Model-based Quantification of the Volatility of Options at Transaction Level with Extended Count Regression Models

Claudia Czado *† Andreas Kolbe ‡

April 20, 2006

This is a preprint of an article published in *Applied Stochastic Models in Business and Industry, Vol. 23, Issue 1, p. 1-21*

Abstract

In this paper we elaborate how Poisson regression models of different complexity can be used in order to model absolute transaction price changes of an exchange-traded security. When combined with an adequate autoregressive conditional duration model, our modelling approach can be used to construct a complete modelling framework for a security's absolute returns at transaction level and thus for a model-based quantification of intraday volatility and risk. We apply our approach to absolute price changes of an option on the XETRA DAX index based on quote-by-quote data from the EUREX exchange and find that within our Bayesian framework a Poisson Generalized Linear Model (GLM) with a latent AR(1) process in the mean is the best model for our data according to the deviance information criterion (DIC). 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.

Keywords: index options, quotation data, absolute returns, Poisson regression, autocorrelation, Markov Chain Monte Carlo, DIC

---

*Center for Mathematical Sciences, Munich University of Technology, Boltzmannstr. 3, D-85747 Garching b. München, Germany, cczado@ma.tum.de, Phone: +49/89/28917428, Fax: +49/89/28917435

†The first author's research was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 386: Statistische Analyse diskreter Strukturen. Both authors would like to thank an anonymous referee for his/her valuable comments.

‡Center for Mathematical Sciences, Munich University of Technology, Boltzmannstr. 3, D-85747 Garching b. München, Germany, kolbe@ma.tum.de, Phone: +49/89/28917417

1 Introduction, motivation and related research

The investigation and statistical modelling of ultra-high frequency financial data has become a new focus of academic interest since quotation and transaction data provide detailed information about the trading and asset pricing process. 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 and Russell (1998), 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. If adequately combined with an ACD model, as e.g. in the recent ACM-ACD model of Engle and Russell (2005), models for transaction price changes can be used to construct a complete modelling framework for a security's price process in continuous time. 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 statistical modelling of the last component of the previously described approaches and consider only absolute non-zero transaction price changes. Our modelling approach can then be combined with an adequate ACD model in order to give a complete modelling framework for a security's absolute returns and thus for a model-based quantification of intraday volatility and risk. In the context of the modelling of asset price volatility, the empirical process of absolute intraday returns (or absolute price changes, when measuring returns in absolute values) have proven to be a good empirical measure for the asset's instantaneous volatility (see e.g. Forsberg and Ghysels (2004)). For example, in order to estimate parameters in a stochastic volatility model, one can observe the asset's absolute return or price change in a time interval $\Delta t := t_1 - t_0$ and consider this empirical measure as a proxy for

$$\sigma_{\Delta t} := \int_{t_0}^{t_1} \sigma(s) ds,$$

where $\sigma(s)$ denotes the instantaneous volatility of the asset. In this paper it is our aim to model absolute price changes at transaction level within a regression framework, which allows us to identify possible drivers of the instantaneous volatility and to quantify their influence.

While Rydberg and Shephard (2003), Liesenfeld and Pohlmeier (2003) as well as Engle and Russell (2005) use observation-driven time series models, according to the classification of Cox (1993), parameter-driven Poisson models are used in this paper. While observation-driven models offer the advantage of an easy to handle parameter estimation and of straightforward prediction in most cases, the regression parameters in parameter-driven models are in general easier to interpret. Since, as previously mentioned, one focus of this paper is on the effects of explanatory variables we choose the class of parameter-driven models for our modelling. In particular we consider a Poisson-GLM with an AR(1) latent process in the mean as specified by Zeger (1988), a Poisson-GLM with heterogeneous variance structure and, for comparison purposes, an ordinary Poisson-GLM. 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 et al. (1999). Davis et al. (2003) discuss observation-driven models for time series of counts (in particular the Poisson-GLARMA model, the Poisson-equivalent of the models used by Rydberg and Shephard (2003) and Liesenfeld and Pohlmeier (2003)) while Zeger (1988) and Chan and Ledolter (1995) explicitly discuss and develop estimation techniques for the parameter-driven Poisson-GLM with an autoregressive latent process in the mean. In recent work, Heinen (2003) provides a framework for the modelling of time series of counts with the double-Poisson distribution and extends the approach in Heinen and Rengifo (2003) to the multivariate case using copulas. In a Bayesian context, time series of counts are also considered in Durbin and Koopman (2000).

