

Czado, Claudia; Kolbe, Andreas

**Working Paper**

## Empirical study of intraday option price changes using extended count regression models

Discussion paper // Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München, No. 403

**Provided in Cooperation with:**

Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

Suggested Citation: Czado, Claudia; Kolbe, Andreas (2004) : Empirical study of intraday option price changes using extended count regression models, Discussion paper // Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München, No. 403, Techn. Univ.; Sonderforschungsbereich 386, Statistische Analyse Diskreter Strukturen, München, <http://nbn-resolving.de/urn:nbn:de:bvb:19-epub-1773-7>

This Version is available at:  
<http://hdl.handle.net/10419/31108>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Empirical Study of Intraday Option Price Changes using extended Count Regression Models

Claudia Czado, Andreas Kolbe  
Center for Mathematical Sciences  
Munich University of Technology  
Boltzmannstr. 3  
D-85747 Garching b. München, Germany  
cczado@ma.tum.de, Andreas.Kolbe@extern.lrz-muenchen.de

December 1, 2004

## Abstract

In this paper we model absolute price changes of an option on the XETRA DAX index based on quote-by-quote data from the EUREX exchange. In contrast to other authors, we focus on a parameter-driven model for this purpose and use a Poisson Generalized Linear Model (GLM) with a latent AR(1) process in the mean, which accounts for autocorrelation and overdispersion in the data. Parameter estimation is carried out by Markov Chain Monte Carlo methods using the WinBUGS software. In a Bayesian context, we prove the superiority of this modelling approach compared to an ordinary Poisson-GLM and to a complex Poisson-GLM with heterogeneous variance structure (but without taking into account any autocorrelations) by using the deviance information criterion (DIC) as proposed by Spiegelhalter et al. (2002). We include a broad range of explanatory variables into our regression modelling for which we also consider interaction effects: While, according to our modelling results, the price development of the underlying, the intrinsic value of the option at the time of the trade, the number of new quotations between two price changes, the time between two price changes and the Bid-Ask spread have significant effects on the size of the price changes, this is not the case for the remaining time to maturity of the option. By giving possible interpretations of our modelling results we also provide an empirical contribution to the understanding of the microstructure of option markets.

*Keywords: index options, quotation data, price change process, poisson regression, latent process, autocorrelation, Markov Chain Monte Carlo, DIC*

*The first author's research was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 386: Statistische Analyse diskreter Strukturen.*

# 1 Introduction, motivation and related 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. The books by Bauwens and Giot (2001) and Dacorogna et al. (2001) are two examples of publications covering various aspects related to this field of research. An introduction to the econometric modelling of ultra-high frequency financial data as well as a survey of current and future topics of research in this field is given by Hautsch and Pohlmeier (2002). 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. Rydberg and Shephard (2003), 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 allow for the incorporation of other marks of the trading process as regressors are used. Rydberg and Shephard (2003) decompose the price process at transaction level into three components:

1. A binary process on  $\{0, 1\}$  modelling activity (meaning the price moves or not)
2. Another binary process on  $\{-1, 1\}$  modelling the direction of the price move
3. A process on strictly positive integers modelling the absolute value (=size) of the price move

Liesenfeld and Pohlmeier (2003) develop quite a similar model called integer count hurdle (ICH) model where the two binary processes of the Rydberg-Shephard model are incorporated into a trinomial autoregressive conditional multinomial (ACM) model (no price change or price movement upwards or price movement downwards). In both publications some explanatory variables (e.g. the time between consecutive transactions) are incorporated into the modelling and the models are applied to stocks traded at the NYSE and the Frankfurt Stock Exchange respectively. In this paper we concentrate on the last component of the previously described modelling approaches and consider only absolute non-zero transaction price changes. While both Rydberg and Shephard (2003) and Liesenfeld and Pohlmeier (2003) use Negative Binomial Generalized Linear ARMA (GLARMA)

models, which following the classification of Cox (1993) belong to the class of observation-driven models, a parameter-driven Poisson-GLM with an AR(1) latent process in the mean is used in this paper. An overview of parameter- and observation-driven regression models for time series of counts with econometric and biometric applications can be found for example in Davis, Dunsmuir, and Wang (1999), while Zeger (1988) and Chan and Ledolter (1995) explicitly discuss and develop estimation techniques for the Poisson-GLM with an autoregressive latent process in the mean.

While applications have had their focus on stock markets in other publications, we consider non-zero transaction price changes of an option on the German XETRA DAX index based on quote-by-quote data from the EUREX exchange in this paper. While price changes of stocks are often limited to a small number of discrete values (mainly  $\pm 3, \pm 2, \pm 1$  ticks), we have to deal with a wide range of values in our option data example. Thus, only a modelling approach with the Poisson distribution or the Negative Binomial distribution (our preference for the Poisson distribution will be elaborated later) can be considered as adequate. In our regression analysis we include a wide range of other marks of the trading process as explanatory variables: In addition to the time between consecutive price changes, the Bid-Ask spread at the time of the trade and the number of new quotations between two price changes we also consider option-specific quantities such as the price development of the underlying, the intrinsic value of the option at the time of the trade and the remaining time to maturity, which also provides an empirical contribution to the understanding of market microstructure in general and option markets in particular.

Beside the Poisson-GLM with an AR(1) latent process in the mean we also fit an ordinary Poisson-GLM and a Poisson-GLM with a heterogeneous variance structure to our data for comparison purposes. 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 for example summarized in Chib (2001) and discussed in detail in Gamerman (1997). Gilks, Richardson, and Spiegelhalter (1996) provide examples of applications. It is, of course, not straightforward to compare the adequacy of different modelling approaches, especially in a Bayesian context. We use an information criterion proposed by Spiegelhalter, Best, Carlin, and van der Linde (2002) in order to finally compare and assess the adequacy of the different models. For the computation the S-Plus and WinBUGS software packages are used.

The paper is organized as follows: The available data is presented in Chapter 2. In Chapter 3 we shortly introduce and discuss the regression models which we use for the modelling of the data. The modelling results are presented in Chapter 4 and the adequacy of the models is comparatively

discussed in Chapter 5. Finally, Chapter 6 concludes, gives some interpretations of the results and discusses weaknesses of our modelling approach.

