Müller, Czado:

Regression Models for Ordinal Valued Time Series with Application to High Frequency Financial Data

Sonderforschungsbereich 386, Paper 301 (2002)

Online unter: http://epub.ub.uni-muenchen.de/

Regression Models for Ordinal Valued Time Series with Application to High Frequency Financial Data

Gernot Müller, Claudia Czado

Zentrum Mathematik
Technische Universität München
Boltzmannstraße 3
85747 Garching bei München
Germany

November 2002

Abstract

In financial time series transaction price changes often occur in discrete increments, for example in eights of a dollar. We consider these price changes as discrete random variables which are assumed to be generated by a latent process which incorporates both exogenous variables and autoregressive components. A standard Gibbs sampling algorithm has been developed to estimate the parameters of the model. However this algorithm exhibits bad convergence properties. To improve the standard Gibbs sampler we utilize methods proposed by Liu and Sabatti (2000, Biometrika 87), based on transformation groups on the sample space. A simulation study will be given to demonstrate the substantial improvement by this new algorithm. Finally we apply our model to the data of the IBM stock on Nov 13, 2000, and estimate the influence of the duration between transactions, the volume, and the bid-offer-spread both to model fit and prediction.

Keywords: Bayesian inference; Discrete-valued time series; High-frequency finance; Markov Chain Monte Carlo; Multigrid Monte Carlo.

1 Introduction

In recent years the availability of intraday financial data had a big impact on the econometrics literature. Many propositions were made to model the behavior of these high frequency data. The models deal either with the price process itself or with the time intervals (the durations) between market events. For a global overview about high frequency finance see Bauwens and Giot (2001) or Dacorogna, Gençay, Müller, Olsen, and Pictet (2001).

Recent developments in the section of the duration models are for example the ACD model of Engle and Russell (1998) and the many extensions such as the fractionally integrated ACD model of Jasiak (1998), the threshold ACD model of Zhang, Russell, and Tsay (2001), the SCD model of Bauwens and Veredas (1999), and the log-ACD model of Bauwens and Giot (2000). In modeling the price process itself one is faced with the feature that transaction price changes often occur in discrete increments. Hausman, Lo, and MacKinlay (1992) captured this feature in their ordered probit model, where they assumed the existence of a latent continuous dependent variable whose conditional mean is a linear function of observed explanatory variables. Another proposition was made by Rydberg and Shephard (2002). They suggest a decomposition model, where the price change is assumed to be a product of three random variables, namely of a price change indicator, of the direction and of the absolute value of the price change.

In this paper we introduce another model that deals with the price process itself and that captures the feature of discrete price changes. We assume that the observed price changes are discrete versions of latent continuous variables. Latent variable approaches were used for many different problems in literature, for example for the data augmentation algorithm in Tanner and Wong (1987) or for the computation of posterior distributions in probit models in Albert and Chib (1993). The latent process in our model has an autoregressive structure and is influenced by exogenous variables. The autoregressive structure is a major difference to the work of Hausman, Lo, and MacKinlay (1992), where the latent process is not autoregressive but only incorporates the lagged observed discrete price changes. Additionally, we try to find appropriate transformations for the exogenous variables whereas the cited work employs only observed covariates themselves.

For estimating the parameters we use a Bayesian approach and develop a standard Gibbs sampling algorithm in Section 2. However, this algorithm exhibits very bad convergence properties. Therefore in Section 3 we follow methods proposed by Liu and Sabatti (2000) to improve the behavior of the standard Gibbs sampler. In particular, we introduce a special transformation group acting on the sample space and

modify the algorithm adding a so-called GM-step (Liu and Sabatti (2000)). The simulation study in Section 4 demonstrates the substantial improvement of the modified Gibbs sampler. Finally, we apply our model to the data of the IBM stock on the 13th of November, 2000, and estimate the influence of some covariates to the price changes with regard to model fit and prediction.

2 Model and the standard Gibbs sampler

2.1 Model formulation

We assume that we can observe a discrete response time series $\{Y_t, t = 1, \ldots, T\}$, where Y_t takes on only K different values, and a $(p+1)$-dimensional vector $\mathbf{X}_t = (1, X_{t1}, \ldots, X_{tp})'$ of real-valued covariates for each $t \in \{1, \ldots, T\}$. To model the time dependency in $\{Y_t, t = 1, \ldots, T\}$ we assume that there exists an underlying unobserved real-valued time series $\{Y_t^*, t = 1, \ldots, T\}$ which produces the discrete valued Y_t by thresholding. In particular, the following latent variable representation holds:

$$Y_t = k \iff Y_t^* \in [\alpha_{k-1}, \alpha_k), \qquad k \in \{1, \ldots, K\}, \tag{2.1}$$

$$Y_t^* = \mathbf{X}_t'\boldsymbol{\beta} + \phi Y_{t-1}^* + \varepsilon_t^*, \qquad t \in \{1, \ldots, T\}, \tag{2.2}$$

where $-\infty = \alpha_0 < \alpha_1 < \ldots < \alpha_{K-1} < \alpha_K = \infty$ are unknown cutpoints, and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$ is a vector of unknown regression cofficients. All latent variables (except parameters) are marked with an asterisk *. We assume that $\varepsilon_t^* \sim N(0, \delta^2)$ i.i.d. Since the vector of covariates contains an intercept, we have to fix α_1 for reasons of identifiability. In particular, we set $\alpha_1 = 0$. For the same reasons we have to fix the variance δ^2, since otherwise we could multiply $\boldsymbol{\alpha} := (\alpha_2, \ldots, \alpha_{K-1})'$, $\boldsymbol{\beta}$ and $\mathbf{Y}^* := (Y_0^*, \ldots, Y_T^*)'$ by a positive constant without changing the likelihood. Therefore we assume $\delta^2 = 1$. It remains to estimate the latent variables $Y_t^*, t = 0, \ldots, T$, the cutpoints $\alpha_j, j = 2, \ldots, K-1$, the regression parameters $\beta_j, j = 0, \ldots, p$ and the autoregression parameter ϕ. For the following we introduce the notations $\boldsymbol{\theta} := (\beta_0, \ldots, \beta_p, \phi)'$, $\mathbf{Y} := (Y_1, \ldots, Y_T)'$, $\mathbf{Y}_{-t}^* := (Y_0^*, \ldots, Y_{t-1}^*, Y_{t+1}^*, \ldots, Y_T^*)'$ and $\boldsymbol{\alpha}_{-k} := (\alpha_2, \ldots, \alpha_{k-1}, \alpha_{k+1}, \ldots, \alpha_{K-1})'$. Univariate normal distributions that are truncated to an interval $[a, b]$ are denoted by $N_{[a,b]}(\mu, \sigma^2)$. For the n-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ we write $N_n(\boldsymbol{\mu}, \Sigma)$. I_n denotes the identity matrix with n rows and columns.

2.2 Bayesian inference using a standard Gibbs sampler

Now we develop a MCMC algorithm that allows us to draw approximate samples from the posterior distribution $[\boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{Y}^* \mid \mathbf{Y}]$. Here $[x|y]$ denotes the conditional distribution function or density of x given y, and $[x]$ the marginal distribution or density of x. For the Bayesian approach we have to specify prior distributions for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\mathbf{Y}^*$. We assume

$$[\mathbf{Y}^*, \boldsymbol{\theta}, \boldsymbol{\alpha}] \propto \exp\left\{-\frac{1}{2}\left[\sigma^{-2}(Y_0^*)^2 + \tau^{-2}\boldsymbol{\beta}'\boldsymbol{\beta} + \rho^{-2}\phi^2\right]\right\}, \tag{2.3}$$

where σ, τ and ρ are known hyperparameters. We assume all parameters to be a priori independent, we choose a noninformative prior for $\boldsymbol{\alpha}$ and normal priors for Y_0^*, $\boldsymbol{\beta}$ and ϕ, respectively. We can take large values for σ, τ and ρ, when there is little prior information about Y_0^* and θ.

Latent variable update:
First we have to determine the conditional distribution $\left[Y_0^* \mid \mathbf{Y}, \mathbf{Y}_{-0}^*, \boldsymbol{\alpha}, \boldsymbol{\theta}\right]$. Since Equation (2.1) holds, the complete vector $\mathbf{Y}$ is known, when $\boldsymbol{\alpha}$ and $\mathbf{Y}_{-0}^*$ are known. Therefore the conditional distribution reduces to $\left[Y_0^* \mid \mathbf{Y}_{-0}^*, \boldsymbol{\alpha}, \boldsymbol{\theta}\right]$. As one can see from Equation (2.2) only the component Y_1^* in the vector $\mathbf{Y}_{-0}^*$ can play a role for the conditional distribution. Together with the $N(0, \sigma^2)$-prior for Y_0^* it follows that

$$\left[Y_0^* \mid \mathbf{Y}, \mathbf{Y}_{-0}^*, \boldsymbol{\alpha}, \boldsymbol{\theta}\right] \sim N\left(\frac{\phi(Y_1^* - \mathbf{X}_1'\boldsymbol{\beta})}{\phi^2 + \sigma^{-2}}, \frac{1}{\phi^2 + \sigma^{-2}}\right).$$

Next we investigate the most complicated case, namely the conditional distribution $\left[Y_t^* \mid \mathbf{Y}, \mathbf{Y}_{-t}^*, \boldsymbol{\alpha}, \boldsymbol{\theta}\right]$ for $t \in \{1, \ldots, T-1\}$. By definition we have immediately

$$[Y_t^*|\mathbf{Y}, \mathbf{Y}_{-t}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}] = \frac{[\mathbf{Y}^*, \mathbf{Y}, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}]}{[\mathbf{Y}, \mathbf{Y}_{-t}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}]} \tag{2.4}$$

$$= \frac{[Y_T^*, Y_T|\mathbf{Y}_{-T}, \mathbf{Y}_{-T}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}][\mathbf{Y}_{-T}, \mathbf{Y}_{-T}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}]}{[\mathbf{Y}, \mathbf{Y}_{-t}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}]}. \tag{2.5}$$

The second factor in the numerator of expression (2.5) has the same form as the numerator on the right-hand side of Equation (2.4), therefore we can iterate this way of proceeding until we get

$$[Y_t^*|\mathbf{Y}, \mathbf{Y}_{-t}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}] = \frac{\prod_{n=1}^{T}[Y_n^*, Y_n|Y_0^*, \ldots, Y_{n-1}^*, Y_1, \ldots, Y_{n-1}, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}][Y_0^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}]}{[\mathbf{Y}, \mathbf{Y}_{-t}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}]}.$$

Since we are interested in a conditional distribution for Y_t^*, we can ignore terms that do not involve Y_t^* to get

$$[Y_t^*|\mathbf{Y}, \mathbf{Y}_{-t}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}] \propto \prod_{n=t}^{T}[Y_n^*, Y_n|Y_0^*, \ldots, Y_{n-1}^*, Y_1, \ldots, Y_{n-1}, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}]$$