Applications in the field of finance and financial econometrics have had their focus on stock markets in other publications. While our approach can of course be applied to a stock traded at a stock exchange as well, 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. To our best knowledge, models for the price process of a security at transaction level have not been applied to exchange-traded options, yet. One may argue that the reason for this is the fact that the price process of an option (and with it the volatility of the option) is determined by the price process of the underlying which is the case in a perfect Black-Scholes world where the (deterministic) volatility of the underlying is known. Since we consider the option price process at transaction level, where market microstructure effects and noise may be present, we do not rely on the assumptions of the Black-Scholes pricing world, but consider directly the price process of the option and incorporate the process of the underlying into the modelling framework through explanatory variables such as the intrinsic value of the option at the time of the trade, the (absolute) price changes of the underlying and the remaining time to maturity of the option. In addition to these "fundamental" explanatory variables closely related to option pricing theory, we also consider "microstructural" explanatory variables, such as 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. With this application we also provide some empirical contributions to the understanding of the market microstructure of option markets which may not be new or too surprising on the qualitative side. However, we deliver a framework for the quantification of some of the effects related to the previously mentioned explanatory variables, also taking into account nonlinear and interaction terms, which is usually not done in other publications.

Since we are dealing with a count data structure, the Poisson distribution is a natural distributional choice (see Section 3 for some further discussion of the adequacy of the Poisson distributional assumption). Particularly for the Poisson-GLM with an AR(1) latent process in the mean, the likelihood features high-dimensional integrals introduced through the latent process, which makes maximum-likelihood based estimation techniques hard to handle. Yet, by using the WinBUGS software package, parameter estimation with Markov Chain Monte Carlo (MCMC) methods is well feasible. We will thus choose the MCMC approach in this paper. MCMC techniques are for example summarized in Chib (2001) and discussed in detail in Gamerman (1997). Gilks et al. (1996) provide examples of applications from a wide range of fields. It is, of course, not straightforward to compare the adequacy of our different model specifications, especially in a Bayesian context. We use an information criterion proposed by Spiegelhalter et al. (2002) in order to finally compare and assess the adequacy of the different models. For the computation the S-Plus software package (in addition to WinBUGS) is used.

The paper is organized as follows: The available data for the application of our modelling approach is presented in Section 2. In Section 3 we shortly introduce and discuss the regression models which we use for the modelling of the data and give empirical results. The adequacy of the different model specifications is comparatively discussed in Section 4. Finally, Section 5 concludes, gives some interpretations of the empirical results and discusses weaknesses of our modelling approach.

2 Data presentation

The available data set for the application 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.

2.1 The variable of interest: Absolute option price changes

We analyze the changes of the quoted last prices of the previously described Call option and relate these changes to a series of explanatory variables. As already elaborated in Section 1, we model the absolute values of the price changes which can be used as an empirical volatility measure. 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 $t = 1, 2, ..., T$ 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.

2.2 Explanatory variables

We now introduce both fundamental and microstructural 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 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 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 τ_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	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	

Histograms of all of the possible explanatory variables are given in Figure 3.

Fundamental explanatory variables related to option pricing theory

- *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 $t - 1$.

- *Intrinsic value of the option at the time of the transaction*:

The intrinsic value of a Call option at the point of time τ 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 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.

- *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 matured on 21-March 13:00:00. Thus, for example, TTM_{t+2} is equal to 2 min 54 sec which yields a value of 0.0020139 [in days] for TTM_{t+2}.

Microstructural explanatory variables

- *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 $DeltaTm_{t+1}$.

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, as expected.

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 regression of count data structures. Of course, there is no clear evidence that the data are actually Poisson-distributed. Figure 2 indicates, however, that a count data approach should be used for the modelling (rather than, e.g., an ordered probit approach since there are too many categories). As an alternative to the Poisson distribution we have also considered the Negative Binomial distribution (in particular, a model of the kind *NB2* as described in Cameron and Trivedi (1998), Chapter 3.3). Yet, 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. Negative Binomial models are usually used to account for overdispersion. However, overdispersion is also accounted for in the extended Poisson models which we use in this paper (see (3.6) and (3.10)). Furthermore, even if the 'true' distribution of the response Y_t given the regressors $\mathbf{x}_t$ is not Poisson, this does not lead to a serious model misspecification since we are mainly interested in modelling the mean and variance. Now, thinking in terms of pseudo maximum-likelihood, it is well-known that the Poisson assumption leads to consistent estimators of those.

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 et al. (1999)). In this paper, as previously mentioned, 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.
As a third model we fit a Poisson-GLM with a non-standard variance, but without any autocorrelation structure to the data. In the following paragraphs we will introduce our modelling approaches and state the results of their application to the option data.

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) with the log-link as canonical link function, i.e.

$$\begin{aligned} Y_t &\sim Poi(\lambda_t) \\ \lambda_t &= exp(\mathbf{x}_t^t \boldsymbol{\beta}),\ t = 1, .., T. \end{aligned} \tag{3.1}$$

The conditional mean of Y_t given the covariates $\mathbf{x}_t$ is given by

$$E(Y_t \mid \mathbf{x}_t) = \exp(\mathbf{x}_t^t \boldsymbol{\beta}),\ t = 1, .., T. \tag{3.2}$$

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

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

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).
Using explorative data analysis methods such as plotting covariate combinations versus the log-response we identify the covariate transformations and interaction effects to be investigated further. In order to be able to compare the results of the different regression approaches we then maintain the transformations of the regressors for the fit of the more complex models in the following. Table

2 shows the classical ML-estimation results for the ordinary Poisson-GLM after transformation and standardization of the covariates and elimination of those regressors that are not significant at 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 has been deleted due to lack of statistical significance. Since we want to compare these results with those of the other modelling approaches in a Bayesian setting, we also estimate the parameters with a MCMC simulation in WinBUGS. Similarly to the simulations carried out for the other models discussed in the following sections, we run 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 virtually identical to those obtained in the classical approach with ML-estimation stated in Table 2.