## 2 Data presentation

The available data set for this paper consists of intraday quote-by-quote data for a European Call option on the XETRA DAX index traded at the Eurex exchange. The strike price of the option is 2600 points and its expiry month is March 2003. We consider in this paper the time period between 10-Feb-2003 and 21-March-2003, the expiration day of the option. Figure 1 shows the last prices of some selected Call options on the XETRA DAX index with different strike prices, including the previously mentioned security, on a quote-by-quote data base during this time period.

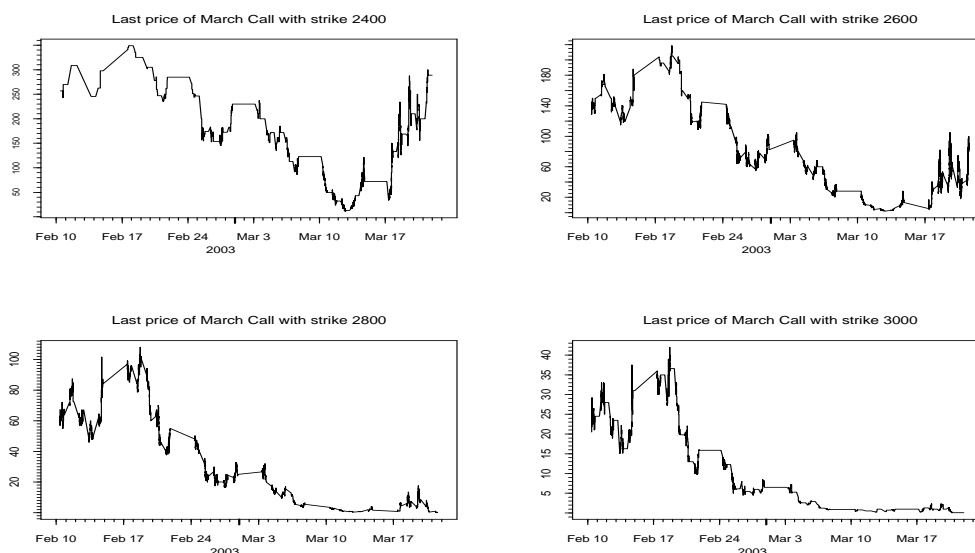


Figure 1: Last prices of some selected Call options on the XETRA DAX with different strike prices.

### 2.1 The variable of interest: Absolute option price changes

In this paper we analyze the changes of the quoted last prices 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. As already mentioned in Section 1, we do not model the direction of the price changes, but consider only their absolute values. 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.

In our data set it can be observed that multiples of 10 ticks (a tick is the smallest incremental change of the price of the security and equals in our data example 0.1 index points or EUR 0.50) 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 2 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  $t = 1, 2, \dots, T$  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.

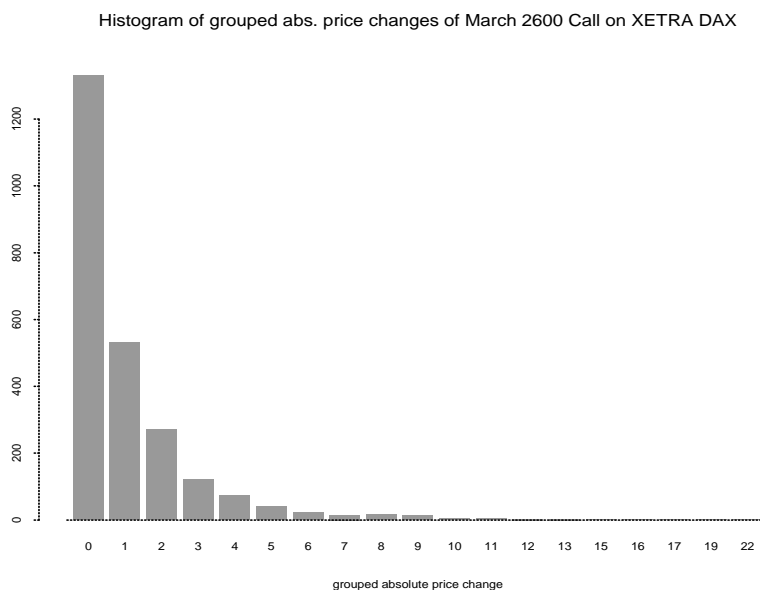


Figure 2: Histogram of grouped non-zero transaction price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003.

## 2.2 Explanatory variables

We now introduce a series of explanatory variables that may have influence on  $Y_t$  as previously defined. 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  $\tau_1$ ,  $\tau_2$  and  $\tau_3$ . The corresponding values of the variable of interest are  $Y_t = 7$  (since the price change at the point of time  $\tau_1$  is equal to 8 points or 80 ticks, which yields a value of 7 for our variable of interest grouped as described earlier),  $Y_{t+1} = 2$  and  $Y_{t+2} = 1$  where the  $t$ -th price change of the option in the time period we consider in this paper occurred at the point of time  $\tau_1$  (a convention we will maintain in the following illustration of the explanatory variables).

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	$\tau_1 = 2003-03-21 12:56:20$
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	$\tau_2 = 2003-03-21 12:57:04$
2003-03-21 12:57:06	90	83	90.8	$\tau_3 = 2003-03-21 12:57:06$
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 sections. 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:**

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

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

where  $\tau$  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  $\tau_1$ .

**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 3. 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  $\Delta Tm$  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, as expected.

