Müller, Czado:

# Stochastic volatility models for ordinal valued time series with application to finance

Sonderforschungsbereich 386, Paper 504 (2006)

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

Projektpartner



# Stochastic Volatility Models for Ordinal Valued Time Series with Application to Finance

Gernot MÜLLER*, Claudia CZADO*

**Abstract**

In this paper we introduce two stochastic volatility models where the response variable takes on only finite many ordered values. Corresponding time series occur in high-frequency finance when the stocks are traded on a coarse grid. For parameter estimation we develop an efficient Grouped Move Multigrid Monte Carlo (GM-MGMC) sampler. We apply both models to price changes of the IBM stock in January, 2001 at the NYSE. Dependencies of the price change process on covariates are quantified and compared with theoretical considerations on such processes. We also investigate whether this data set requires modeling with a heavy-tailed Student-t distribution.

**Keywords:** Grouped move; High-frequency finance; Markov chain Monte Carlo; Multigrid Monte Carlo; Price process;

*Center for Mathematics, Munich University of Technology, Boltzmannstraße 3, 85747 Garching/Munich, Germany, emails: {mueller,cczado}@ma.tum.de.

# 1 INTRODUCTION

Stochastic volatility (SV) models offer a useful tool of modeling time series with non-constant volatility. Perhaps the most important field for the application of these models is finance, and in the last years many different versions of SV models for financial applications appeared, cf. for example Taylor (1994) or Shephard (1996). Recently, also many continuous-time SV models are discussed, cf. for example Barndorff-Nielsen and Shephard (2002).

In modeling stock prices one usually decides to model either the log-returns or the price change process itself. However, when a stock is traded on a coarse grid, for example a 1/16 dollar grid, and when the range of the stock prices in the data set at hand is not too large, it seems not adequate to consider the log-returns and to apply a continuous-response model. The reason for this is that the possible log-returns then occur in clusters. Because of the small price range, these clusters are clearly separated from each other. This feature of the data set can only be captured by discrete-response models. The aim of this paper is therefore to suggest a new model for the price change process together with an efficient estimation procedure. Moreover, we will provide volatility estimates for this process and use this model to prove empirically some theoretical considerations on the price change process.

Up to now only few discrete-response models have been suggested to model price changes. The approach by Hausman, Lo, and MacKinlay (1992) applied the common ordered probit model to such data. Another approach was made by Rydberg and Shephard (2003). They suggested 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. Recently, Russell and Engle (2005) introduced the ACM-ACD model which combines an autoregressive conditional multinomial model for the price changes and an ACD model for the durations following Engle and Russell (1998). In their context our models could be viewed as an alternative for the ACM model, since we consider the durations as given covariates and model the price changes conditioned on these durations and other covariates.

We introduce two SV models for ordinal valued time series where also exogenous variables are involved. We call these models ordinal-response stochastic volatility (OSV) models. Whereas the first model incorporates normally distributed errors, the second model allows for Student-t distributed errors. To distinguish between these two models, we abbreviate in following the first one by OSV and the second

one by OSVt.

For fitting the OSV and the OSVt model, we apply Markov chain sampling methods to simulate the posterior distribution of the parameters and latent variables. In addition, we utilize the method of Liu and Sabatti (2000) to improve the efficiency of the sampling procedure. In particular, a randomly drawn element from a transformation group is used to scale some of the parameters and latent variables in each iteration. We show that the use of this scaling leads to a very satisfying mixing of the MCMC output. Procedures which involve such grouped moves are called Grouped Move Multigrid Monte Carlo (GM-MGMC) algorithms.

We apply the OSV and OSVt model to price changes of the IBM stock at the New York Stock Exchange in January 2001. We detect and quantify the impact of covariates on several levels of the OSV process. The feature of the data that the observations are irregularly spaced is taken into account by using a corresponding covariate in the volatility equation. Further we investigate whether the data set requires modeling with a heavy-tailed Student-t distribution. For this, we use the OSVt model, modify the GM-MGMC algorithm for the OSVt case, and estimate also the degrees of freedom of the additional Student-t distribution in the IBM data set.