Using the definition for conditional distributions each of these factors can be further decomposed as

$$\begin{aligned}
&[Y_n^*, Y_n | Y_0^*, \ldots, Y_{n-1}^*, Y_1, \ldots, Y_{n-1}, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}] \\
&\quad = [Y_n | Y_n^*, Y_0^*, \ldots, Y_{n-1}^*, Y_1, \ldots, Y_{n-1}, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}][Y_n^* | Y_0^*, \ldots, Y_{n-1}^*, Y_1, \ldots, Y_{n-1}, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}] \\
&\quad = [Y_n | Y_n^*, \boldsymbol{\alpha}][Y_n^* | Y_{n-1}^*, \boldsymbol{\beta}, \phi].
\end{aligned}$$

For the last equality, (2.1) and (2.2) are used. Now ignoring again all terms that do not involve Y_t^* we get

$$\begin{aligned}
&[Y_t^* | \mathbf{Y}, \mathbf{Y}_{-t}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}] \\
&\quad \propto [Y_t | Y_t^*, \boldsymbol{\alpha}][Y_{t+1}^* | Y_t^*, \boldsymbol{\beta}, \phi][Y_t^* | Y_{t-1}^*, \boldsymbol{\beta}, \phi] \\
&\quad \propto 1_{\left[\alpha_{Y_t-1}, \alpha_{Y_t}\right)}(Y_t^*) \cdot \exp\left\{-\frac{1}{2}(Y_{t+1}^* - \mathbf{X}_{t+1}'\boldsymbol{\beta} - \phi Y_t^*)^2\right\} \\
&\qquad\qquad \cdot \exp\left\{-\frac{1}{2}(Y_t^* - \mathbf{X}_t'\boldsymbol{\beta} - \phi Y_{t-1}^*)^2\right\} \\
&\quad \propto 1_{\left[\alpha_{Y_t-1}, \alpha_{Y_t}\right)}(Y_t^*) \cdot \\
&\qquad \exp\left\{-\frac{1}{2}\left[\left(1+\phi^2\right)(Y_t^*)^2 - 2\left(\phi(Y_{t+1}^* - \mathbf{X}_{t+1}'\boldsymbol{\beta}) + (\mathbf{X}_t'\boldsymbol{\beta} + \phi Y_{t-1}^*)\right) Y_t^*\right]\right\}.
\end{aligned}$$

Obviously this is a truncated normal distribution,

$$[Y_t^* | \mathbf{Y}, \mathbf{Y}_{-t}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}] \sim N_{[\alpha_{Y_t-1}, \alpha_{Y_t})}\left(\frac{\phi(Y_{t+1}^* - \mathbf{X}_{t+1}'\boldsymbol{\beta}) + (\mathbf{X}_t'\boldsymbol{\beta} + \phi Y_{t-1}^*)}{1+\phi^2}, \frac{1}{1+\phi^2}\right).$$

In a similar way as the full conditional for Y_0^* we get

$$[Y_T^* | \mathbf{Y}, \mathbf{Y}_{-T}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}] \sim N_{\left[\alpha_{Y_T-1}, \alpha_{Y_T}\right)}\left(\mathbf{X}_T'\boldsymbol{\beta} + \phi Y_{T-1}^*, 1\right).$$

Joint regression and autoregressive parameter update:
Here we can update all the parameters β_j, $j = 0, \ldots, p$, and ϕ in one block. The derivation of the full conditional is completely analogous to the well-known probit model (e.g. Liu and Sabatti (2000), p.364). In particular we have for $\boldsymbol{\theta} = (\beta_0, \ldots, \beta_p, \phi)'$

$$[\boldsymbol{\theta} | \mathbf{Y}, \mathbf{Y}^*, \boldsymbol{\alpha}] \sim N_{p+2}(\Sigma Z' \mathbf{Y}_{-0}^*, \Sigma),$$

a $(p+2)$-dimensional normal distribution with covariance matrix $\Sigma = \left(Z'Z + \tau^{-1} I_{p+2}\right)^{-1}$ and

$$Z = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} & Y_0^* \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{T1} & \cdots & x_{Tp} & Y_{T-1}^* \end{bmatrix}.$$

Cutpoint parameter update:
We are now interested in the full conditionals $[\alpha_k | \mathbf{Y}, \mathbf{Y}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}_{-k}]$ for $k \in \{2, \ldots, K-$

1}. As one can see from Equation (2.1), α_k is the cutpoint that separates category k and category $k+1$. Additionally, to fulfill the order condition, α_k has to lie between α_{k-1} and α_{k+1}. Since we have no other information about α_k and have assumed a non-informative prior for it, it follows that $[\alpha_k|\mathbf{Y}, \mathbf{Y}^*, \boldsymbol{\beta}, \phi, \boldsymbol{\alpha}_{-k}]$ is uniformly distributed in the interval (l_k, r_k), where

$$l_k = \max\left\{\alpha_{k-1}, \max_{t=1,\ldots,T}\{Y_t^*|Y_t = k\}\right\} \tag{2.6}$$

$$r_k = \min\left\{\alpha_{k+1}, \min_{t=1,\ldots,T}\{Y_t^*|Y_t = k+1\}\right\}. \tag{2.7}$$

3 A Grouped Move Multigrid Monte Carlo (GM-MGMC) algorithm

Simulation experiments with the standard Gibbs sampler developed in Section 2.2 show that the produced MCMC-chains converge very slowly to the region around the true value especially for the cutpoints α_j and the regression intercept β_0. This behavior was also observed by Cowles (1996) for the independent multinomial case, which can be explained as follows. The parameter α_j is drawn from a uniform distribution with boundaries l_j and r_j given in Equations (2.6) and (2.7). If the observed dataset is large (e.g. $T = 2000$), the difference $r_j - l_j$ is very small, so α_j has very little room to move in one iteration. Therefore we look for some possibilities to speed up the convergence of the standard Gibbs sampler.

One general possibility is, of course, to update some variables in one block, for example all latent variables Y_t^*. In this case we have to draw a sample of a $(T+1)$-dimensional truncated normal distribution instead of drawing $T+1$ samples from univariate normal distributions. However, to get a sample from a multivariate truncated normal distribution one has to employ a Gibbs sampler itself (cf. Geweke (1991) or Robert (1995)). Simulations show that one reaches an improvement in convergence, but the computational cost is very high. So if one uses this method fewer iterations are needed for a comparable result, but the time used for each iteration increases in such a way that the overall improvement is negligible. In addition this blocking of the latent variables does not involve the update of the cutpoints α_k which seems to be most important.

Therefore, we use now a method that was proposed by Liu and Sabatti (2000). It is based on a method to sample from a distribution using group transformations. In particular, if Γ is a locally compact group of transformations defined on the sample

space $\mathbf{S}$, L its left-Haar measure (as defined in Rao (1987), p. 492), $\mathbf{x} \in \mathbf{S}$ follows a distribution with density π, and $\gamma \in \Gamma$ is drawn from $\pi(\gamma(\mathbf{x}))|J_\gamma(\mathbf{x})|L(d\gamma)$, with $J_\gamma(\mathbf{x}) = \det\left(\frac{\partial\gamma(\mathbf{x})}{\partial\mathbf{x}}\right)$, $\frac{\partial\gamma(\mathbf{x})}{\partial\mathbf{x}}$ the Jacobian matrix, then $\mathbf{x}^* = \gamma(\mathbf{x})$ has density π, too (Liu and Sabatti (2000), Theorem 1). Typical examples of such transformation groups are the *translation group on* $\mathbf{S}$ *along an arbitrary direction*, $\Gamma = \{\gamma \in \mathbb{R}^1 : \gamma(\mathbf{x}) = \mathbf{x} + \gamma\mathbf{e} = (x_1 + \gamma e_1, \ldots, x_d + \gamma e_d)\}$ with Lebesque measure as left-Haar measure (here γ has to be drawn from $\pi(\mathbf{x} + \gamma\mathbf{e})$), or the *scale group on* $\mathbf{S}$, $\Gamma = \{\gamma > 0 : \gamma(\mathbf{x}) = (\gamma x_1, \ldots, \gamma x_d)\}$ with $\gamma^{-1}d\gamma$ as left-Haar measure (here γ has to be drawn from $\gamma^{d-1}\pi(\gamma\mathbf{x})$). In both cases the form of the left-Haar measure follows directly from its definition.
We apply the method by Liu and Sabatti (2000) where $\mathbf{x}$ is a vector with $T+p+K+1$ components, namely

$$\mathbf{x} = (Y_0^*, \ldots, Y_T^*, \beta_0, \ldots, \beta_p, \alpha_2, \ldots, \alpha_{K-1}, \phi)\,,$$

and π is the posterior density of $\mathbf{x}$. The difficulty in the choice of a suitable transformation group is to find one where the resulting distribution allows to draw samples very fast. Unfortunately, in our problem standard transformation groups as the translation group or the scale group do not lead to an easy sampling distribution. Therefore we use the group

$$\Gamma_m := \{\gamma > 0 : \gamma(\mathbf{x}) = (\gamma x_1, \ldots, \gamma x_m, x_{m+1}, \ldots, x_d)\}\,,$$

which we call a *partial scale group on* $\mathbf{S}$. Here only m components are transformed, the others remain fixed. The left-Haar measure for this group is again $\gamma^{-1}d\gamma$ as for the (total) scale group. We easily compute $\det\left(\frac{\partial\gamma(\mathbf{x})}{\partial\mathbf{x}}\right) = \gamma^m$. Therefore

$$\pi(\gamma(\mathbf{x}))|J_\gamma(\mathbf{x})|L(d\gamma) = \gamma^{m-1}\pi(\gamma\mathbf{x})\,d\gamma.$$

In order to get an easy sampling distribution for our problem we take $m = T + p + K$ and let only the parameter ϕ remain fixed. Therefore

$$\gamma(\mathbf{x}) = (\gamma Y_0^*, \ldots, \gamma Y_T^*, \gamma\beta_0, \ldots, \gamma\beta_p, \gamma\alpha_2, \ldots, \gamma\alpha_{K-1}, \phi)\,.$$

The posterior distribution in our problem is given by

$$\begin{aligned}
\pi(\mathbf{x}) &= [Y_0^*, \ldots, Y_T^*, \boldsymbol{\beta}, \boldsymbol{\alpha}, \phi | Y_1, \ldots, Y_T] \\
&\propto \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T}(Y_t^* - \boldsymbol{\beta}'\mathbf{X}_t - \phi Y_{t-1}^*)^2 + \sigma^{-2}(Y_0^*)^2 + \tau^{-2}\boldsymbol{\beta}'\boldsymbol{\beta} + \rho^{-2}\phi^2\right]\right\} \cdot \\
&\quad \cdot \prod_{t=1}^{T} 1_{[\alpha_{Y_t-1}, \alpha_{Y_t})}(Y_t^*)
\end{aligned}$$