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 2 with all other explanatory variables set to their median values. 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 the fundamentals of option pricing theory.

3.2 A Poisson-GLM 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:

$$\begin{aligned} Y_t &\sim Poi(\lambda_t) \\ \lambda_t &= exp(Z_t) \\ Z_t &= \mathbf{x}_t^t \boldsymbol{\beta} + u_t \\ u_t &= \rho \cdot u_{t-1} + \varepsilon_t \\ \varepsilon_t &\overset{iid}{\sim} N(0, \sigma_\varepsilon^2), \end{aligned} \tag{3.3}$$

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

$$\begin{aligned} E(u_t) &= 0 \\ \sigma_u^2 := Var(u_t) &= \frac{\sigma_\varepsilon^2}{1-\rho^2} \\ \gamma_u(k) := Cov(u_t, u_{t+k}) &= \rho^k \sigma_u^2. \end{aligned} \tag{3.4}$$

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

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

Using these results it is easy to see that

$$E(Y_t|\mathbf{x}_t) = E(E(Y_t|u_t, \mathbf{x}_t)) = E(exp(u_t)) \cdot exp(\mathbf{x}_t^t \boldsymbol{\beta}) = exp\left(\frac{\sigma_u^2}{2} + \mathbf{x}_t^t \boldsymbol{\beta}\right). \tag{3.5}$$

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:

$$Var(Y_t|\mathbf{x}_t) = exp(\frac{\sigma_u^2}{2}) \cdot \nu_t + exp(\sigma_u^2)(exp(\sigma_u^2) - 1) \cdot \nu_t^2. \tag{3.6}$$

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

$$Cov(Y_t, Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}) = \nu_t \cdot \nu_{t+k} \cdot exp(\sigma_u^2) \cdot (exp(\rho^k \sigma_u^2) - 1). \tag{3.7}$$

These results are also stated in Zeger (1988) and Davis et al. (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 $\gamma_u(k) \neq 0 \Longrightarrow Cor(Y_t, Y_{t+k}|\mathbf{x}_t, \mathbf{x}_{t+k}) \neq 0$, 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 et al. (1999) show 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).

The likelihood of the Model (3.3) is given by:

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

where $\alpha_t := exp(u_t)$ and f^{Poi} and f^{LogN} denote the probability function of the Poisson distribution and the p.d.f of the log-normal distribution respectively. Due to the high-dimensional integrals introduced through the latent process, a direct evaluation of the likelihood function is hard to handle. We therefore resort to MCMC estimation in WinBUGS, as previously mentioned. For this purpose the following transformations have to be made: From the model specification (3.3) 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,

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

Due to (3.4) 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 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 statistics (plots not shown) 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 be assumed to be negligible. Table 3 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). Here, the parameters β_7 and β_9 are not different from 0 at 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 different from 0 at a 5%-credible level either, it has to be investigated whether the explanatory variable 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 be expected to be rather small. This can be confirmed by considering in our MCMC simulation the 2419 posterior mean estimates of u_t, $t = 1, ..., 2419$ 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 $\hat{u}_t$, $t = 1, ..., 2419$, 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 (3.5), (3.6) and (3.7):

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

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

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

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

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:

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

with two vectors of explanatory variables $\mathbf{x}_t$ and $\tilde{\mathbf{x}}_t$ for $t = 1, ..., T$. 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. In a first MCMC simulation the parameter ρ is not different from 0 at a 5%-credible level and we therefore discard this model in favour of the less complex model

$$\begin{aligned} Y_t &\sim Poi(\lambda_t) \\ \lambda_t &= exp(\mathbf{x}_t^t \boldsymbol{\beta} + u_t) \\ u_t &= exp\left(\frac{\tilde{\mathbf{x}}_t^t \boldsymbol{\alpha}}{2}\right) \cdot \varepsilon_t \\ \varepsilon_t &\overset{iid}{\sim} N(0,1), \end{aligned} \tag{3.8}$$

without the autoregressive component. Since in the Model (3.8) $u_t \sim N\left(0, exp(\tilde{\mathbf{x}}_t^t \boldsymbol{\alpha})\right)$, 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 the expectation and the variance of Y_t conditional on the regressors can be calculated similarly to the calculation of these quantities in the Poisson-GLM with an AR(1) latent process in (3.6) and (3.7):

$$E(Y_t|\mathbf{x}_t) = E(E(Y_t|u_t, \mathbf{x}_t)) = E(e^{u_t}) \cdot e^{\mathbf{x}_t^t \boldsymbol{\beta}} = exp\left(0.5e^{\tilde{\mathbf{x}}_t^t \boldsymbol{\alpha}} + \mathbf{x}_t^t \boldsymbol{\beta}\right) \tag{3.9}$$

