. *Scandinavian Journal of Statistics 8*, 93–115.

Davis, R., W. Dunsmuir, and S. Streett (2003). Observation Driven Models for Poisson Counts. *Biometrika 9*, 777–790.

Davis, R., W. Dunsmuir, and Y. Wang (1999). *"Modelling Time Series of Count Data"* Asymptotics, Nonparametrics and Time Series (Subir Gosh, editor)*, 63-114*. New York: Marcel Dekker.

Diamond, D. and R. Verecchia (1987). Contraints on Short-Selling and Asset Price Adjustment to Private Information. *Journal of Financial Economics 18*, 277–311.

Engle, R. (2000). The Econometrics of Ultra-High-Frequency Data. *Econometrica 68*, 1–22.

Fahrmeir, L., H. Kaufmann, and F. Ost (1981). *Stochastische Prozesse. Eine Einführung in Theorie und Anwendungen*. Muenchen: Hanser Fachbuch.

Gamerman, D. (1997). *Markov Chain Monte Carlo- Stochastic simulation for Bayesian inference*. London: Chapman & Hall.

Gelman, A. and D. Rubin (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science 7*, 457–472.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *Bayesian Statistics 4*, 169–193.

Gilks, W., S. Richardson, and D. Spiegelhalter (1996). *Markov Chain Monte Carlo in practice*. London: Chapman & Hall.

Harris, L. (1994). Minimum price variations, discrete bid-ask spreads, and quotation sizes. *Review of Financial Studies 7*, 149–178.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97–109.

Hull, J. C. (2003). *Options, Futures and other derivatives*. New Jersey: Prentice Hall.

Johnson, N., S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions, Vol. 1, 2nd edition*. New York: John Wiley & Sons.

Liesenfeld, R. and W. Pohlmeier (2003). A Dynamic Integer Count Data Model for Financial Transaction Prices. *Working Paper, University of Konstanz*.

McCullagh, P. (1983). Quasi-likelihood functions. *Ann. Statist. 11*, 59–67.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models - Second edition*. London: Chapman & Hall.

McCulloch, C. and S. Searle (2001). *Generalized, Linear and Mixed Models*. New York: John Wiley & Sons.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953). Equation of state calculations by fast computing machine. *Journal of Chemical Physics 21*, 1087–1091.

O'Hara, M. (1995). *Market Microstructure Theory*. Cambridge, MA: Blackwell Publishers.

Rade, L., B. Westergren, and P. Vachenauer (1997). *Springers Mathematische Formeln*. Berlin: Springer-Verlag.

Raftery, A. and S. Lewis (1992). How Many Iterations in the Gibbs Sampler? *Bayesian Statistics 4*, 763–773.

Resnick, S. (1992). *Adventures in Stochastic Processes*. Boston: Birkhaeuser.

Shephard, N. (1995). Generalized Linear Autoregressions. *Unpublished Paper, Oxford University*.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society 64*, 583–639.

Stoehr, J. (1999). *Numerische Mathematik 1*. Berlin: Springer-Verlag.

Tanner, M. (1996). *Tools for Statistical Inference*. New York: Springer-Verlag.

Venables, W. and B. Ripley (1999). *Modern Applied Statistics with S-PLUS*. Springer-Verlag.

Zeger, S. (1988). A Regression Model for Time Series of Counts. *Biometrika 75*, 621–629.