The density $\gamma^{m-1}\pi(\gamma\mathbf{x})$ is therefore proportional to

$$\begin{aligned}
&\gamma^{m-1}\exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T}(\gamma Y_t^* - \gamma\boldsymbol{\beta}'\mathbf{X}_t - \phi\gamma Y_{t-1}^*)^2 + \sigma^{-2}(\gamma Y_0^*)^2 + \tau^{-2}\gamma^2\boldsymbol{\beta}'\boldsymbol{\beta} + \rho^{-2}\phi^2\right]\right\}\cdot \\
&\quad\cdot\prod_{t=1}^{T}1_{[\gamma\alpha_{Y_t-1},\gamma\alpha_{Y_t})}(\gamma Y_t^*) \\
\propto\ &\gamma^{m-1}\exp\left\{-\frac{1}{2}\gamma^2\left[\sum_{t=1}^{T}(Y_t^* - \boldsymbol{\beta}'\mathbf{X}_t - \phi Y_{t-1}^*)^2 + \sigma^{-2}(Y_0^*)^2 + \tau^{-2}\boldsymbol{\beta}'\boldsymbol{\beta}\right]\right\}\cdot \\
&\quad\cdot\prod_{t=1}^{T}1_{[\alpha_{Y_t-1},\alpha_{Y_t})}(Y_t^*) \\
\propto\ &\left(\gamma^2\right)^{\frac{m-1}{2}}\exp\left\{-\frac{1}{2}\gamma^2\left[\sum_{t=1}^{T}(Y_t^* - \boldsymbol{\beta}'\mathbf{X}_t - \phi Y_{t-1}^*)^2 + \sigma^{-2}(Y_0^*)^2 + \tau^{-2}\boldsymbol{\beta}'\boldsymbol{\beta}\right]\right\}
\end{aligned}$$

which is proportional to a Gamma distribution $\Gamma(a,b)$ for γ^2 with parameters

$$a = \frac{T+K+p+1}{2} \tag{3.1}$$

$$b = \frac{\sum_{t=1}^{T}(Y_t^* - \mathbf{X}_t'\boldsymbol{\beta} - \phi Y_{t-1}^*)^2 + \sigma^{-2}(Y_0^*)^2 + \tau^{-2}\boldsymbol{\beta}'\boldsymbol{\beta}}{2}. \tag{3.2}$$

Here the $\Gamma(a,b)$ density is given by

$$f_{\Gamma(a,b)}(x) = \frac{b^a}{\Gamma(a)}x^{a-1}e^{-bx},\ \ x \geq 0.$$

In this way we get a new algorithm that lies in the class of the grouped move multigrid Monte Carlo (GM-MGMC) algorithms (Liu and Sabatti (2000)). Each iteration consists of the following two parts:

(a) **MCMC-Step**: Generate an iteration from the standard Gibbs-sampler using
- latent variable update
- joint regression and autoregressive parameter update
- cutpoint parameter update

to get $\mathbf{Y}_c^*, \boldsymbol{\beta}_c, \phi_c, \boldsymbol{\alpha}_c$ as current values.

(b) **GM-Step**: Draw γ^2 from $\Gamma(a,b)$ with a and b defined in (3.1) and (3.2) respectively, and update the current values by multiplication with the group element $\gamma = \sqrt{\gamma^2}$,

$$\begin{aligned}
\mathbf{Y}_c^* &\leftarrow \gamma\mathbf{Y}_c^* \\
\boldsymbol{\beta}_c &\leftarrow \gamma\boldsymbol{\beta}_c \\
\boldsymbol{\alpha}_c &\leftarrow \gamma\boldsymbol{\alpha}_c
\end{aligned}$$

Note that ϕ_c does not need to be updated since it remains unchanged under the partial scale group. Chen, Shao, and Ibrahim (2000) point out that GM-MGMC algorithms do not always guarantee faster convergence than its parent MCMC-algorithm. Therefore in the following section we test the new algorithm in several situations.

4 Simulation study

4.1 Design of the simulation study

First we mention that the smaller the number K of categories is the better the parameter estimates are when the Gibbs sampler of Section 3 is used. Therefore we investigate the behavior of our improved Gibbs sampler for data sets with a for our practical concerns relative high number of categories, $K = 7$. Further we take $T = 2000$.

We investigate the behavior of the GM-MGMC Gibbs sampler for six different parameter sets that differ in the sign of the autoregressive parameter ϕ, in the frequencies that occur in the categories, and in the covariates. See Table 1 for specific choices made. Especially concerning the frequencies we are interested in whether the parameter estimates are better if the frequencies in the categories are nearly identical than in situations where the majority of the observations lies in one or two categories and only few observations in the other categories. This is important for many practical problems. In particular for our application in Section 5 we observe that more than 80% of the observations lie in the categories 1 and 2, and only 3.5% in category 4. In the following, 'very different' frequencies means that there is at least one category with more than 40% of the observations and at least one category with less than 5% of the observations. 'Nearly identical' means that in each category lie between 11% and 18% of the observations.

Parameter set	ϕ	freq. in categories	covariates
A	negative	nearly identical	2 (normally distributed)
B	positive	nearly identical	2 (normally distributed)
C	negative	very different	2 (normally distributed)
D	positive	very different	2 (normally distributed)
E	negative	nearly identical	2 (exp.trend,Bernoulli(0.6))
F	negative	nearly identical	4 (normally distributed)

Table 1: Parameter settings for the simulation study.

	$X_{1,t}$		$X_{2,t}$		$X_{3,t}$		$X_{4,t}$	
Parameter set	mean	stdd.	mean	stdd.	mean	stdd.	mean	stdd.
A	-0.40	2.20	0.02	0.02				
B	-1.00	1.00	-0.25	0.18				
C	0.00	0.10	0.30	0.90				
D	-1.00	0.90	-0.25	0.11				
F	0.00	0.80	1.60	1.10	0.40	0.60	-0.50	0.40

Table 2: Means and standard deviations of the normal distributions to generate the covariates.

The parameter sets A to D are connected in the sense that in each we use two covariates generated from normal distributions with specified mean and variance values, and that there appear all combinations of positive/negative ϕ and nearly identical/very different frequencies. In parameter set E, covariate $X_{1,t}$ is an exponential trend with $X_{1,t} = e^{t/T} = e^{t/2000}$ and $X_{2,t},\ t = 1, \ldots, 2000$ is drawn from a Bernoulli-distribution with success probability 0.6. In parameter set F we use four covariates generated from normal distributions.

For each parameter set we first chose the value of the autoregressive parameter ϕ. Then we tried not only different cutpoints to get nearly identical or very different frequencies in the categories, but also different standard deviations of the normal distributions the covariates were generated from. This is because otherwise we would have had very small or very high values for the cutpoint $\alpha_{K-1} = \alpha_6$, which would mean that the influence of the noise would have been very large or very small. The means and standard deviations of the normal distributions used in parameter sets A, B, C, D, and F are given in Table 2. Figure 1 shows the densities of $\beta_j X_{tj},\ j = 1, 2$, the linear predictor $\mathbf{X}_t'\boldsymbol{\beta}$, and the error component ε_t^* for parameter sets A to D. In the fourth and fifth column one can compare the influence of $\mathbf{X}_t'\boldsymbol{\beta}$ and of the noise ε_t^*. The specific settings of the cutpoints, regression coefficients, and autoregressive parameter are given in Table 3.

The simulation consists of the following three steps:

1. Generation of one design matrix per parameter set.
2. Simulation of 100 data sets per parameter set using the design of step 1.
3. 15000 iterations of the GM-MGMC Gibbs sampler for each of the data sets.

Parameter set	α_2	α_3	α_4	α_5	α_6	β_0	β_1	β_2	β_3	β_4	ϕ
A	1.3	2.4	3.5	4.6	5.7	4.9	0.8	-60.0			-0.2
B	1.2	2.2	3.1	4.1	5.3	2.9	-0.6	9.0			0.5
C	0.7	1.7	3.0	4.0	4.7	2.9	7.9	0.6			-0.3
D	0.8	1.9	3.5	4.7	5.5	3.3	-0.6	9.0			0.4
E	0.6	1.3	2.4	3.1	3.7	2.7	0.2	-1.1			-0.3
F	0.7	1.8	3.5	4.6	5.3	2.8	-0.8	0.4	0.5	0.4	-0.3

Table 3: Settings of the cutpoints, regression coefficients, and autoregressive parameter.

4.2 Results of the simulation study

The computing time is about 0.08 seconds per iteration in presence of two covariates and about 0.12 seconds in presence of four covariates on an 850 MHz Pentium III processor, so that we had an over-all computing time of about 220 hours. As start values we took $\beta_j = 0,\ j = 0,\ldots,p,\ \phi = 0,\ (\alpha_2, \alpha_3, \alpha_4, \ldots) = (2, 4, 6, \ldots)$, and randomly drawn values uniformly distributed between α_{Y_t-1} and α_{Y_t} for $Y_t^*,\ t = 1,\ldots,2000$. After running the GM-MGMC Gibbs sampler we estimated the parameters by the posterior means using the iterations 3001 to 15000, and computed estimates of the relative bias, the standard deviation for the relative bias, the relative MSE, and the standard deviation for the relative MSE for each parameter in each parameter set. The estimates of the relative bias ($\widehat{\mathrm{B}}_{\mathrm{rel}}$), the relative MSE ($\widehat{\mathrm{MSE}}_{\mathrm{rel}}$), and of the corresponding standard deviations are

$$
\begin{aligned}
\widehat{\mathrm{B}}_{\mathrm{rel}} &= \frac{1}{\psi}\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\psi}_r - \psi\right), \\
\widehat{\mathrm{MSE}}_{\mathrm{rel}} &= \frac{1}{\psi^2}\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\psi}_r - \psi\right)^2, \\
\widehat{\mathrm{std}}(\widehat{\mathrm{B}}_{\mathrm{rel}}) &= \frac{1}{\psi}\sqrt{\frac{1}{R(R-1)}\sum_{r=1}^{R}\left[\hat{\psi}_r - \psi - \psi\widehat{\mathrm{B}}_{\mathrm{rel}}\right]^2}, \\
\widehat{\mathrm{std}}(\widehat{\mathrm{MSE}}_{\mathrm{rel}}) &= \frac{1}{\psi^2}\sqrt{\frac{1}{R(R-1)}\sum_{r=1}^{R}\left[\left(\hat{\psi}_r - \psi\right)^2 - \psi^2\widehat{\mathrm{MSE}}_{\mathrm{rel}}\right]^2},
\end{aligned}
$$

where ψ is the parameter to estimate and $\hat{\psi}_r$ the posterior mean estimate of ψ for the r^{th} data set based on 12000 iterations of the GM-MGMC algorithm. Here we used $\widehat{\mathrm{B}}_{\mathrm{rel}} = \psi^{-1}\hat{B}$, where $\hat{B}$ is the common bias estimate, and the independence of the estimated posterior means for different data sets.