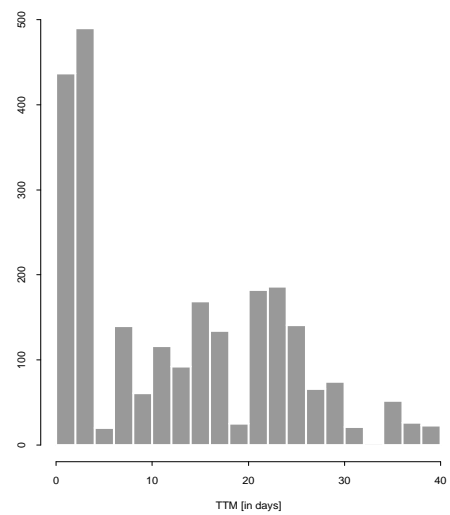
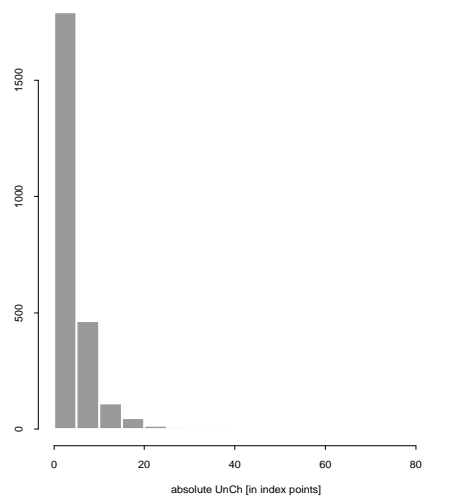
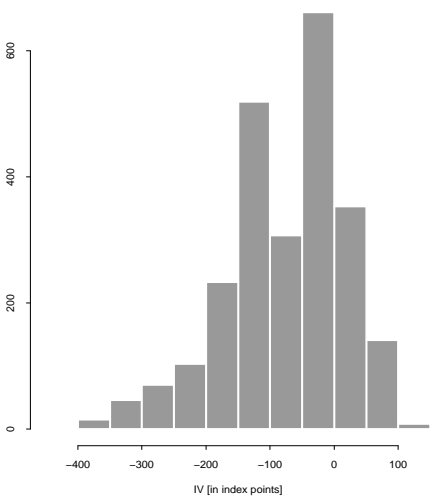
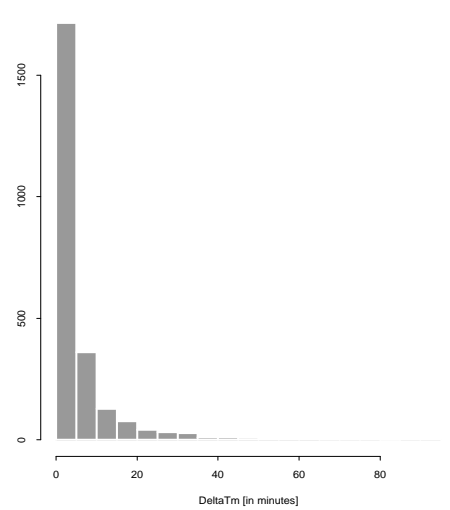
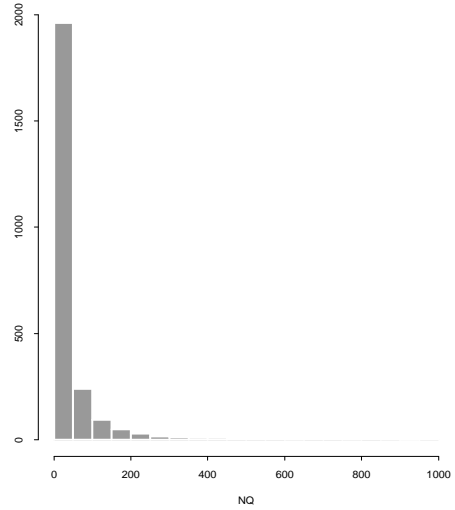
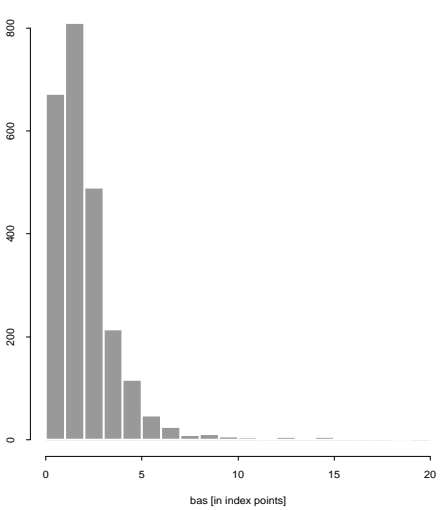


Figure 3: Histograms of explanatory variables.

### 3 Regression models for the statistical analysis

Before we take into account the time series structure of the price changes and fit a Poisson-GLM with an AR(1) latent process in the mean to our data, we first consider an ordinary Poisson regression model which seems to be a natural starting point for any count data regression. Since the Poisson regression model is a special case of a Generalized Linear Model, its statistical properties are well-known and parameter estimation can easily be handled (see for example McCullagh and Nelder (1989)). The incorporation of an unobservable latent process in the mean of the Poisson-GLM introduces autocorrelation into the model and therefore accounts for its time series structure. Parameter estimation in this class of models is, however, much more complicated than in the ordinary Poisson-GLM. In the literature several estimation techniques have been proposed (see for example Zeger (1988), Chan and Ledolter (1995) or Davis, Dunsmuir, and Wang (1999)). In this paper we will use the WinBUGS software (freely available on the BUGS-project website <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>) which estimates the parameters of the model with Markov Chain Monte Carlo (MCMC) methods. MCMC methods can be used in Bayesian Inference. In contrast to the classical approach, model parameters are considered to be random variables in Bayesian Inference. They are distributed according to a *prior distribution* which summarizes all a-priori knowledge about the parameters obtained for example by previous experiments or experts' assumptions. Then, given a random sample, the knowledge about the model parameters can be updated by using the new sample and by calculating the conditional distribution of the parameters given the new sample. This conditional distribution is commonly called *posterior distribution*. MCMC methods can be used to obtain an approximate sample from a complex posterior distribution by recording values of a simulated Markov Chain with a limiting distribution which is equal to the posterior distribution of the model parameters. The mean of these values generated by the Markov Chain can then be used as estimate for the model parameters. For a detailed introduction to MCMC estimation see for example Gamerman (1997) or Gilks, Richardson, and Spiegelhalter (1996). We will assess convergence of the MCMC simulation with the method proposed by Gelman and Rubin (1992).