$$\begin{aligned} Var(Y_t|\mathbf{x}_t) \quad = \quad & exp\left(0.5e^{\tilde{\mathbf{x}}_t^t \boldsymbol{\alpha}} + \mathbf{x}_t^t \boldsymbol{\beta}\right) + \\ & + exp\left(e^{\tilde{\mathbf{x}}_t^t \boldsymbol{\alpha}}\right) \cdot \left(exp\left(e^{\tilde{\mathbf{x}}_t^t \boldsymbol{\alpha}}\right) - 1\right) \cdot \left(exp(\mathbf{x}_t^t \boldsymbol{\beta})\right)^2. \end{aligned} \tag{3.10}$$

Equation (3.10) shows that this model allows for a heterogeneous variance, accounting for possible overdispersion in the data. Concerning the parameter vector $\boldsymbol{\alpha}$, our MCMC simulations yield that only the intercept term and the Bid-Ask-spread are significant at a 5%-credible level, so that we reduce the vector $\tilde{\mathbf{x}}_t$ to $(1, \text{stand. } log(bas_t))^t$. We finally run 3 independent chains in order to estimate the parameters of the model (3.8) 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 $\boldsymbol{\alpha}$:

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

We had to choose slightly more 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 4 shows the results of the final MCMC simulation.

4 Assessment of model adequacy

While the estimation results of all of the three previously discussed models lead to the same general conclusions about the influence of the explanatory variables on the absolute option price changes (in fact, the fitted regression surfaces as shown in Figure 4 for the simple Poisson-GLM look very similar in shape for each of the three models), the question arises which of the discussed models is the most adequate one. A first explorative answer is given by the consideration of residuals. In

Figure 5 we plot the expected values of the variable of interest (conditional on the explanatory variables) in each model according to (3.2), (3.5) and (3.9) against the actual observations. As a further illustration of the goodness of fit of our models we show the last 50 residuals (i.e. the difference between the observation and the expected value of the variable of interest according to the respective model). The plots suggest that the more complex models, particularly the Poisson-GLM with the heterogeneous variance structure seem to fit our data better. However, 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 quantitatively 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)):

$$D(\mathbf{y}, \hat{\boldsymbol{\lambda}}) := -2 \cdot [l(\mathbf{y}, \hat{\boldsymbol{\lambda}}) - l(\mathbf{y}, \mathbf{y})], \tag{4.11}$$

where $Y_t \sim Poi(\lambda_t)$, $\lambda_t = exp(\mathbf{x}_t^t \boldsymbol{\beta})$, independent for $t = 1, 2, ..., T$. 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:

$$D(\boldsymbol{\lambda}) := -2 \cdot \log[p(\mathbf{y}|\boldsymbol{\lambda})] + 2 \cdot \log[p(\mathbf{y}|\mathbf{y})], \tag{4.12}$$

where $p(\mathbf{y}|\boldsymbol{\lambda})$ denotes the probability function of the data given the vector of mean parameters $\boldsymbol{\lambda}$. Note that (4.11) and (4.12) 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:

$$DIC := D(\overline{\boldsymbol{\lambda}}) + 2p_D = \overline{D(\boldsymbol{\lambda})} + p_D. \tag{4.13}$$

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. The DIC is a recent alternative to Bayesian model comparison by Bayes factors. Bayes factors have the disadvantage that they are sometimes hard to compute, which is especially true for our Poisson-GLM with an AR(1) latent process in the mean. Computation of Bayes factors for this model would again involve integrating out the latent variables. We therefore choose the DIC, which is readily available within the MCMC simulation, as our criterion for model choice.

In Table 1 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. Note 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 (4.11). The deviances in Column 2 and Column 3 of Table 1 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	$D(\overline{\boldsymbol{\lambda}})$	$\overline{D(\boldsymbol{\lambda})}$	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 1: Deviances and DIC values for the Poisson option data models.

5 Conclusions and discussion

In this paper we have elaborated how Poisson regression models of different complexity can be used to model the absolute returns of an exchange-traded security at transaction level. We suggest using them together with an adequate ACD model in order to give a complete modelling framework for a security's absolute returns and thus for a model-based quantification of intraday volatility and risk. We have applied our modelling approaches to 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 identified drivers of the instantaneous volatility of the option and quantified their influence. 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. In some applications a prediction of the instantaneous volatility may also be of interest. We have not addressed this issue in this paper. Prediction of instantaneous volatility based on our modelling approach would imply the prediction of the processes underlying the explanatory variables in a first step. Whether this may finally lead to a good prediction of absolute price changes within our modelling approach remains a topic for further research.

We finally discuss (albeit shortly) our empirical results in the light of option pricing and market microstructure theory. 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. 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 be expected, according to our modelling results, 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. These results are, of course, no surprise and completely in line with basic option pricing theory. This is not the case for the remaining time to maturity, which in our data example does not have any significant influence on the absolute option price changes at transaction level. 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. Further research with more data (of different securities) is needed to obtain reliable evidence for this conclusion. On the microstructural side 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 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 Verecchia (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. Furthermore, 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, $t = 1, ..., T$ and the explanatory variables $\mathbf{X}_t$, $t = 1, ..., T$ can be written in the following form:

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

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 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.