Parameter set	rel. bias	stdd. of rel. bias	rel. MSE	stdd. of rel. MSE
A	-0.0274	0.002932	0.0016	0.000132
B	0.0027	0.002513	0.0006	0.000078
C	0.0063	0.004196	0.0018	0.000206
D	0.0005	0.000048	0.0012	0.000119
E	0.0005	0.000049	0.0016	0.000165
F	0.0144	0.005020	0.0027	0.000393

Table 4: Estimates of the relative bias, relative MSE, and their standard deviations for the autoregressive parameter ϕ.

The autoregressive parameter ϕ is always estimated quite well, as one can see in Table 4. Figures 2 and 3 show the results for the cutpoints and the regression parameters. The relative bias is between -1% and +1% for the cutpoints as well as for the regression parameters. The relative MSE is less than 0.001 for most of the parameters in the parameter sets A to D. That means that on average the estimates are less than 3% away from the true values. The relative MSE is worse for the parameter sets E and F. The worst value we get for β_1 in set E, what was the regression parameter for the exponential trend. Further we like to mention that all these estimates are similar if one uses the iterations 5001 to 15000 instead of iterations 3001 to 15000.
Finally in Figure 4 we have a look at the autocorrelations of the chains produced by the Gibbs sampler with GM-step and the Gibbs sampler without GM-step for a specific data set from parameter set A. However, a similar behavior of the autocorrelations was also observed for the other parameter sets (not shown). For the cutpoints the chains produced by the the standard Gibbs sampler show very slowly decreasing autocorrelations, even for lag 100 they are mostly above 0.8. In contrast, the GM-MGMC Gibbs sampler produces chains with much lower autocorrelations for the cutpoints. The chains for the autoregressive parameter ϕ and the regression parameters $\boldsymbol{\beta}$ show the same pattern of autocorrelations in both algorithms.
We conclude that the GM-MGMC Gibbs sampler works very well in most situations, especially there is no difference in the performance whether the autoregressive parameter is positive or negative and whether there are categories with very different frequencies or not. As expected, the fewer covariates we use, the better is the estimation. The autocorrelations in the chains produced by the GM-MGMC Gibbs sampler are explicitly better than those in the chains produced by the standard Gibbs sampler.

5 Application to financial data

5.1 General remarks

In this chapter we want to detect and to quantify the influence of covariates information on price changes of stocks. For this we took data of the IBM stock traded on Nov 13, 2000 at the NYSE. This is a very frequently traded stock (about 450 transactions per hour), so that we have enough data even if we do not use data from the opening and closing period which might exhibit different behavior. The data is taken from the TAQ2 database of the NYSE, which contains the following covariates:

- TIMEDIFF, the time elapsed between two following transactions in seconds.
- SIZE, the volume of the transaction.
- BOS, the last available bid-offer-spread in US$.

As response Y_t we take the absolute values of the price differences. The price differences take on only values which are integer multiples of $\frac{1}{16}$ US$, and 99.5% of them lie between $-\frac{3}{16}$ US$ and $+\frac{3}{16}$ US$. The absolute price differences are a reasonable quantity to consider, since first exploratory analyses show that longer time differences between two transactions generally tend to lead to higher price jumps, however partly upwards and partly downwards: As can be seen from Figure 5, price changes of $\frac{3}{16}$ US$ or more upwards or downwards do not occur for time differences of less than 4 seconds. The frequency of higher price changes, upwards and downwards, increases with the time elapsed since the last transaction. Therefore taking the signed values would nullify the effect of the time difference. For comparison all computations will also be made without covariates, that means only with the intercept and the AR(1) component.

First we want to conduct an exploratory analysis to choose appropriate transformations of the covariates. However, two problems arise here. The first problem arises since the response variable is discrete and takes on only few values. Therefore ordinary scatter plots are not informative, especially when the regressor is also discrete or categorial. Instead, we group the covariate data in intervals of the same length or use categories and then compute the average response per interval or per category. Now we can look for a linear (quadratic, logarithmic ...) relationship.

This relationship, however, is between Y_t in Equation (2.1) and X_t in Equation (2.2). Therefore linearity can be destroyed by the (a priori unknown) cutpoints α_k, which is the second problem. Only when the cutpoints are estimated to be nearly equidistant, we have a validation for the chosen transformation of the covariate. Otherwise one should use other transformations that take into account the different distances

price change	category	frequency
0$	1	855
$\pm 1/16$$	2	781
$\pm 2/16$$	3	293
$\leq -3/16$$ and $\geq 3/16$$	4	71

Table 5: Absolute price changes: associated categories and observed frequencies.

between the cutpoints. At this point further research will be necessary to develop iterative methods for choosing appropriate transformations.

5.2 Exploratory analysis

On the Nov 13, 2000, between 10:00:01 am and 2:29:45 pm there was a total of 2001 transactions of the IBM stock at the NYSE. Because we want to consider only the absolute values of the price changes from one transaction to the next one, we associate the signed price changes to the response categories as shown in Table 5.

EXPLORATORY ANALYSIS FOR TIMEDIFF
Considering the data it seems to be useful to take the logarithm of TIMEDIFF to get a nearly linear dependency of the response. As described in Section 5.1, we group the data in intervals of the same length and compute the average response for each interval. The result can be seen in Figure 6. The relationship is quite linear. For small (logarithmic) time differences we have an average response of 1.35, for big ones an average response of 1.96. The relatively high difference of $1.96 - 1.35 = 0.61$ is a first hint at the significance of this covariate. Therefore we use scored values of ln(TIMEDIFF) as covariate. Values greater than 2.6 were set equal to 2.6.

EXPLORATORY ANALYSIS FOR SIZE
Here we use categories to achieve a linear relationship. A SIZE of less than 500 stocks corresponds to category 0, 500 to 1000 stocks are associated with category 1, and 1000 or more with category 2. A plot of the average response per category is given in Figure 7. Obviously, the dependency is linear, however the difference between the maximal and minimal average response is only $1.84 - 1.69 = 0.15$. So this categorized version of SIZE is probably not a significant covariate. Estimations of the corresponding regression coefficient in a model with ln(TIMEDIFF) and category of SIZE as covariates affirm this result: A 95% posterior credible interval contains zero. Using other classifications for the categories do not lead to smooth dependencies

	Model A			Model B		
	estimate	std.err.	90% cred.int.	estimate	std.err.	90% cred.int.
α_2	1.141	0.034	(1.085,1.197)	1.091	0.080	(0.961,1.225)
α_3	2.108	0.059	(2.014,2.208)	2.004	0.113	(1.822,2.195)
β_0	-0.936	0.110	(-1.117,-0.755)	0.175	0.044	(0.097,0.247)
β_1	2.617	0.409	(1.944,3.283)			
β_2	0.340	0.040	(0.274,0.407)			
ϕ	0.034	0.030	(-0.015,0.083)	0.045	0.040	(-0.021,0.110)

Table 6: Estimated posterior means and corresponding estimated standard deviations and 90% posterior credible intervals for parameters in Models A and B.

or have the disadvantage of big unbalances in the frequencies for different categories. Therefore we do not use the covariate SIZE (or any transformation) in the following analysis.

EXPLORATORY ANALYSIS FOR BOS
The bid and offer prices provided by the TAQ2 database have different time stamps than the transactions. We use the last available bid and offer prices for computing the bid-offer-spread. This covariate takes on values from US$ 0.0625 to US$ 0.75 in steps of US$ 0.0625. However, bid-offer-spreads greater than US$ 0.3125 occur only 19 times in the 2000 observations. Therefore spreads of greater than US$ 0.3125 were set equal to US$ 0.3125. The average responses for spreads between US$ 0.0625 and US$ 0.3125 can be found in Figure 8. The relationship is quite linear, and the difference between the maximal and minimal average response is $2.04 - 1.53 = 0.51$. So we expect the covariate BOS to be significant.

5.3 Model estimation and verification

MODEL ESTIMATION
All computations were done for two models specified by:

$$\text{Model A:} \quad Y_t^* = \beta_0 + \beta_1 \cdot \text{BOS}_t + \beta_2 \cdot \ln(\text{TIMEDIFF})_t + \phi Y_{t-1}^*$$
$$\text{Model B:} \quad Y_t^* = \beta_0 + \phi Y_{t-1}^*$$

The estimated posterior means for the cutpoints, the regression coefficients, and the autoregressive parameter using each of the iterations 5001 to 15000 of the GM-MGMC Gibbs sampler are given in Table 6 together with their corresponding estimated standard deviations and 90% credible intervals. In Figure 9 the estimated marginal densities for the parameters in model A are shown, in Figure 10 the estimated marginal

densities for the parameters in Model B. We conclude that the intercept, the bid-offer-spread, and ln(TIMEDIFF) are highly significant, whereas the autoregressive component with estimated parameter 0.034 is less significant. This low value means that the response is mainly dependent of the covariates. Because of the positive sign of the estimates for β_1 and β_2, and because of the estimate for β_0, which lies clearly left of the smallest cutpoint $\alpha_1 = 0.0$, we expect no price change, if the bid-offer-spread is small and only few seconds elapse until the next transaction. The higher the bid-offer-spread is and the more time elapses, the higher is our expected price change.