As a third model we fit a Poisson-GLM with a non-standard variance, but without any autocorrelation structure to the data. We have also considered the use of the Negative Binomial distribution instead of the Poisson distribution. The fit of an ordinary Negative Binomial model (a model of the kind *NB2* as described in Cameron and Trivedi (1998), Chapter 3.3) leads to fairly similar results as the Poisson models concerning the influence of the explanatory variables on the variable

of interest as well as the estimated overdispersion of the data. The incorporation of an autoregressive latent process into the Negative Binomial model causes parameter identification problems. Since additional identification conditions would have to be completely arbitrary we discard the modelling approach with the Negative Binomial distribution and concentrate on the use of the Poisson distribution. We will shortly introduce the models considered in this paper in the following paragraphs.

### 3.1 The Poisson Generalized Linear Model

As already mentioned earlier, the Poisson regression model is a special case of the Generalized Linear Model (GLM). In a GLM, given the covariates  $\mathbf{x}_i$ , the response variables  $y_i$ ,  $i = 1, \dots, n$  are independent and identically distributed with a density of the form

$$f(y; \theta, \phi) = \exp \left\{ \frac{\theta y - b(\theta)}{a(\phi)} + c(y, \phi) \right\}, \quad (3.1)$$

which is (for a known  $\phi$ ) a density from the exponential family.  $\theta$  is a location parameter (not necessarily the mean) and  $\phi$  is a dispersion parameter. The probability density function of the Poisson distribution

$$f_{poi}(y) = \exp\{-\lambda + y \log \lambda - \log y!\}$$

can be rewritten in the form (3.1) with  $a(\phi) = 1$ ,  $\theta = \log \lambda$ ,  $b(\theta) = \exp(\theta) = \exp(\log \lambda) = \lambda$ , and  $c(y, \phi) = -\log y!$ . The function which links the regressors to the parameter(s) of the distribution of the variable of interest is called *link function*. In the case of the Poisson regression model the canonical link function is the log-link, i.e.  $\eta_t := \mathbf{x}_t^t \boldsymbol{\beta} = \log(\lambda_t)$  for  $t = 1, \dots, T$ . Therefore, the conditional mean of  $y_t$  given the covariates  $\mathbf{x}_t$  is given by

$$E(Y_t | \mathbf{x}_t) = \exp(\eta_t) = \exp(\mathbf{x}_t^t \boldsymbol{\beta}), \quad t = 1, \dots, T.$$

One important property of the Poisson distribution is equidispersion, i.e.

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

If in a data set the variance exceeds the mean, the data set is said to show *overdispersion*. For a detailed discussion of count data regression in general and the modelling of overdispersed count data in particular see for example Cameron and Trivedi (1998) or Winkelmann (2003).

### 3.2 Poisson-GLMs with an AR(1) latent process in the mean

This class of models can be considered as an extension of the Poisson-GLM described in the previous section when the observations  $Y_t$  come from a time series and are unlikely to be independent. The

general model framework as considered for example by Chan and Ledolter (1995) is given by:

$$\begin{aligned}
Y_t &\sim Poi(\lambda_t), \\
\lambda_t &= \exp(u_t + \mathbf{x}_t^t \boldsymbol{\beta}), \\
u_t &= \rho \cdot u_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\varepsilon^2)
\end{aligned} \tag{3.2}$$

where the observations  $Y_t$  conditional on  $\mathbf{x}_t$  and on the latent variables  $u_t$  are independent for  $t = 1, \dots, T$ . Since  $\{u_t\}$  is an ordinary AR(1) process, it follows that

$$\begin{aligned}
E(u_t) &= 0 \\
\sigma_u^2 := Var(u_t) &= \frac{\sigma_\varepsilon^2}{1 - \rho^2} \\
\gamma_u(k) := Cov(u_t, u_{t+k}) &= \rho^k \sigma_u^2.
\end{aligned}$$

Thus, in this model  $u_t \sim N(0, \sigma_u^2)$  and using the common formulas for the expectation and variance of the log-normal distribution it follows that

$$\begin{aligned}
E(\exp(u_t)) &= \exp\left(\frac{\sigma_u^2}{2}\right) \\
Var(\exp(u_t)) &= \exp(\sigma_u^2)(\exp(\sigma_u^2) - 1).
\end{aligned}$$

Using these results it is easy to see that

$$E(Y_t | \mathbf{x}_t) = E(E(Y_t | u_t, \mathbf{x}_t)) = E(\exp(u_t)) \cdot \exp(\mathbf{x}_t^t \boldsymbol{\beta}) = \exp\left(\frac{\sigma_u^2}{2} + \mathbf{x}_t^t \boldsymbol{\beta}\right). \tag{3.3}$$

Using the abbreviation  $\nu_t := \exp(\mathbf{x}_t^t \boldsymbol{\beta})$  the variance of  $Y_t$  conditional on the regressors (but not conditional on the latent process) can be calculated as:

$$Var(Y_t | \mathbf{x}_t) = \exp\left(\frac{\sigma_u^2}{2}\right) \cdot \nu_t + \exp(\sigma_u^2)(\exp(\sigma_u^2) - 1) \cdot \nu_t^2. \tag{3.4}$$

Furthermore, for  $k = 1, 2, \dots$  the autocovariance between  $Y_t$  and  $Y_{t+k}$  conditional only on the regressors can be calculated as :

$$Cov(Y_t, Y_{t+k} | \mathbf{x}_t, \mathbf{x}_{t+k}) = \nu_t \cdot \nu_{t+k} \cdot \exp(\sigma_u^2) \cdot (\exp(\rho^k \sigma_u^2) - 1). \tag{3.5}$$

These results are also stated in Zeger (1988) and Davis, Dunsmuir, and Wang (1999) for a slightly different parameterisation of the latent process.

Since  $Var(Y_t | \mathbf{x}_t) \geq E(Y_t | \mathbf{x}_t)$  the model allows for overdispersion. Note also that if  $\gamma_u(k) \neq 0 \implies$

$Cor(Y_t, Y_{t+k} | \mathbf{x}_t, \mathbf{x}_{t+k}) \neq 0$ , i.e. the latent process induces autocorrelation into  $Y_t$ . Since  $\gamma_u(k) = 0$  only when  $\rho = 0$ , autocorrelation in  $Y_t$  given the regressors is present if  $\rho \neq 0$ . Davis, Dunsmuir, and Wang (1999) elaborate that parameter-driven Poisson-GLMs with an AR(1) latent process are, using some approximations, closely related to observation-driven Poisson-GLARMA models, the Poisson equivalent of the models used by Rydberg and Shephard (2003) and Liesenfeld and Pohlmeier (2003).