Durbin, J. and S. Koopman (2000). Time series of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society B 62*, 3–56.

Engle, R. and J. Russell (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Data. *Econometrica 66*, 1127–1162.

Engle, R. and J. Russell (2005). A Discrete-State Continuous-Time Model of Financial Transaction Prices and Times: The ACM-ACD Model. *Journal of Business and Economic Statistics 23*, 166–180.

Forsberg, L. and E. Ghysels (2004). Why Do Absolute Returns Predict Volatility So Well? *Working Paper*.

Gamerman, D. (1997). *Markov Chain Monte Carlo- Stochastic simulation for Bayesian inference*. London: Chapman & Hall.

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.

Heinen, A. (2003). Modelling Time Series Count Data: An Autoregressive Conditional Poisson Model. *CORE Discussion Paper 2003*.

Heinen, A. and E. Rengifo (2003). Multivariate Autoregressive Modelling of Time Series Count Data Using Copulas. *CORE Discussion Paper 2003*.

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.

	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
inter-action 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 2: 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 ($c_1 = 80$, $c_2 = 1$).

	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
inter-action 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 3: 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 ($c_1 = 80$, $c_2 = 1$).

	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
inter-action 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 4: 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 (3.8) 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 .

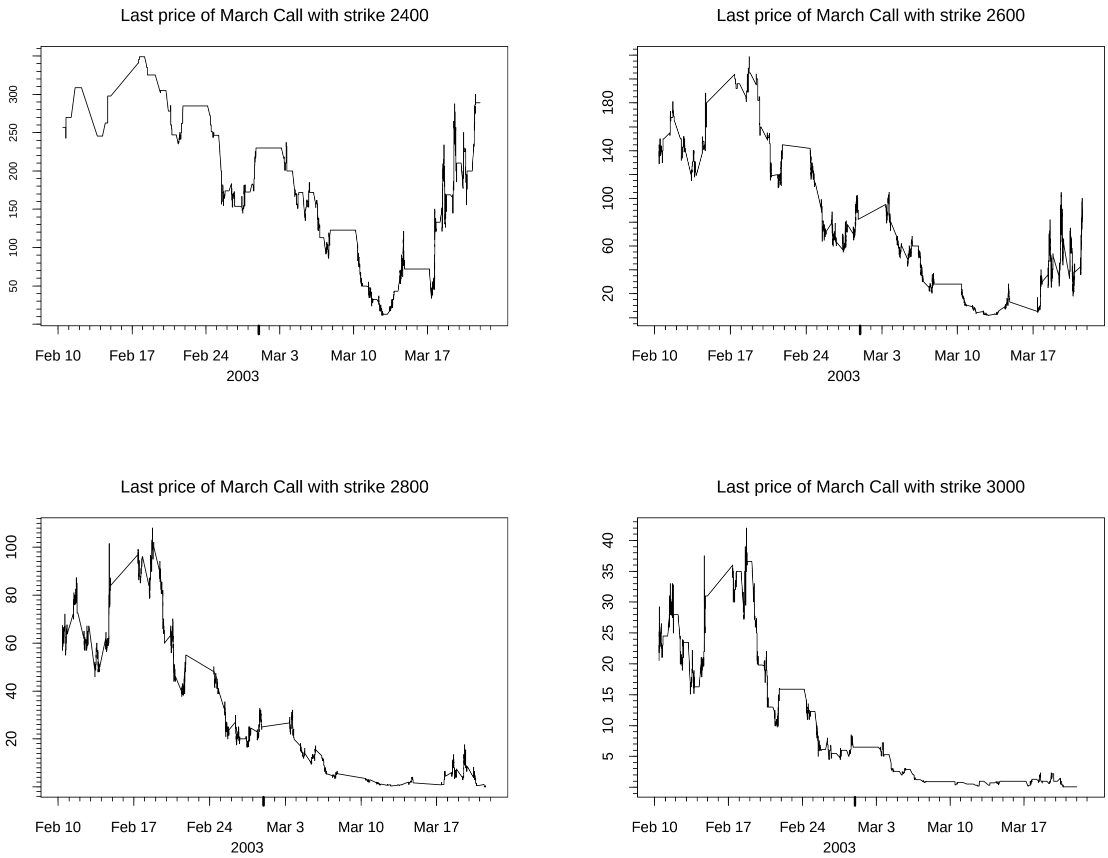

Figure 1: Last prices of some selected Call options on the XETRA DAX with different strike prices.