MODEL VERIFICATION

We want to compare model A with model B. Running the GM-MGMC Gibbs sampler we get parameter estimates $\hat{\boldsymbol{\alpha}}_i$, $\hat{\boldsymbol{\beta}}_i$, $\hat{\phi}_i$ and $\hat{Y}^*_{t-1,i}$ for each iteration i. For each $i \in \{5001, \ldots, 15000\}$ and each $t \in \{1, \ldots, 2000\}$ we can then compute the following estimated probabilities:

$$\hat{\mathrm{P}}(Y_t = 1 \mid \hat{\boldsymbol{\beta}}_i, \hat{\phi}_i, \hat{\boldsymbol{\alpha}}_i, \hat{Y}^*_{t-1,i}) := \Phi(\hat{\alpha}_{1,i} - \mathbf{X}'_t\hat{\boldsymbol{\beta}}_i - \hat{\phi}_i\hat{Y}^*_{t-1,i}) \tag{5.1}$$

$$\hat{\mathrm{P}}(Y_t = 2 \mid \hat{\boldsymbol{\beta}}_i, \hat{\phi}_i, \hat{\boldsymbol{\alpha}}_i, \hat{Y}^*_{t-1,i}) := \Phi(\hat{\alpha}_{2,i} - \mathbf{X}'_t\hat{\boldsymbol{\beta}}_i - \hat{\phi}_i\hat{Y}^*_{t-1,i}) - \Phi(\hat{\alpha}_{1,i} - \mathbf{X}'_t\hat{\boldsymbol{\beta}}_i - \hat{\phi}_i\hat{Y}^*_{t-1,i}) \tag{5.2}$$

$$\hat{\mathrm{P}}(Y_t = 3 \mid \hat{\boldsymbol{\beta}}_i, \hat{\phi}_i, \hat{\boldsymbol{\alpha}}_i, \hat{Y}^*_{t-1,i}) := \Phi(\hat{\alpha}_{3,i} - \mathbf{X}'_t\hat{\boldsymbol{\beta}}_i - \hat{\phi}_i\hat{Y}^*_{t-1,i}) - \Phi(\hat{\alpha}_{2,i} - \mathbf{X}'_t\hat{\boldsymbol{\beta}}_i - \hat{\phi}_i\hat{Y}^*_{t-1,i}) \tag{5.3}$$

$$\hat{\mathrm{P}}(Y_t = 4 \mid \hat{\boldsymbol{\beta}}_i, \hat{\phi}_i, \hat{\boldsymbol{\alpha}}_i, \hat{Y}^*_{t-1,i}) := 1 - \Phi(\hat{\alpha}_{3,i} - \mathbf{X}'_t\hat{\boldsymbol{\beta}}_i - \hat{\phi}_i\hat{Y}^*_{t-1,i}) \tag{5.4}$$

Now we set

$$\hat{P}_{t,k} := \frac{1}{10000} \sum_{i=5001}^{15000} \hat{P}(Y_t = k \mid \hat{\boldsymbol{\beta}}_i, \hat{\phi}_i, \hat{\boldsymbol{\alpha}}_i, \hat{Y}^*_{t-1,i})$$

to get the 1-step-predictions for the response given by

$$\hat{Y}_t := \mathrm{argmax}_{k=1,\ldots,4} \hat{P}_{t,k}.$$

The frequencies of the predicted categories and the sum of squared errors defined by

$$SSE = \sum_{t=1}^{2000} (\hat{Y}_t - Y_t)^2$$

are shown in Table 7.

In both models categories 3 and 4 were nearly never predicted, a result that was somehow expected: The estimated value of the highest cutpoint $\hat{\alpha}_3$ (= 2.108) is relatively small compared to the noise variance (= 1). Therefore the influence of the noise to the process is relatively high. Because most responses lie in category 1 and only few in category 3 or 4, the emergence of these categories is connected to high values of ε^*_t. The noise, however, does not play any role in the predictions, and the

	Model A				Model B			
true category	predicted category							
	1	2	3	4	1	2	3	4
1	509	346	0	0	816	39	0	0
2	410	371	0	0	766	15	0	0
3	88	205	0	0	283	10	0	0
4	6	64	0	1	64	7	0	0
sum of sq. errors	1623				2551			

Table 7: Frequencies of predicted categories classified by true categories in Models A and B.

impact of the covariates on the predictions is too small to reach the rarely occurring categories 3 or 4.
Further we observe that Model A (with covariates) is clearly better than Model B (without covariates): The tendency of our predictions is quite good in Model A as the percentage of category-2-predictions raises with increasing true categories. In Model B, however, nearly always category 1 is predicted independent of the true category. Further the overall sums of squared errors show the superiority of Model A.
We like to note that the computation of these 1-step-predictions is very computer intensive. A first approximation however can be gotten by using the posterior mean estimates $\overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}}, \overline{\phi}$, and $\overline{Y}^*_{t-1}$ to compute probability estimates $\overline{\mathrm{P}}_{t,k} = \hat{\mathrm{P}}(Y_t = k \mid \overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}}, \overline{\phi}, \overline{Y}^*_{t-1})$ where $\hat{\mathrm{P}}$ is defined similarly to (5.1) to (5.4) and finally to set $\hat{\overline{Y}}_t := \mathrm{argmax}_{k=1,\ldots,4} \overline{P}_{t,k}$. Computations show that one gets hardly worse predictions if one uses this approximation.

5.4 Predictions

We are now interested in fast predictions based on posterior mean estimates for $\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi$, and $\mathbf{Y}^*$ which date back up to one hour, and based on the current values of the covariates. For this we use the approximation approach mentioned in Section 5.3. In particular, we split our data in different sections. The data of one fixed section contains only the response and covariates information from t_{min} to t_{max}. Employing the GM-MGMC Gibbs sampler and using again its iterations 3001 to 15000 we compute the posterior mean estimates $\overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}}, \overline{\phi}$, and $\overline{Y}^*_{t_{max}}$ for this fixed data section from t_{min} to t_{max}. Afterwards, we want to make predictions for $Y_{t_{max}+n}$, $n \geq 1$. For this

purpose we first compute estimates $\hat{Y}^*_{t_{max}+n}$ for $Y^*_{t_{max}+n},\ n \geq 1$, given by

$$\begin{aligned}\hat{Y}^*_{t_{max}+1} &:= \mathbf{X}'_{t_{max}+1}\,\overline{\boldsymbol{\beta}} + \overline{\phi}\,\overline{Y}^*_{t_{max}} \\ \hat{Y}^*_{t_{max}+n} &:= \mathbf{X}'_{t_{max}+n}\,\overline{\boldsymbol{\beta}} + \overline{\phi}\,\hat{Y}^*_{t_{max}+n-1},\ \ n \geq 2.\end{aligned}$$

Similar to Section 5.3 we get the following probability estimates for $k \in 1, \ldots, K$ (set $\overline{\alpha}_0 = -\infty$ and $\overline{\alpha}_K = \infty$):

$$\begin{aligned}\hat{\mathrm{P}}(Y_{t_{max}+1} = k \mid \overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}}, \overline{\phi}, \overline{Y}^*_{t_{max}}) &:= \Phi(\overline{\alpha}_k - \mathbf{X}'_{t_{max}+1}\overline{\boldsymbol{\beta}} - \overline{\phi}\,\overline{Y}^*_{t_{max}}) - \\ &\quad -\Phi(\overline{\alpha}_{k-1} - \mathbf{X}'_{t_{max}+1}\overline{\boldsymbol{\beta}} - \overline{\phi}\,\overline{Y}^*_{t_{max}}) \\ \hat{\mathrm{P}}(Y_{t_{max}+n} = k \mid \overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}}, \overline{\phi}, \hat{Y}^*_{t_{max}+n-1}) &:= \Phi(\overline{\alpha}_k - \mathbf{X}'_{t_{max}+n}\overline{\boldsymbol{\beta}} - \overline{\phi}\,\hat{Y}^*_{t_{max}+n-1}) - \\ &\quad -\Phi(\overline{\alpha}_{k-1} - \mathbf{X}'_{t_{max}+n}\overline{\boldsymbol{\beta}} - \overline{\phi}\,\hat{Y}^*_{t_{max}+n-1}),\ \ n \geq 2.\end{aligned}$$

That leads to the predictions

$$\hat{Y}_{t_{max}+n} := \mathrm{argmax}_{k=1,\ldots,4}\hat{\mathrm{P}}(Y_{t_{max}+n} = k \mid \ldots),\ n \geq 1.$$

We computed these predictions for different combinations of t_{min} and t_{max} and for $n = 1, \ldots, 500$. To have the possibility to set $t_{max} = 2000$ we took the data of further 500 transactions between 2:29 pm and 3:26 pm on Nov 13, 2000. The corresponding frequencies classified by predicted and true categories are given in Table 8.
Because of the small values of the estimated ϕ, the predictions in Model B depend mainly on the intercept. Therefore nearly always the same category was predicted in each subtable for this model. Except for the data section $\{t_{min} = 1, \ldots, t_{max} = 500\}$, where the sum of squared errors is nearly the same, model A is clearly better than model B.

We repeated all computations in Sections 5.3 and 5.4 also with a model containing the covariates ln(TIMEDIFF), BOS, and, in addition, with the interaction of these both (results not shown). However, this did not lead to a better fit to the data.
In addition, the computations were also be made for the direction of the price change as response with categories 1=down, 2=no change, and 3=up. The covariates TIMEDIFF, BOS and SIZE were all non-significant, the estimate for the autoregressive parameter, however, was about -0.31. This affirms the conjecture of a negative autocorrelation of lag 1 that appears immediately when looking on the stock prices: Very often a negative price change is followed by a positive one and vice versa.

6 Summary and discussion

In Chapter 2 we introduced a new threshold model for discrete-valued time series, where we assumed the existence of a latent process that includes both covariates and

$t_{min} = 1$, $t_{max} = 2000$	Model A				Model B			
true category	predicted category							
	1	2	3	4	1	2	3	4
1	118	100	0	0	218	0	0	0
2	116	86	0	0	202	0	0	0
3	16	47	0	0	63	0	0	0
4	2	15	0	0	17	0	0	0
sum of sq. errors	405				607			

$t_{min} = 1$, $t_{max} = 500$	Model A				Model B			
true category	predicted category							
	1	2	3	4	1	2	3	4
1	98	130	0	0	0	228	0	0
2	57	129	0	0	1	185	0	0
3	18	48	0	0	0	66	0	0
4	0	20	0	0	0	20	0	0
sum of sq. errors	387				375			

$t_{min} = 501$, $t_{max} = 1000$	Model A				Model B			
true category	predicted category							
	1	2	3	4	1	2	3	4
1	129	86	0	0	215	0	0	0
2	105	100	0	0	205	0	0	0
3	22	51	0	0	73	0	0	0
4	1	6	0	0	7	0	0	0
sum of sq. errors	363				560			

$t_{min} = 1001$, $t_{max} = 1500$	Model A				Model B			
true category	predicted category							
	1	2	3	4	1	2	3	4
1	120	98	0	0	218	0	0	0
2	94	87	0	0	181	0	0	0
3	32	49	0	0	81	0	0	0
4	3	17	0	0	20	0	0	0
sum of sq. errors	464				685			

$t_{min} = 1501$, $t_{max} = 2000$	Model A				Model B			
true category	predicted category							
	1	2	3	4	1	2	3	4
1	153	65	0	0	218	0	0	0
2	146	56	0	0	202	0	0	0
3	28	35	0	0	63	0	0	0
4	9	8	0	0	17	0	0	0
sum of sq. errors	471				607			

Table 8: Frequencies classified by predicted and true categories for different sections of the data.

an autoregressive component. The standard Gibbs sampler for the estimation of the parameters in this model shows an extremely slow convergence behavior in the produced chains. Therefore we developed a new GM-MGMC Gibbs sampler, using a so-called partial scale transformation group, whose elements operate on the random samples of the interesting posterior distribution. In contrast to the standard Gibbs sampler, this MG-MGMC Gibbs sampler shows a very satisfying behavior in most cases, which is supported through the results presented in Section 4. Finally we used the new Gibbs sampler to detect and to quantify significant covariates for the price development of the IBM stock. For the absolute values of the price changes mainly the bid-offer-spread and the logarithm of the elapsed time between two following transactions are important covariates.