### 3.3 A Poisson-GLM with heterogeneous variance structure

In the Poisson-GLM with an AR(1) latent process discussed in the previous paragraph 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 discuss another model which allows for a heterogeneous variance structure.

The model is given by:

$$\begin{aligned} Y_t &\sim Poi(\lambda_t) \\ \lambda_t &= \exp(\mathbf{x}_t^t \boldsymbol{\beta} + u_t) \\ u_t &= \exp\left(\frac{\tilde{\mathbf{x}}_t^t \boldsymbol{\alpha}}{2}\right) \cdot \varepsilon_t \\ \varepsilon_t &\stackrel{iid}{\sim} N(0, 1). \end{aligned} \tag{3.6}$$

with two vectors of explanatory variables  $\mathbf{x}_t$  and  $\tilde{\mathbf{x}}_t$  for  $t = 1, \dots, T$ . Since

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

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 the expectation and the variance of  $Y_t$  conditional on the regressors in the model (3.6) can be calculated similarly to the calculation of these quantities in the Poisson-GLM with an AR(1) latent process in (3.4) and (3.5):

$$\begin{aligned} E(Y_t | \mathbf{x}_t) &= E(E(Y_t | u_t, \mathbf{x}_t)) \\ &= E(\exp(u_t)) \cdot \exp(\mathbf{x}_t^t \boldsymbol{\beta}) \\ &= \exp\left(0.5e^{\tilde{\mathbf{x}}_t^t \boldsymbol{\alpha}} + \mathbf{x}_t^t \boldsymbol{\beta}\right) \end{aligned} \tag{3.7}$$

$$\begin{aligned} \text{Var}(Y_t|\mathbf{x}_t) &= \exp\left(0.5e^{\tilde{\mathbf{x}}_t^t\boldsymbol{\alpha}} + \mathbf{x}_t^t\boldsymbol{\beta}\right) + \\ &+ \exp\left(e^{\tilde{\mathbf{x}}_t^t\boldsymbol{\alpha}}\right) \cdot \left(\exp\left(e^{\tilde{\mathbf{x}}_t^t\boldsymbol{\alpha}}\right) - 1\right) \cdot \left(\exp(\mathbf{x}_t^t\boldsymbol{\beta})\right)^2 \end{aligned} \quad (3.8)$$

Equation (3.8) shows that this model allows for a heterogeneous variance, accounting for possible overdispersion in the data.

## 4 Model building and modelling results

We first fit an ordinary Poisson-GLM to the data presented in Section 2 and conduct a stepwise regressor selection (with the *step* function in S-Plus) to determine main- and interaction effects that have a significant influence on the absolute price changes of our option example on the 5%-level. We also investigate whether transformations of the explanatory variables (e.g. taking logarithms) improve the model fit. In order to be able to compare the results of the different regression approaches discussed in Section 3 we then maintain the transformations of the regressors for the fit of the more complex models.

### 4.1 Fit of an ordinary Poisson-GLM

The following table shows the results of an ordinary least squares estimation for the ordinary Poisson-GLM as described in Section 3.1 after transformation and standardization of the covariates and elimination of those regressors that are not significant on the 5%-level. The common S-Plus notation  $A:B$  for the interaction terms denotes the product of the regressors  $A$  and  $B$ . Note that the explanatory variable  $TTM$  denoting the remaining time to maturity of the option tested out completely.

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 4) 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 empirical 10%- and 90%-quantiles. The expected absolute option price change is calculated according to the results of Table 1 with all other explanatory variables set to their median values.