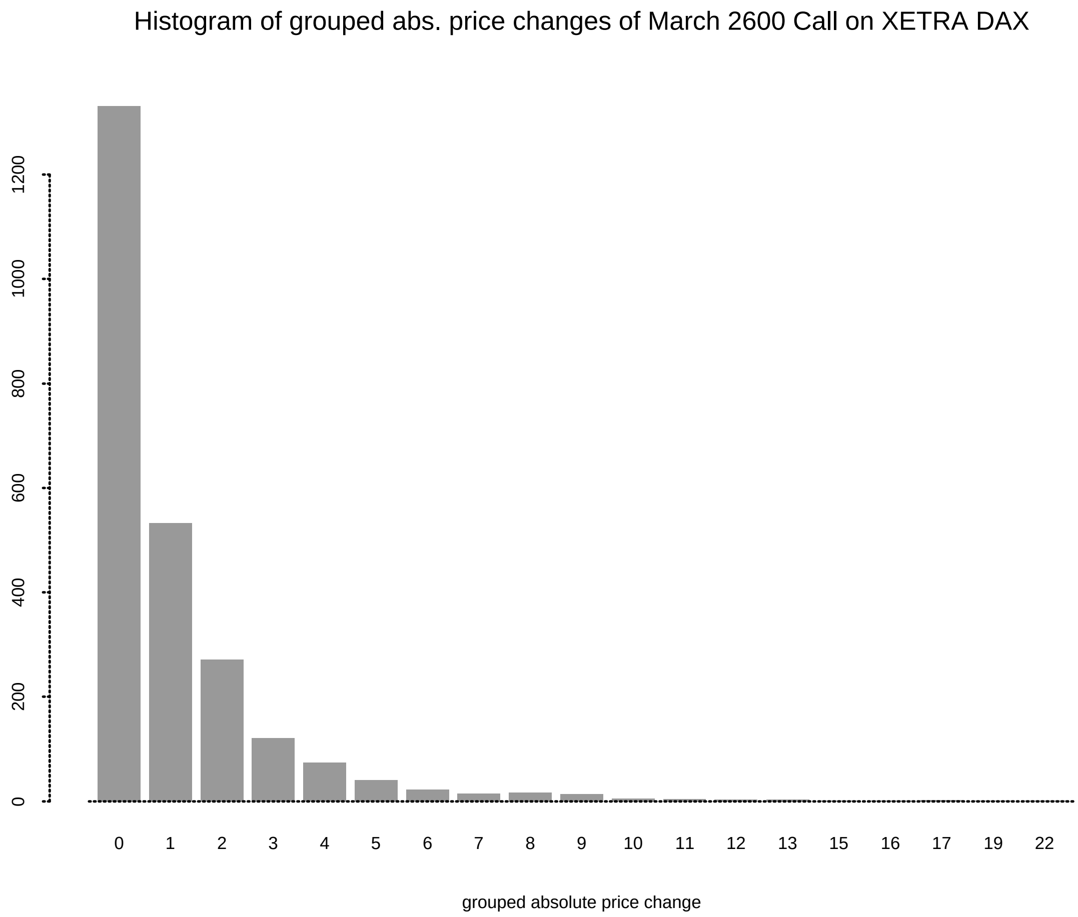

Figure 2: Histogram of grouped non-zero transaction price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003.