Of course, other applications of the model are possible. For example, one could think about pain patients where the feeling of pain is measured on an ordinal scale, and where besides an autoregressive structure some covariates are supposed to influence the feeling of the patients.

We plan to extend the model in the direction that temporal effects as opening and closing periods are covered. Further, one could drop the assumption that the variance of the noise is constant. Then, of course, one will have to find again a method that guarantees an acceptable convergence of the Gibbs sampler. In addition, we would like to cover other time dependence structures as arising for example in stochastic volatility models. See for example Taylor (1994), Shephard (1996), or Ghysels, Harvey, and Renault (1996) for a discussion of these models. Chib, Nardari, and Shephard (2002) considered MCMC algorithms for SV models with real valued response.

References

Albert, J. H. and S. Chib (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association 88*, 669–679.

Bauwens, L. and P. Giot (2000). The Logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique 60*, 117–149.

Bauwens, L. and P. Giot (2001). *Econometric Modelling of Stock Market Intraday Activity*. Boston: Kluwer Academic Publishers.

Bauwens, L. and D. Veredas (1999). The stochastic conditional duration model: a latent factor model for the analysis of financial durations. Discussion Paper 9958. CORE, Université catholique de Louvain. Forthcoming in *Journal of Econometrics*.

Chen, M.-H., Q.-M. Shao, and J. G. Ibrahim (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics 108*, 281–316.

Cowles, M. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing 6*, 101–111.

Dacorogna, M. M., R. Gençay, U. A. Müller, R. B. Olsen, and O. V. Pictet (2001). *An Introduction to High Frequency Finance*. San Diego: Academic Press.

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration; a new model for irregularly spaced transaction data. *Econometrica 66*, 1127–1162.

Geweke, J. (1991). Efficient Simulation from the Multivariate Normal and Student-t Distribution Subject to Linear Constraints and the Evaluation of Constraint Probabilities. *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface 23*, 571–578.

Ghysels, E., A. Harvey, and E. Renault (1996). Stochastic volatility, in *Statistical Methods in Finance* (eds. C.R.Rao and G.S.Maddala), 119-191. Amsterdam: North-Holland.

Hausman, J. A., A. W. Lo, and A. C. MacKinlay (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics 31*, 319–379.

Jasiak, J. (1998). Persistence in intertrade durations. *Finance 19*, 166–195.

Liu, J. S. and C. Sabatti (2000). Generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika 87*, 353–369.

Rao, M. M. (1987). *Measure Theory and Integration*. New York: Wiley.

Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and Computing 5*, 121–125.

Rydberg, T. H. and N. Shephard (2002). Dynamics of trade-by-trade price movements: decomposition and models. Forthcoming in *Journal of Econometrics*.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In *Time Series Models with Econometric, Finance and Other Applications* (eds. D.R.Cox, D.V.Hinkley and O.E.Barndorff-Nielson), 1-67. London: Chapman and Hall.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association 82*, 528–540.

Taylor, S. (1994). Modelling stochastic volatility. *Mathematical Finance 4*, 183–204.

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

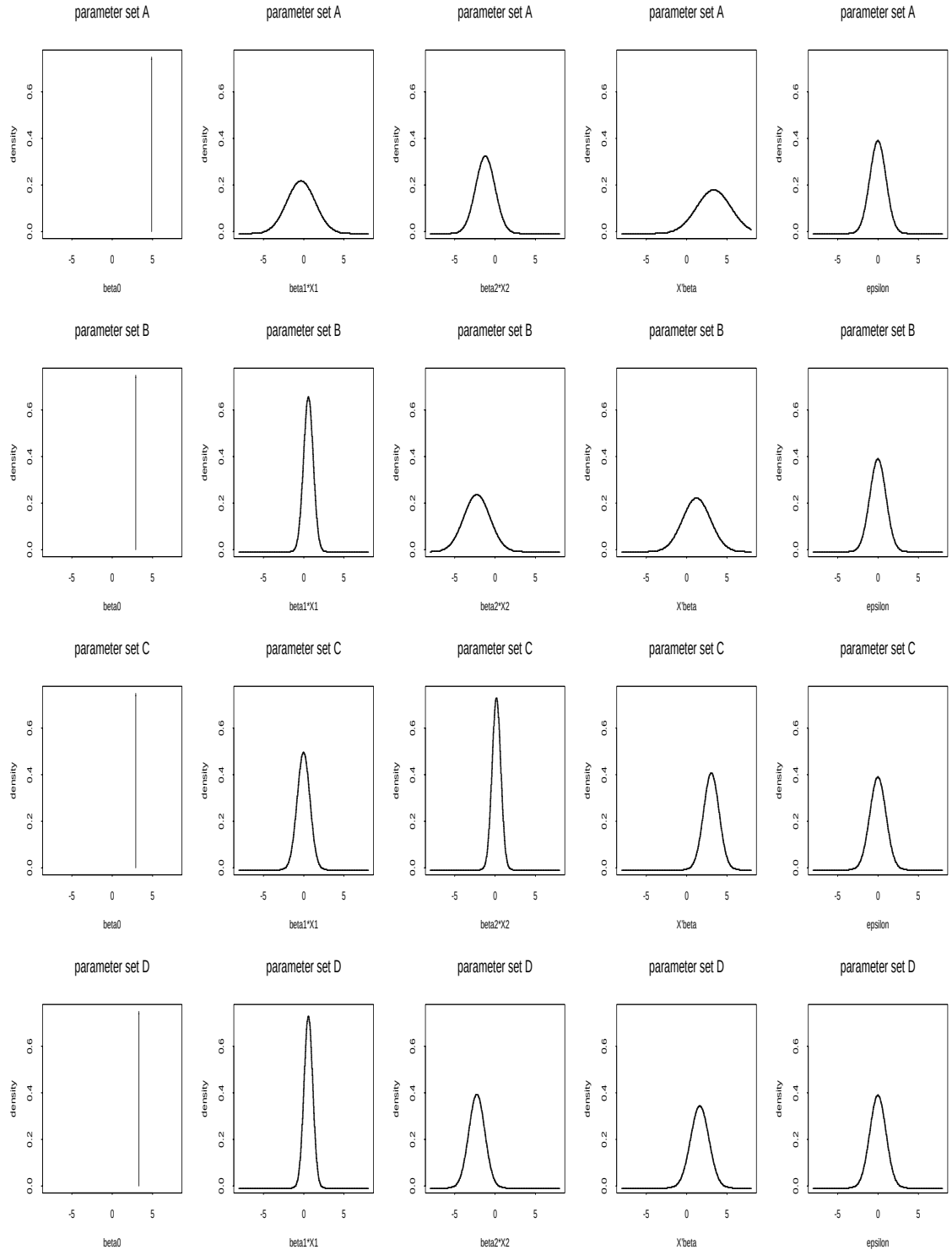

Figure 1: Generating densities for $\beta_j X_{tj},\ j = 1, 2$, in the parameter sets A to D, densities for the linear predictor $\mathbf{X}_t'\boldsymbol{\beta}$ and the error component $\varepsilon_t^* \sim N(0, 1)$.

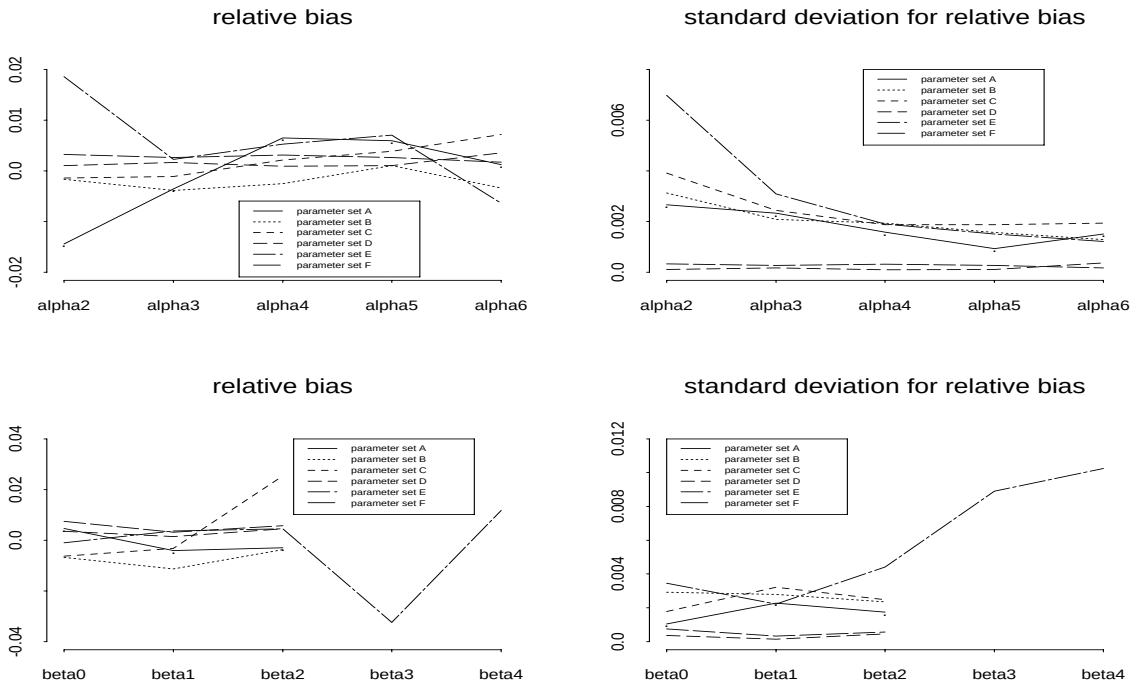

Figure 2: Relative bias and its standard deviation for the cutpoints and the regression coefficients.