The paper is organized as follows. In Section 2 we introduce the OSV model (with normally distributed errors), and construct a GM-MGMC sampler for parameter estimation. In Section 3 we apply the OSV model to the IBM data set. In Section 4 the OSV model is extended to the OSVt model with Student-t errors. This OSVt model is then applied to the same IBM data set. Section 5 provides a short summary and discussion. Since the main application of these models is to high-frequency financial data we finally refer to Bauwens and Giot (2001) or Dacorogna, Gençay, Müller, Olsen, and Pictet (2001) for a global overview about models for such data.

# 2 OSV MODEL AND PARAMETER ESTIMATION

## 2.1 Model formulation

To cover the main features of the price change process as ordinal response, non-constant volatility, and dependence on covariates we introduce the **Ordinal-response Stochastic Volatility (OSV) Model**

defined by the following three equations:

$$y_t = k \quad \Leftrightarrow \quad y_t^* \in [c_{k-1}, c_k)\,, \tag{2.1}$$

$$y_t^* = \boldsymbol{x}_t'\boldsymbol{\beta} + \exp(h_t^*/2)\varepsilon_t^*, \tag{2.2}$$

$$h_t^* = \mu + \boldsymbol{z}_t'\boldsymbol{\alpha} + \phi(h_{t-1}^* - \mu - \boldsymbol{z}_{t-1}'\boldsymbol{\alpha}) + \sigma\eta_t^*, \tag{2.3}$$

where $\varepsilon_t^* \overset{\text{i.i.d.}}{\sim} N(0,1)$ independent of $\eta_t^* \overset{\text{i.i.d.}}{\sim} N(0,1)$, $k \in \{1,\ldots,K\}$, $t \in \{1,\ldots,T\}$, and $h_0^* = \mu$. $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$ are vectors of covariates, for $t = 0$ we set

$$\boldsymbol{z}_0 := (0,\ldots,0)'. \tag{2.4}$$

The parameters $c_1,\ldots,c_{K-1}$ are cutpoints. For notational convenience we set $c_0 := -\infty$ and $c_K := +\infty$. Additionally we define $\boldsymbol{y} := (y_1,\ldots,y_T)$.

We call the latent variables $y_t^*$ the continuous versions of the observations $y_t$. Given the covariate vector $\boldsymbol{x}_t$ and the latent variable $h_t^*$, $y_t^*$ is normally distributed with conditional mean $\boldsymbol{x}_t'\boldsymbol{\beta}$ and conditional variance $\exp(h_t^*)$. The log-volatilities $h_t^*$ form an autoregressive process of order one with impact of another covariate vector $\boldsymbol{z}_t$. Since the expression $\boldsymbol{z}_{t-1}'\boldsymbol{\alpha}$ is subtracted in the brace in Equation (2.3), the vector $\boldsymbol{z}_t$ has an impact only on $h_t^*$ but not on future log-volatilities $h_s^*$, $s > t$.

For reasons of identifiability we set $c_1 = 0$. Furthermore, we have to fix another cutpoint or, alternatively, the intercept $\mu$ in the log-volatility evolution equation (2.3). Since some of the equations in Section 2.2, where we develop the GM-MGMC sampler, would not longer be true if we would fix another cutpoint, we fix $\mu$. Obviously a large value for $\mu$ would heavily increase the volatility $\exp(h_t^*)$ and therefore the cutpoints would also become very large. For this we set $\mu = -0.6$ which leads to non-extreme parameter estimates in the data sets considered later.

In addition to the parameters which appear in the continuous-response SV-model of Chib et al. (2002) we have to estimate the cutpoints $\boldsymbol{c} := (c_2,\ldots,c_{K-1})$, and the latent continuous versions of $y_t$, $\boldsymbol{y}^* := (y_1^*,\ldots,y_T^*)$. We emphasize that $\boldsymbol{x}_t = (1, x_{t1},\ldots,x_{tp})'$ contains an intercept, whereas $\boldsymbol{z}_t = (z_{t1},\ldots,z_{tq})'$ does not, since Equation (2.3) already contains the (fixed) intercept $\mu$.