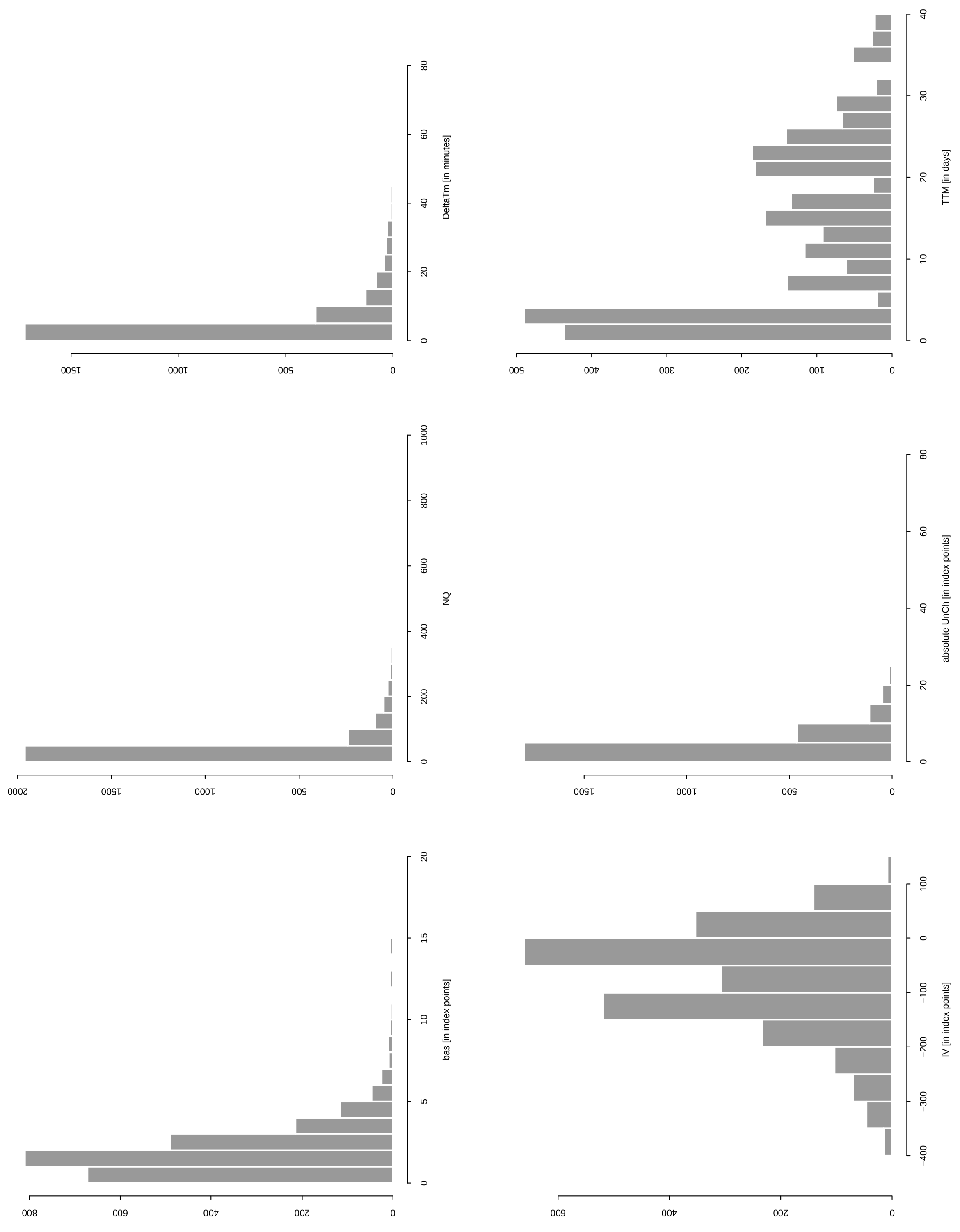

Figure 3: Histograms of fundamental and microstructural explanatory variables.

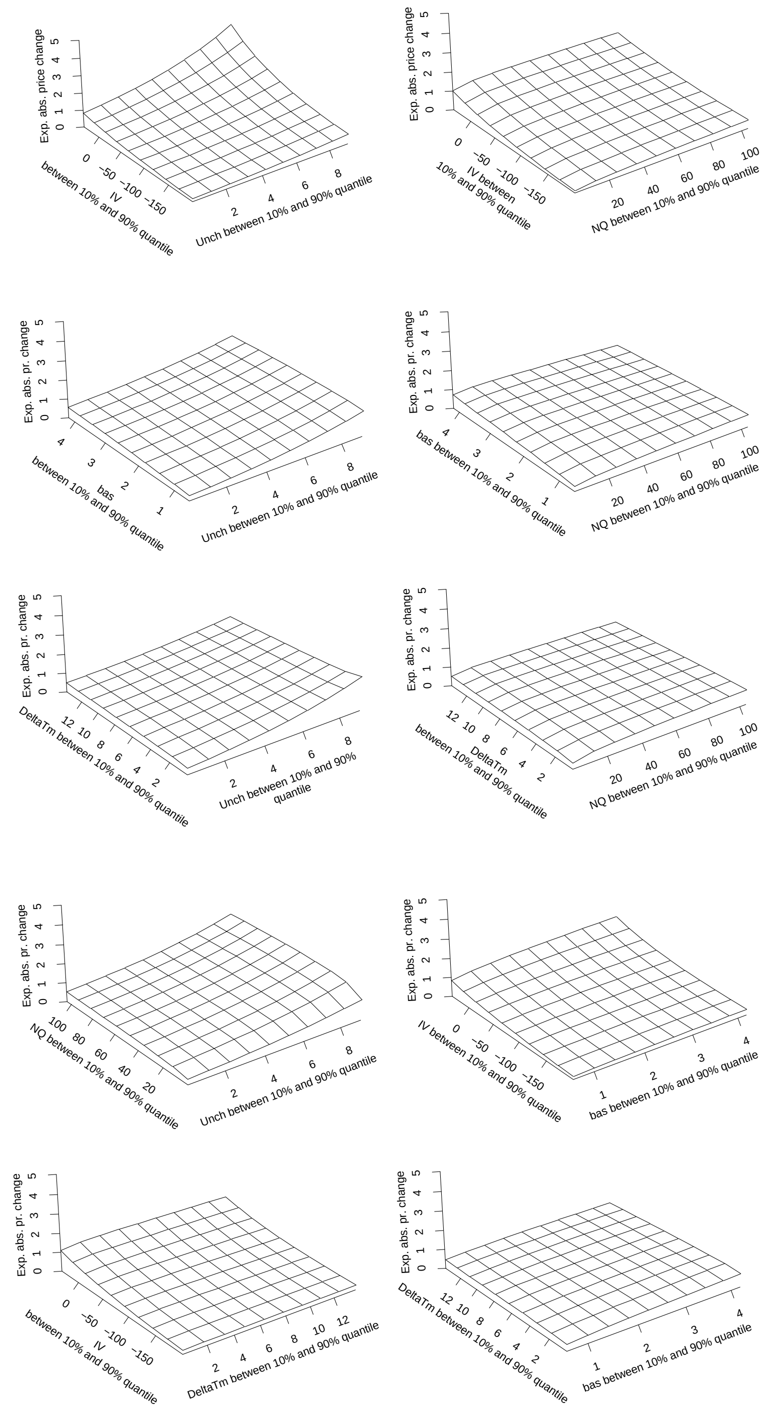

Figure 4: Fitted regression surfaces of the Poisson-GLM for the absolute price changes of the Call option on the XETRA DAX with strike price 2600 and expiration month March 2003 when two regressors vary and the remaining regressors are set equal to their median values.