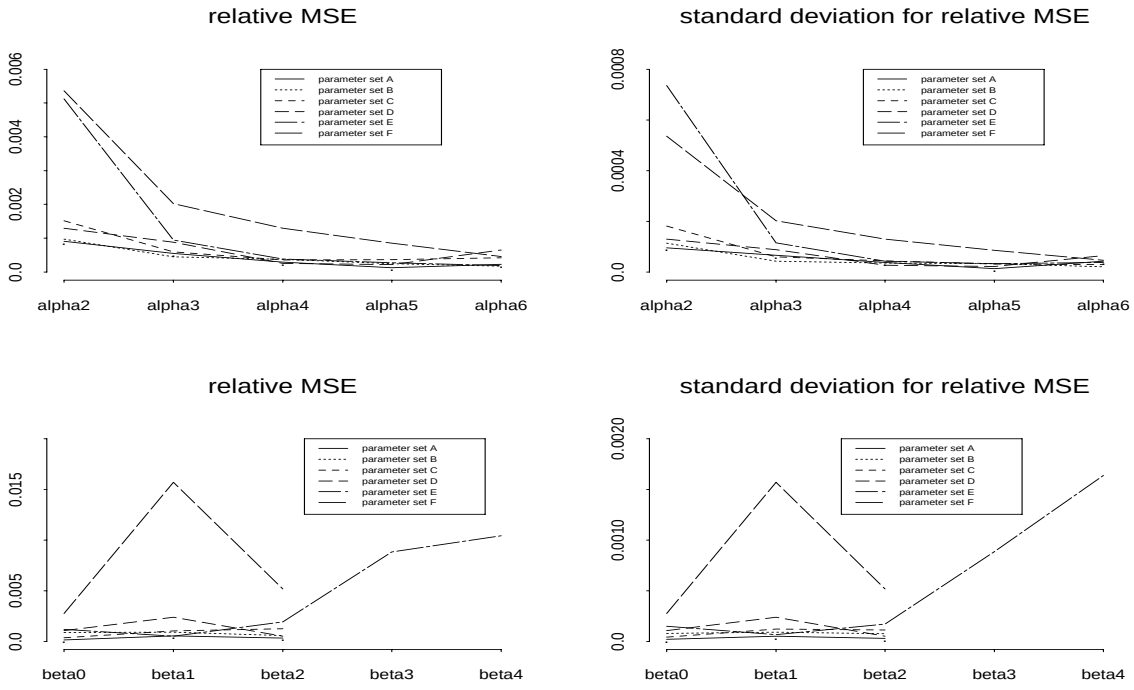

Figure 3: Relative MSE and its standard deviation for the cutpoints and the regression coefficients.

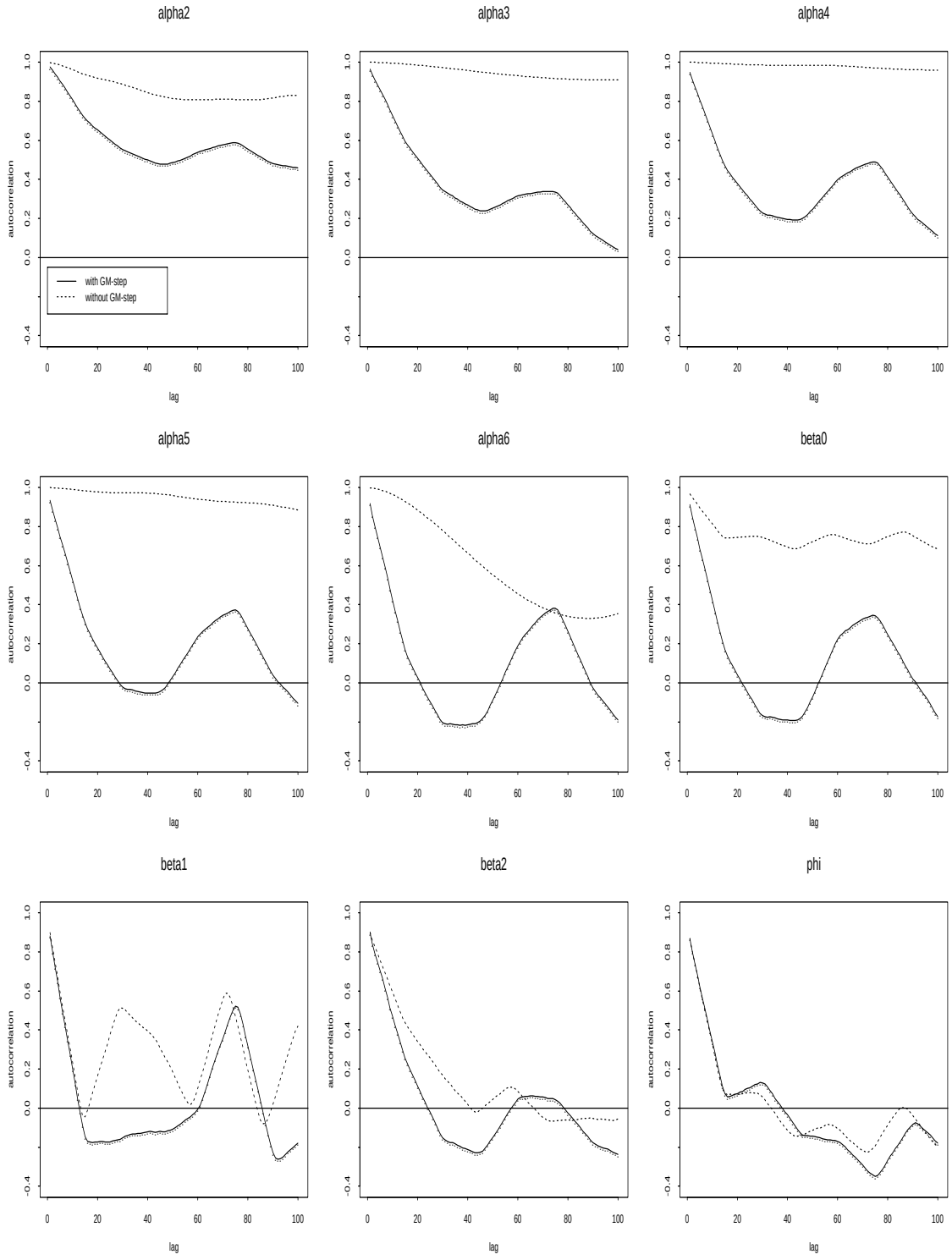

Figure 4: Autocorrelations produced by the Gibbs sampler with GM-step (solid line) and the Gibbs sampler without GM-step (dotted line) for a specific data set from parameter set A.

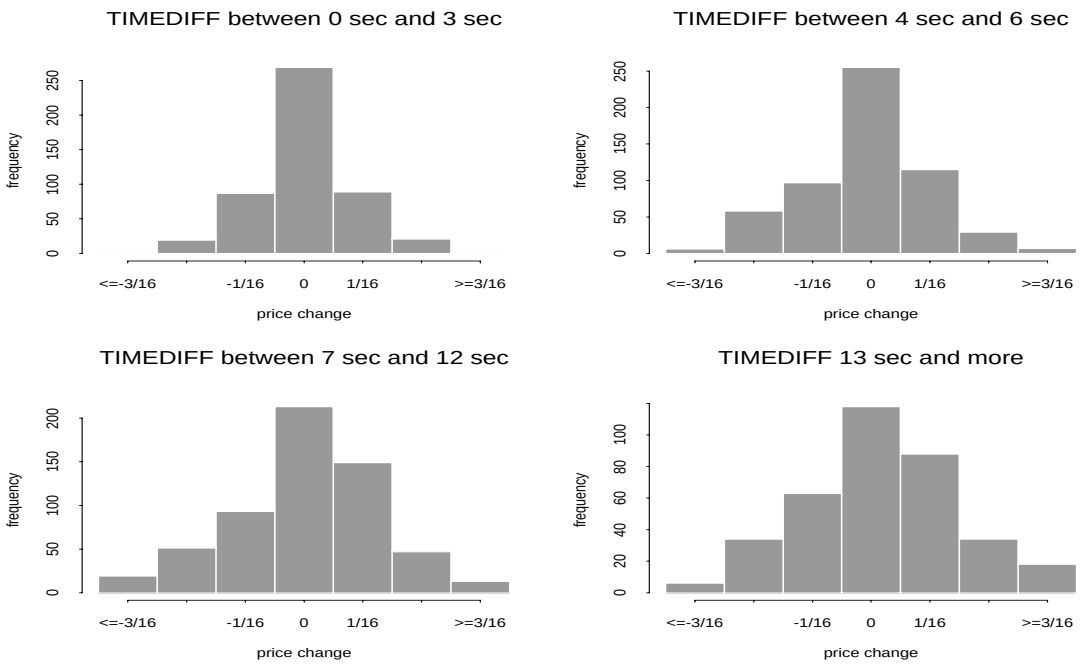

Figure 5: Price change histograms grouped by elapsed time since last transaction.

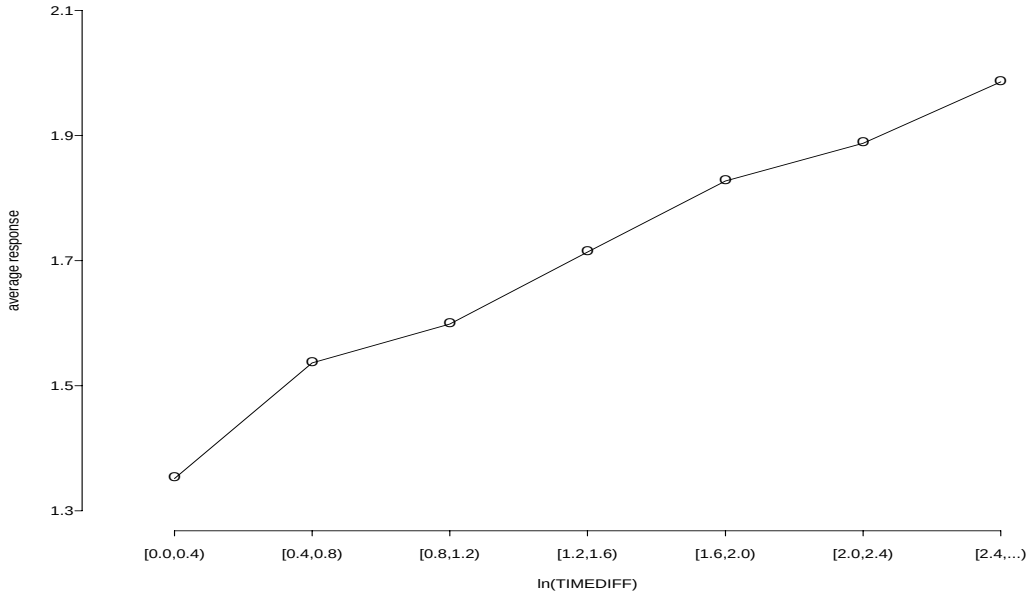

Figure 6: Relationship between response and the covariate ln(TIMEDIFF).