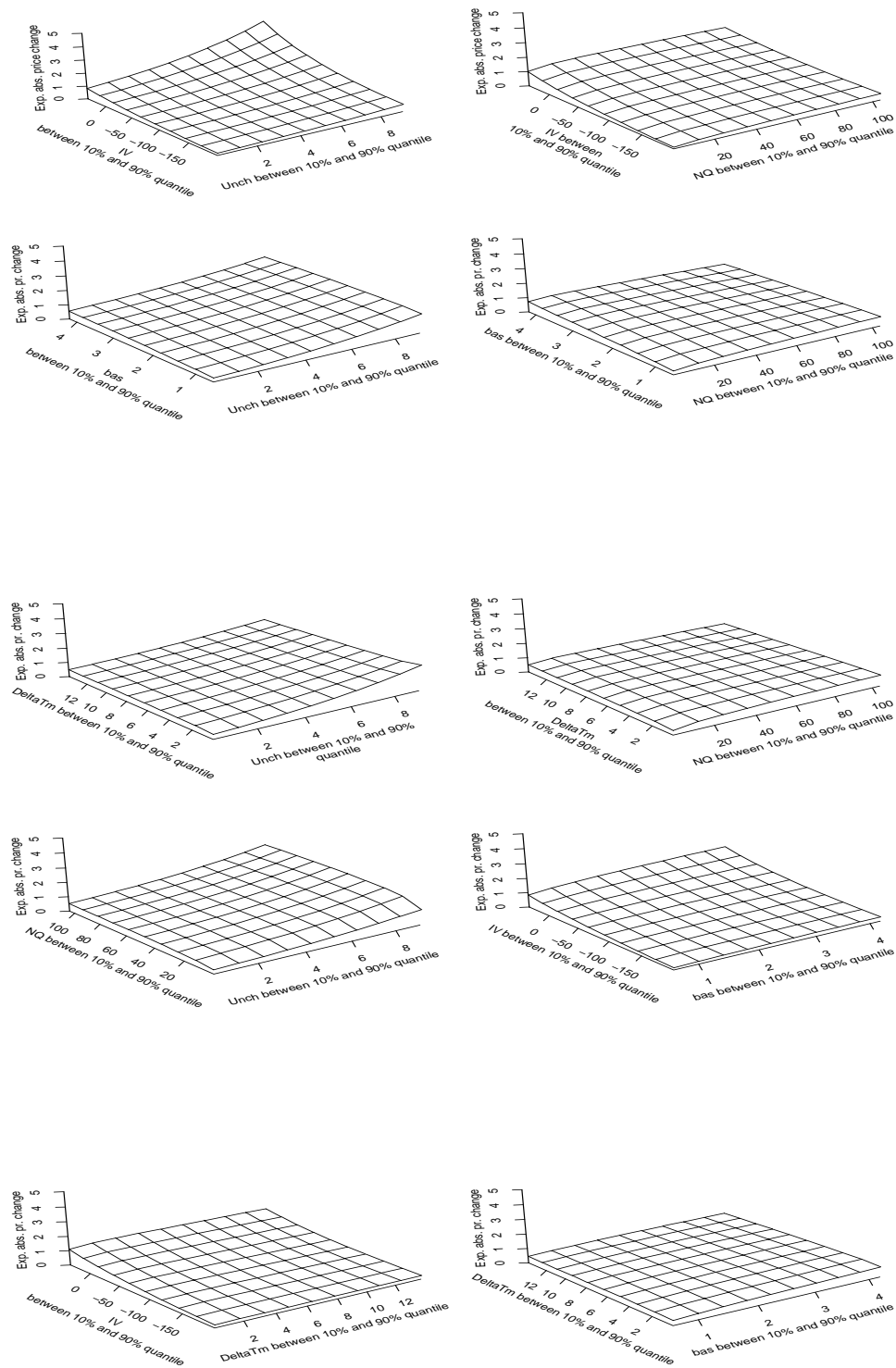


Figure 4: Fitted regression surfaces of the Poisson-GLM for the absolute price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003 when two regressors vary and the remaining regressors are set to their median values.



	Parameter	Est.	Std. error	t-value
main effects	$\beta_1$ (Intercept)	-0.51	0.03	-16.03
	$\beta_2$ (stand. $\log(UnCh + c_1)$ )	0.99	0.04	23.95
	$\beta_3$ (stand. $IV$ )	-2.50	1.41	-1.77
	$\beta_4$ (stand. $\log(NQ + c_2)$ )	0.29	0.07	4.36
	$\beta_5$ (stand. $\log(bas)$ )	9.64	0.93	10.37
	$\beta_6$ (stand. $\log(DeltaTm)$ )	10.06	1.41	7.15
inter-action effects	$\beta_7$ (stand. $IV : \log(NQ + c_2)$ )	-0.28	0.06	-4.38
	$\beta_8$ (stand. $\log(bas) : \log(UnCh + c_1)$ )	-9.29	0.93	-9.98
	$\beta_9$ (stand. $IV : \log(UnCh + c_1)$ )	3.46	1.43	2.42
	$\beta_{10}$ (stand. $\log(bas) : \log(NQ + c_2)$ )	-0.13	0.05	-2.73
	$\beta_{11}$ (stand. $\log(DeltaTm) : \log(UnCh + c_1)$ )	-9.91	1.42	-7.00
	$\beta_{12}$ (stand. $\log(NQ + c_2) : \log(DeltaTm)$ )	-0.14	0.05	-2.94

Table 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 and transformed regressors ( $c_1 = 80$ ,  $c_2 = 1$ ).

Figure 4 shows that, as a very general rule, it can be said that higher values of all of the significant explanatory variables ( $UnCh$ ,  $IV$ ,  $NQ$ ,  $bas$  and  $DeltaTm$ ) lead to higher expected absolute option price changes. 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, of course, in line with economic considerations.

## 4.2 Fit of a Poisson-GLM with a latent AR(1)-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 and estimate the parameters with MCMC methods in WinBUGS. The model as given in (3.2) can be written as:

$$Y_t \sim Poi(\lambda_t) \quad (4.9)$$

$$\lambda_t = exp(Z_t) \quad (4.10)$$

$$Z_t = \mathbf{x}_t^t \boldsymbol{\beta} + u_t \quad (4.11)$$

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t \quad (4.12)$$

$$\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

In order to implement the model in WinBUGS the following transformations have to be made:

From Equations (4.11) and (4.12) 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,

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

Since, as already discussed in Section 3.2

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

we get:

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

These considerations at hand, the model can be implemented in WinBUGS. As priors we choose 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  $\rho$ :

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

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 statistics (plots not shown) justify these choices. In a second simulation we investigate prior sensitivity for the parameter  $\rho$  and use an informative  $N(0, 0.3^2)$  prior truncated to  $[-1, 1]$  for  $\rho$ . The MCMC simulation leads to similar results (with slower convergence of the simulation), so that the prior sensitivity of  $\rho$  can be assumed to be negligible.

Table 2 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  $\rho$ ).

	Parameter	mean	Std. err.	2.5%	med.	97.5%
main effects	$\beta_1$ (Intercept)	-0.59	0.09	-0.76	-0.59	-0.42
	$\beta_2$ (stand. $\log(UnCh + c_1)$ )	0.99	0.05	0.90	0.99	1.08
	$\beta_3$ (stand. $IV$ )	-0.85	1.59	-3.95	-0.81	2.13
	$\beta_4$ (stand. $\log(NQ + c_2)$ )	0.34	0.07	0.21	0.34	0.48
	$\beta_5$ (stand. $\log(bas)$ )	8.89	1.13	6.65	8.92	11.06
	$\beta_6$ (stand. $\log(DeltaTm)$ )	8.52	1.50	5.50	8.55	11.35
inter-action effects	$\beta_7$ (stand. $IV : \log(NQ + c_2)$ )	0.12	0.19	-0.23	0.11	0.50
	$\beta_8$ (stand. $\log(bas) : \log(UnCh + c_1)$ )	-8.56	1.13	-10.73	-8.56	-6.35
	$\beta_9$ (stand. $IV : \log(UnCh + c_1)$ )	1.61	1.58	-1.35	1.57	4.66
	$\beta_{10}$ (stand. $\log(bas) : \log(NQ + c_2)$ )	-0.14	0.05	-0.24	-0.14	-0.05
	$\beta_{11}$ (stand. $\log(DeltaTm) : \log(UnCh + c_1)$ )	-8.26	1.52	-11.13	-8.30	-5.17
	$\beta_{12}$ (stand. $\log(NQ + c_2) : \log(DeltaTm)$ )	-0.22	0.05	-0.33	-0.22	-0.13
	$\rho$	0.97	0.02	0.93	0.97	0.99
	$\sigma_\varepsilon^2$	0.009	0.002	0.006	0.009	0.014