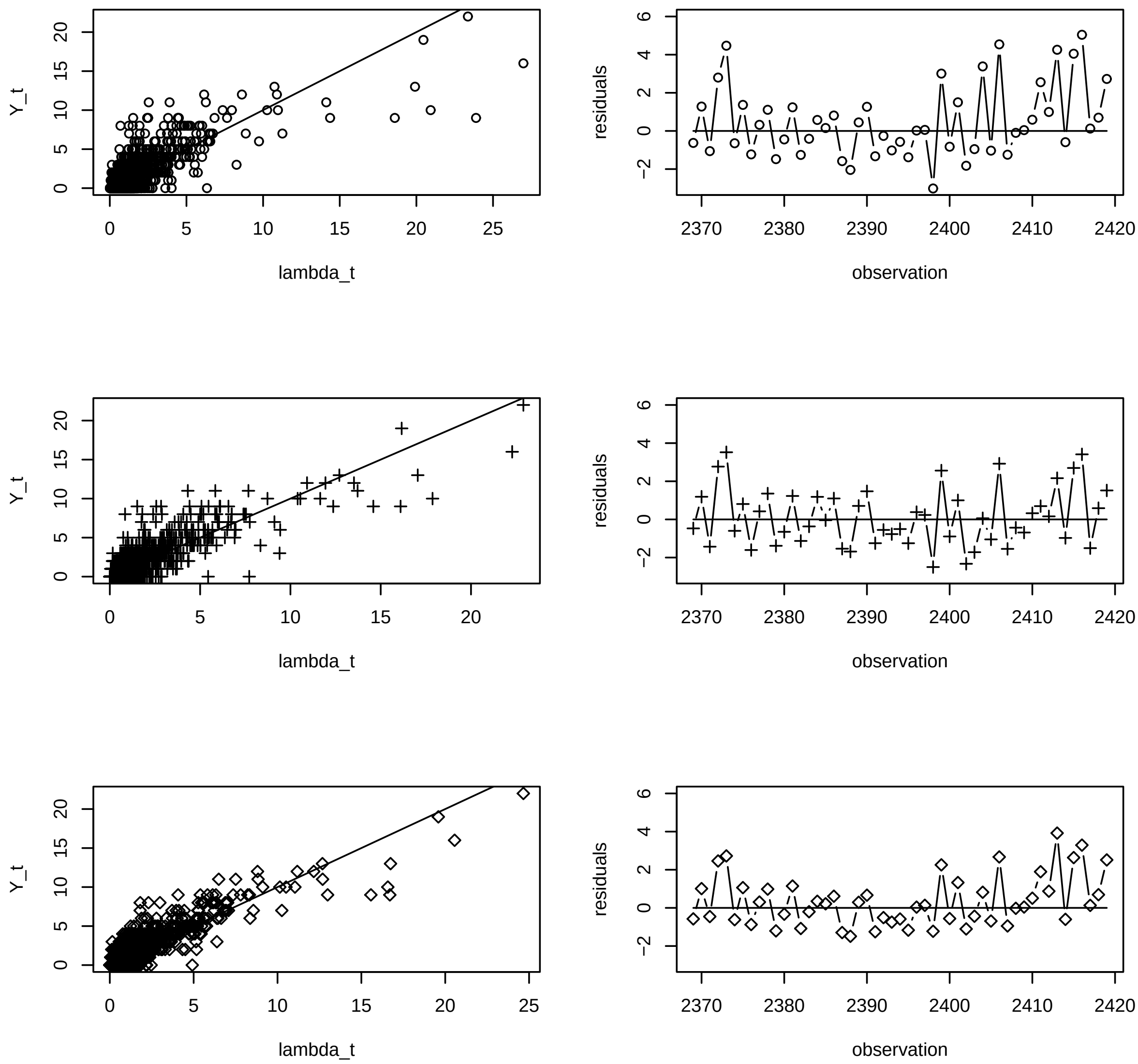

Figure 5: Observations Y_t plotted against $\hat{\lambda}_t = E(Y_t|\mathbf{x}_t)$ (with bisecting line) and residual time series plots (for the last 50 observations of the available sample) in the simple Poisson-GLM (top row), in the Poisson-GLM with an AR(1) latent process in the mean (centre row) and in the Poisson-GLM with a heterogeneous variance structure (bottom row).