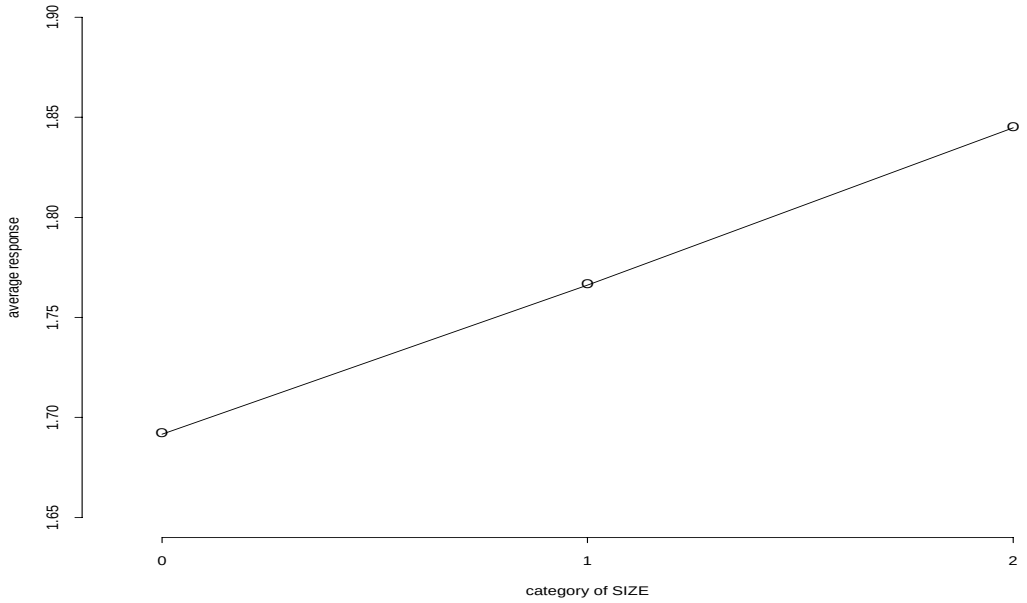

Figure 7: Relationship between response and category of covariate SIZE.

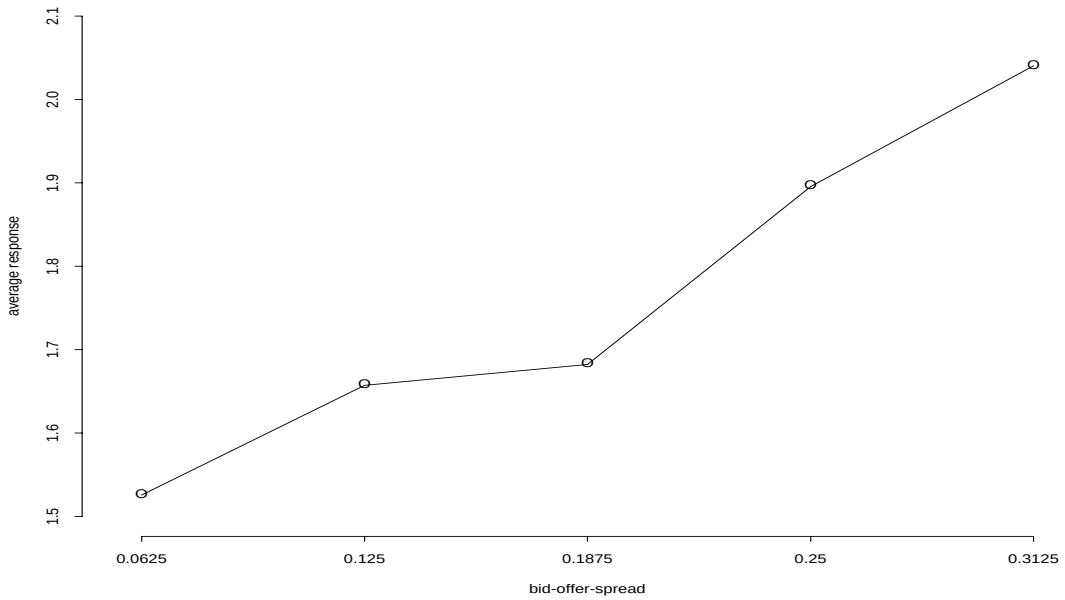

Figure 8: Relationship between response and covariate bid-offer-spread.

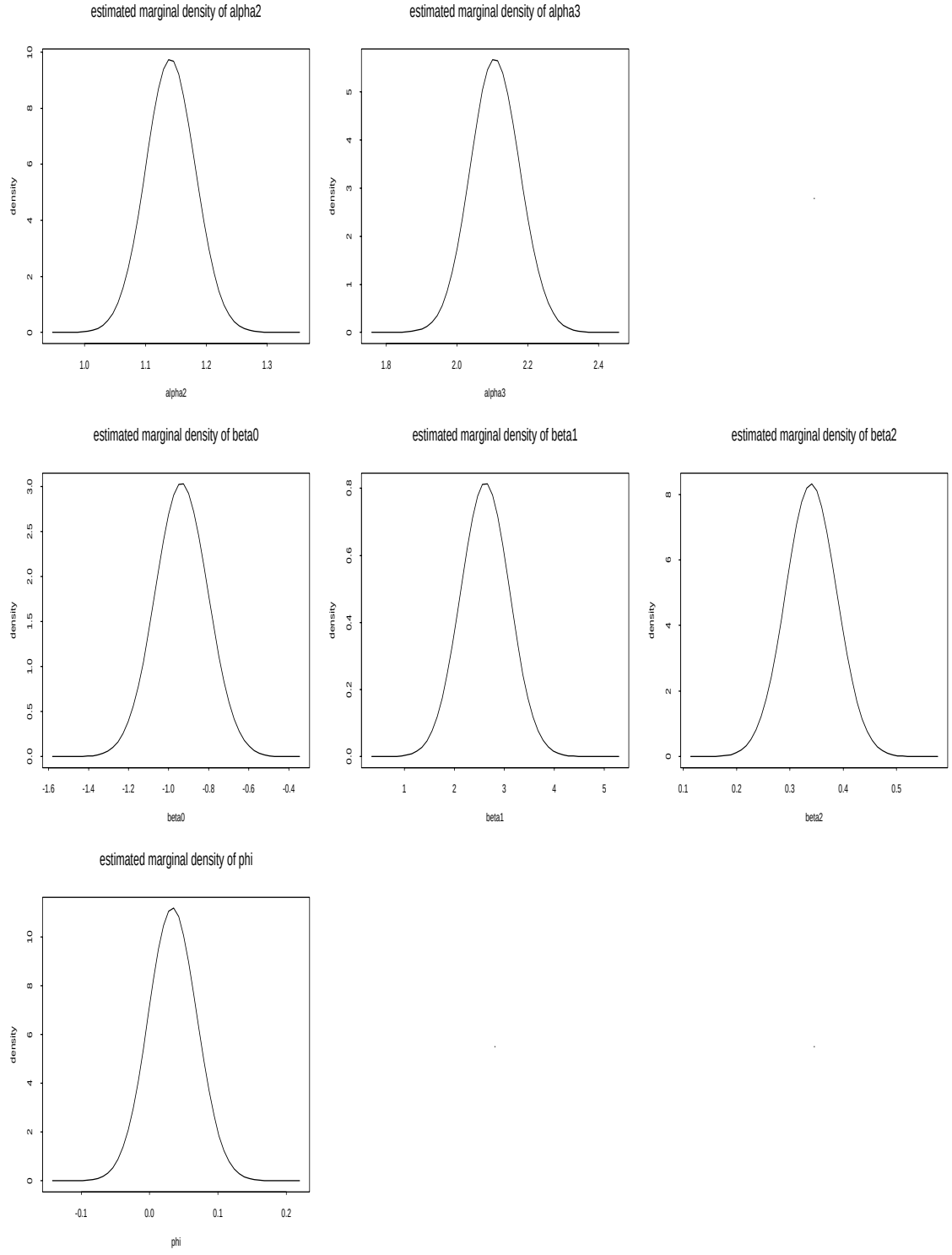

Figure 9: Estimated marginal densities for Model A: $Y_t^* = \beta_0 + \beta_1 \cdot \mathrm{BOS}_t + \beta_2 \cdot \ln(\mathrm{TIMEDIFF})_t + \phi Y_{t-1}^*$.

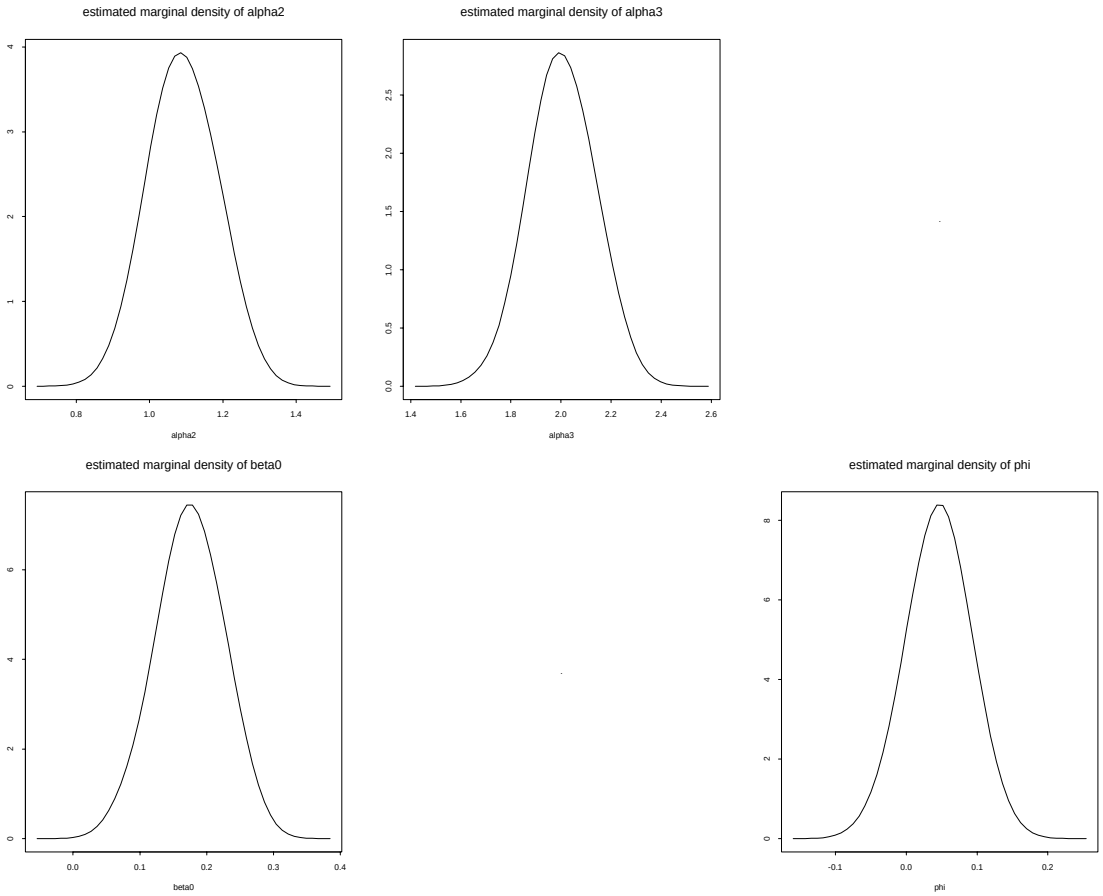

Figure 10: Estimated marginal densities for Model B: $Y_t^* = \beta_0 + \phi Y_{t-1}^*$.