## 2.2 GM-MGMC algorithm

For estimating the parameters in the OSV model we constructed first an Hybrid MCMC sampler without any steps to speed up the convergence. However, simulations showed that this Hybrid MCMC

sampler produces chains which converge very slowly to the region around the true values, especially for the cutpoints $c_k$ and the regression intercept $\beta_0$. We encountered similar problems already for the autoregressive ordered probit (AOP) model in Müller and Czado (2005). There we achieved a fast convergence by an appropriate grouped move step which was inserted after each iteration of the standard Gibbs sampler.

The idea of grouped move steps is based on the following Theorem in Liu and Sabatti (2000): If $\Gamma$ is a locally compact group of transformations defined on the sample space $\boldsymbol{S}$, $L$ its left-Haar measure, $\boldsymbol{w} \in \boldsymbol{S}$ follows a distribution with density $\pi$, and $\gamma \in \Gamma$ is drawn from $\pi(\gamma(\boldsymbol{w}))|J_\gamma(\boldsymbol{w})|L(d\gamma)$, with $J_\gamma(\boldsymbol{w}) = \det(\partial\gamma(\boldsymbol{w})/\partial\boldsymbol{w})$, $\partial\gamma(\boldsymbol{w})/\partial\boldsymbol{w}$ the Jacobian matrix, then $\boldsymbol{w}^* = \gamma(\boldsymbol{w})$ has density $\pi$, too (Liu and Sabatti (2000), Theorem 1). The difficulty in developing a suitable grouped move step is how to choose the distribution $\pi$ and the transformation group $\Gamma$. Obviously, an improvement in convergence without too high computational cost can only be expected, when on the one hand the problematic parameters are transformed and on the other hand the distribution $\pi(\gamma(\boldsymbol{x}))|J_\gamma(\boldsymbol{x})|L(d\gamma)$ allows to draw samples very fast. We develop an appropriate grouped move step in Section 2.2.2.

The GM-MGMC algorithm which we present here consists of three parts. The first and the third part contain the updates of the Hybrid sampler. In the first part, the regression parameter vector $\boldsymbol{\beta}$, the latent variables $y_t^*$, $t = 1, \ldots, T$, and the cutpoints $c_2, \ldots, c_{K-1}$ are updated. In the second part we apply a grouped move step to achieve fast convergence. The third part updates the remainung parameters. Here we use ideas of the MCMC sampler developed in Chib, Nardari, and Shephard (2002), in particular a state space approximation of the latent process (2.2) and (2.3). This state space approximation is equivalent to that one used in Chib et al. (2002) and is therefore only briefly discussed in Section 2.2.1. However, there are two differences to the sampler of Chib, Nardari, and Shephard (2002). First, the SV model considered there uses the equation $h_t^* = \mu + \boldsymbol{z}_t'\boldsymbol{\alpha} + \phi(h_{t-1}^* - \mu) + \sigma\eta_t^*$ instead of Equation (2.3), so that the covariate vector $\boldsymbol{z}_t$ also has an impact on future log-volatilities $h_s^*$ for $s > t$. Secondly, we use a multivariate normal proposal distribution instead of the multivariate t-distribution in Chib et al. (2002) for the update of $\boldsymbol{\alpha}$, $\phi$, and $\sigma$. Since the updates in the first and third part of the algorithm do not require any sophisticated ideas, we skip further details on them to the appendix.

### 2.2.1 State space approximation and prior distributions

Obviously, Equation (2.2) is equivalent to

$$\log\left(y_t^* - \boldsymbol{x}_t'\boldsymbol{\beta}\right)^2 = h_t^* + \log\varepsilon_t^{*2}. \tag{2.5}$$