Table 2: Estimated posterior mean , std. error, 2.5% quantile, median and 97.5% quantile for the parameters of 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  $\beta_7$  and  $\beta_9$  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  $\beta_3$  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  $\beta_3$  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  $\rho$  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 be expected to be rather small. This can be confirmed by considering in our MCMC simulation the 2419 posterior mean estimates of  $u_t$ ,  $t = 1, \dots, 2419$  which lie within the interval  $[-0.84, 0.68]$ , while those for  $\mathbf{x}_t^t \boldsymbol{\beta}$  lie within the considerably larger interval  $[-4.92, 3.68]$ . The lowest lower bound of all 2419 95%-credible intervals of  $\hat{u}_t$ ,  $t = 1, \dots, 2419$ , 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  $\sigma_u^2$  and  $\mu_t$  in (3.3), (3.4) and (3.5):

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

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

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

which yields

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

for  $k = 1$ .

These results indicate that there is overdispersion and autocorrelation among the responses  $Y_t$  induced by the large autocorrelation present in the latent process.

In order to be able to interpret the estimation results plots similar to those for the ordinary Poisson-GLM in Figure 4 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 can be considered. These plots (not shown for the Poisson-GLM with an Ar(1) latent process) look very similar to those in Figure 4 and thus, the general conclusions about the influence of the explanatory variables on the absolute option price changes remain valid.

### 4.3 Fit of a Poisson-GLM with heterogeneous variance structure

We finally fit a model of the form (3.6) to the data. As regressors  $\tilde{\mathbf{x}}_t$  for the heterogeneous variance term 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  $(1, \text{stand. } \log(\text{bas}_t))^t$ .

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 (3.6) to

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

In the simulation the parameter  $\rho$  is not  $\neq 0$  on a 5%-credible level and we therefore discard this model in favour of the less complex Model (3.6).

We finally run 3 independent chains in order to estimate the parameters of the model (3.6) and again, reduce autocorrelations within the chains by only storing every 100th value. The trajectories of the chains and the Gelman-Rubin statistic 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  $\boldsymbol{\alpha}$ :

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

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 3 shows the results of the final MCMC simulation.

Again the plots similar to those in Figure 4 (ordinary Poisson-GLM) for the model (3.6) with heterogeneous variance structure (plots not shown) 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 paragraphs.

	Parameter	mean	Std. err.	2.5%	med.	97.5%
main effects	$\beta_1$ (Intercept)	-0.56	0.04	-0.63	-0.56	-0.49
	$\beta_2$ (stand. $\log(UnCh + c_1)$ )	0.90	0.04	0.82	0.90	0.98
	$\beta_3$ (stand. $IV$ )	-2.01	1.34	-4.62	-2.01	0.69
	$\beta_4$ (stand. $\log(NQ + c_2)$ )	0.41	0.07	0.27	0.41	0.56
	$\beta_5$ (stand. $\log(bas)$ )	4.51	1.49	1.44	4.51	7.38
	$\beta_6$ (stand. $\log(DeltaTm)$ )	6.35	1.48	3.52	6.41	9.17
inter-action effects	$\beta_7$ (stand. $IV : \log(NQ + c_2)$ )	-0.27	0.07	-0.41	-0.27	-0.13
	$\beta_8$ (stand. $\log(bas) : \log(UnCh + c_1)$ )	-4.12	1.49	-7.01	-4.15	-1.07
	$\beta_9$ (stand. $IV : \log(UnCh + c_1)$ )	2.99	1.36	0.24	2.99	5.65
	$\beta_{10}$ (stand. $\log(bas) : \log(NQ + c_2)$ )	-0.24	0.07	-0.37	-0.24	-0.09
	$\beta_{11}$ (stand. $\log(DeltaTm) : \log(UnCh + c_1)$ )	-6.16	1.50	-9.04	-6.22	-3.32
	$\beta_{12}$ (stand. $\log(NQ + c_2) : \log(DeltaTm)$ )	-0.24	0.05	-0.35	-0.24	-0.14
	$\alpha_1$ (Intercept)	-3.02	0.38	-3.96	-2.97	-2.41
	$\alpha_2$ (stand. $\log(bas)$ )	0.94	0.19	0.60	0.93	1.32

Table 3: Estimated posterior mean , std. error, 2.5% quantile, median and 97.5% quantile for the parameters of the Poisson-GLM with heterogeneous error term variance (3.6) 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 .

## 5 Assessment of model adequacy

When comparatively assessing the adequacy of different statistical models for a given data set, the complexity of a certain model should be traded off against its goodness of fit.

One possibility to assess the goodness of fit of the previously discussed models is the consideration of deviances. The deviance of an ordinary Poisson-GLM is defined by (see for example McCullagh and Nelder (1989)):

$$D(\mathbf{y}, \hat{\boldsymbol{\lambda}}) := -2 \cdot [l(\mathbf{y}, \hat{\boldsymbol{\lambda}}) - l(\mathbf{y}, \mathbf{y})] \quad (5.13)$$

where  $Y_t \sim Poi(\lambda_t)$ ,  $\lambda_t = \exp(\mathbf{x}_t^t \boldsymbol{\beta})$ , independent for  $t = 1, 2, \dots, T$ . Here  $l$  denotes the log-likelihood of the Poisson-GLM and  $\hat{\boldsymbol{\lambda}}$  is the corresponding maximum-likelihood estimate. 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. Spiegelhalter et al. (2002) extend the concept of residual deviances to Bayesian models and define a Bayesian deviance. For all Poisson-models considered in this paper the Bayesian deviance as defined by Spiegelhalter et al. (2002) is given by:

$$D(\boldsymbol{\lambda}) := -2 \cdot \log[p(\mathbf{y}|\boldsymbol{\lambda})] + 2 \cdot \log[p(\mathbf{y}|\mathbf{y})] \quad (5.14)$$

where  $p(\mathbf{y}|\boldsymbol{\lambda})$  denotes the probability function of the data given the vector of mean parameters  $\boldsymbol{\lambda}$ . Note that (5.13) and (5.14) are equal in the Poisson-GLM if  $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}$ . Spiegelhalter et al. (2002)

consider the mean deviance

$$\overline{D(\boldsymbol{\lambda})} := \frac{1}{R-B} \sum_{r=B}^R D(\boldsymbol{\lambda}^{(r)}),$$

where  $\boldsymbol{\lambda}^{(r)}$  denotes the  $r$ -th MCMC iterate of  $\boldsymbol{\lambda}$ , as a Bayesian measure of fit and the quantity

$$p_D := \overline{D(\boldsymbol{\lambda})} - D(\overline{\boldsymbol{\lambda}}),$$

the mean deviance minus the deviance of the means, as the "effective number of parameters" of the Bayesian model. Here,  $\overline{\boldsymbol{\lambda}} := \frac{1}{R-B} \sum_{r=B}^R \boldsymbol{\lambda}^{(r)}$  denotes the posterior mean estimate of  $\boldsymbol{\lambda}$ . As a criterion for the comparison of two Bayesian models they finally suggest the *deviance information criterion* which is given by:

$$DIC := D(\overline{\boldsymbol{\lambda}}) + 2p_D = \overline{D(\boldsymbol{\lambda})} + p_D. \quad (5.15)$$

Thus, the *DIC* is the Bayesian measure of fit  $\overline{D(\boldsymbol{\lambda})}$ , penalized by an additional complexity term  $p_D$ , giving a Bayesian analogue to Akaike's AIC criterion. 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  $D(\overline{\boldsymbol{\lambda}})$  of the Bayesian Poisson-GLM is equal to the deviance of the classical model setting as calculated according to (5.13). In Table 4 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 this paper. The deviances in Column 2 and Column 3 of Table 4 indicate that 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 considerably 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.

Model	$D(\bar{\theta})$	$\overline{D(\theta)}$	$p_D$	$DIC$
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 4: Deviances and DIC values for the Poisson option data models.

## 6 Conclusions and discussion

In this paper we have statistically analyzed the process of the absolute non-zero price changes of an option on the XETRA DAX index based on quote-by-quote data from the EUREX exchange 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 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" of our models, suggesting that overdispersion in our data is better modelled by an autoregressive component than by a heterogeneous variance structure.

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 plausible 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 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. In the context of market microstructure theory Diamond and Verrecchia (1987) give a possible explanation for this. Provided that short-selling is not feasible, they consider longer times between transactions as a possible signal for 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 pure 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. 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 paper. The incorporation of the direction of the price changes in order to construct a "parameter-driven equivalent" of the models proposed by Rydberg and Shephard (2003) and Liesenfeld and Pohlmeier (2003), which should then be applied to real trade-by-trade data (including trades that do not change the price of the security), can be an interesting topic for further research. Another interesting result is the insignificance of the remaining time to maturity of the option. While this variable plays an important role in the determination of the price of an option calculated according to the Black and Scholes formula, it has not had a significant impact on

the size of the price changes of our option based on the available quotation data. Thus, according to our results, the driving force for a rise in short-time volatility when the option gets close to maturity (a phenomenon that can often be observed in option markets) must be the frequency of the price changes rather than their size.

A major weakness of our modelling approach is the fact that we do not model the price changes jointly with other marks of the trading process, but consider the price changes only conditional on a set of information coming from other marks of the trading process, which enter as regressors into our modelling. Using the prediction decomposition the joint distribution of the price changes  $Y_t$ ,  $t = 1, \dots, T$  and the explanatory variables  $\mathbf{X}_t$ ,  $t = 1, \dots, T$  can be written in the following form:

$$f(y_1, \dots, y_T, \mathbf{x}_1, \dots, \mathbf{x}_T) = \prod_{t=1}^T f(y_t | \mathbf{x}_t, F_{t-1}^{y, \mathbf{x}}) f(\mathbf{x}_t | F_{t-1}^{y, \mathbf{x}})$$

where

$$F_t^{y, \mathbf{x}} := \sigma(Y_s, \mathbf{X}_s; s \leq t)$$

This approach is also stated in Rydberg and Shephard (2003). In this paper we have only focused on the modelling of  $f(y_t | \mathbf{x}_t, F_{t-1}^{y, \mathbf{x}})$ . The extension of our modelling approach into this broader context will be a topic of further research.

## References

- Bauwens, L. and P. Giot (2001). *Econometric Modelling of Stock Market Intraday Activity*. Boston: Kluwer Academic Publishers.
- Cameron, A. and P. Trivedi (1998). *Regression analysis of count data*. Cambridge: Cambridge University Press.
- 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.
- Chib, S. (2001). Markov Chain Monte Carlo Methods: Computation and Inference. *In: Handbook of Econometrics* 5, Chapter 57.
- Cox, D. (1993). Statistical analysis of time series, some recent developments. *Scandinavian Journal of Statistics* 8, 93–115.
- Dacorogna, M., R. Gencay, U. Müller, R. Olsen, and O. Pictet (2001). *An Introduction to High Frequency Finance*. San Diego: Academic Press.

- 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). Constraints 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.
- 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.
- Gilks, W., S. Richardson, and D. Spiegelhalter (1996). *Markov Chain Monte Carlo in practice*. London: Chapman & Hall.
- Hautsch, N. and W. Pohlmeier (2002). Econometric Analysis of Financial Transaction Data: Pitfalls and Opportunities. *Allg. Stat. Archiv* 86, 5–30.
- Liesenfeld, R. and W. Pohlmeier (2003). A Dynamic Integer Count Data Model for Financial Transaction Prices. *Working Paper, University of Konstanz*.
- McCullagh, P. and J. Nelder (1989). *Generalized Linear Models - Second edition*. London: Chapman & Hall.
- Rydberg, T. and N. Shephard (2003). Dynamics of Trade-by-Trade Price Movements: Decomposition and Models. *Journal of Financial Econometrics* 1, 2–25.
- 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.
- Winkelmann, R. (2003). *Econometric Analysis of Count Data*. New York: Springer.
- Zeger, S. (1988). A Regression Model for Time Series of Counts. *Biometrika* 75, 621–629.