The distribution of $\log\varepsilon_t^{*2}$ can be approximated by a seven-component mixture of normals, as in Kim, Shephard, and Chib (1998). In particular, $\log\varepsilon_t^{*2} \approx \sum_{i=1}^{7} q_i u_t^{*(i)}$ where $u_t^{*(i)}$ is normally distributed with mean $m_i$ and variance $v_i^2$ independent of $t$. Moreover, the random variables $\{u_t^{*(i)} \,|\, t = 1,\ldots,T, i = 1,\ldots,7\}$ are independent. The quantity $q_i$ denotes the probability that the mixture component $i$ occurs. These probabilities are also independent of $t$ and are given in Chib et al. (2002), Table 1, together with the corresponding means and variances. Now let $s_t \in \{1,\ldots,7\}$ denote the component of the mixture that occurs at time $t$ and let $\pi(s_t)$ denote the prior for $s_t$, where $\pi(s_t = i) = q_i$. With $\tilde{y}_t^* := \log\left(y_t^* - \boldsymbol{x}_t'\boldsymbol{\beta}\right)^2$, Equation (2.5) leads to the following state space approximation of the latent process (2.2) and (2.3):

$$\tilde{y}_t^* = h_t^* + u_t^{*(s_t)}, \tag{2.6}$$

$$h_t^* = \mu + \boldsymbol{z}_t'\boldsymbol{\alpha} + \phi(h_{t-1}^* - \mu - \boldsymbol{z}_{t-1}'\boldsymbol{\alpha}) + \sigma\eta_t^*. \tag{2.7}$$

We will use this approximation for sampling $\phi$, $\sigma$, $\boldsymbol{\alpha}$, and $h_t^*$, $t = 1,\ldots,T$. Of course, we now have to sample in addition the unknown mixture indices $s_t$ for $t = 1,\ldots,T$. For notational convenience we define $\tilde{\boldsymbol{y}}^* := (\tilde{y}_1^*,\ldots,\tilde{y}_T^*)$, $\boldsymbol{s} := (s_1,\ldots,s_T)$, and $\boldsymbol{s}_{-t} := (s_1,\ldots,s_{t-1},s_{t+1},\ldots,s_T)$.

For the Bayesian approach we now specify the prior distributions for $\boldsymbol{c}$, $\boldsymbol{\beta}$, $h_0^*$, $\boldsymbol{\alpha}$, $\phi$, and $\sigma$. We assume prior independence so that the joint prior density can be written as

$$\pi(\boldsymbol{c},\boldsymbol{\beta},h_0^*,\boldsymbol{\alpha},\phi,\sigma) = \pi(\boldsymbol{c})\pi(\boldsymbol{\beta})\pi(h_0^*)\pi(\alpha_1)\cdots\pi(\alpha_q)\pi(\phi)\pi(\sigma).$$

For $\boldsymbol{\beta}$ we choose a multivariate normal prior distribution, for $h_0^*$ the Dirac measure at $\mu$, and for the remaining parameters uniform priors. In particular,

$$\pi(\boldsymbol{c}) = \mathbf{1}_{\{0<c_2<\ldots<c_{K-1}<C\}}, \qquad \pi(\boldsymbol{\beta}) = N_{p+1}(\boldsymbol{\beta}\,|\,\boldsymbol{b}_0,B_0), \qquad \pi(h_0^*) = \mathbf{1}_{\{h_0^*=\mu\}},$$

$$\pi(\alpha_j) = \mathbf{1}_{(-C_\alpha,C_\alpha)}(\alpha_j),\ j = 1,\ldots,q, \qquad \pi(\phi) = \mathbf{1}_{(-1,1)}(\phi), \qquad \pi(\sigma) = \mathbf{1}_{(0,C_\sigma)}(\sigma),$$

where $C > 0$, $C_\alpha > 0$, and $C_\sigma > 0$ are (known) hyperparameters, as well as the mean vector $\boldsymbol{b}_0$ and the covariance matrix $B_0$.

### 2.2.2 Development of an appropriate grouped move step

For the AOP model in Müller and Czado (2005) one succeeded by using the posterior distribution and the partial scale group for $\pi$ and $\Gamma$, respectively. Here, however, it seems to be impossible to find a suitable transformation group when $\pi$ is the full posterior distribution. Therefore we apply the Theorem of Liu and Sabatti (2000) not to the full posterior distribution, but on a certain conditional one. In particular, we set

$$\boldsymbol{w} := (y_1^*, \ldots, y_T^*, c_2, \ldots, c_{K-1}, \beta_0, \ldots, \beta_p)\,,$$

combine the remaining parameters to $\boldsymbol{R} := (\boldsymbol{h}^*, \boldsymbol{\alpha}, \phi, \sigma)$, and let $\pi(\boldsymbol{w})$ be the conditional distribution $f(\boldsymbol{w}|\boldsymbol{y}, \boldsymbol{R})$. With $\mathcal{F}_t$ denoting the observations until time $t$, i.e. $\mathcal{F}_t = (y_1, \ldots, y_t)$, we see that

$$f(y_1^*, \ldots, y_t^*, \boldsymbol{y}, \boldsymbol{c}, \boldsymbol{\beta}, \boldsymbol{R}) \;=\; f(y_t|y_t^*, \boldsymbol{c})\, f(y_t^*|\boldsymbol{\beta}, \boldsymbol{R})\, f(y_1^*, \ldots, y_{t-1}^*, \mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\beta}, \boldsymbol{R}). \tag{2.8}$$

Using Equation (2.8) recursively we have the following proportionality for $\pi(\boldsymbol{w})$:

$$\pi(\boldsymbol{w}) \;=\; f(y_1^*, \ldots, y_T^*, \boldsymbol{c}, \boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{R}) \;\propto\; \left[\prod_{t=1}^{T} f(y_t|y_t^*, \boldsymbol{c})\right] \left[\prod_{t=1}^{T} f(y_t^*|\boldsymbol{\beta}, \boldsymbol{R})\right] \pi(\boldsymbol{c}, \boldsymbol{\beta}, \boldsymbol{R}).$$

At this point we now set the mean of the normal prior for $\boldsymbol{\beta}$ to zero, i.e. $\pi(\boldsymbol{\beta}) = N_{p+1}(\boldsymbol{\beta}\,|\,\mathbf{0}, B_0)$. Otherwise some of the transformations in the following cannot be made. Since we further assume $\boldsymbol{c}$, $\boldsymbol{\beta}$, and $\boldsymbol{R}$ to be a priori independent of each other and a noninformative prior for the cutpoints, we get

$$\pi(\boldsymbol{w}) \;\propto\; \left[\prod_{t=1}^{T} \mathbf{1}_{[c_{y_t-1}, c_{y_t})}(y_t^*)\right] \mathbf{1}_{\{0<c_2<\ldots<c_{K-1}<C\}} \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T} \frac{(y_t^* - \boldsymbol{x}_t'\boldsymbol{\beta})^2}{\exp(h_t^*)} + \boldsymbol{\beta}' B_0^{-1} \boldsymbol{\beta}\right]\right\}.$$

In order to get an easy sampling distribution we now use the scale group

$$\Gamma = \{\gamma > 0 : \gamma(\boldsymbol{w}) = (\gamma w_1, \ldots, \gamma w_d)\}$$

with $\gamma^{-1} d\gamma$ as left-Haar measure. In this case $\gamma$ has to be drawn from $\gamma^{d-1}\pi(\gamma \boldsymbol{w})$, where $d$ denotes the dimension of $\boldsymbol{w}$. Since $\boldsymbol{w}$ contains all the latent variables $y_t^*$, $t = 1, \ldots, T$, the cutpoints $c_k$, $k = 2, \ldots, K-1$, and $\beta_j$, $j = 0, \ldots, p$, we have $d = T + K + p - 1$. Therefore we get the proportionality

$$\begin{aligned}\gamma^{d-1}\pi(\gamma\boldsymbol{w}) \;\propto\; \gamma^{T+K+p-2} &\left[\prod_{t=1}^{T} \mathbf{1}_{[\gamma c_{y_t-1}, \gamma c_{y_t})}(\gamma y_t^*)\right] \mathbf{1}_{\{0<\gamma c_2<\ldots<\gamma c_{K-1}<C\}} \\ &\cdot \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T} \frac{(\gamma y_t^* - \boldsymbol{x}_t'\gamma\boldsymbol{\beta})^2}{\exp(h_t^*)} + \gamma\boldsymbol{\beta}' B_0^{-1} \gamma\boldsymbol{\beta}\right]\right\}.\end{aligned} \tag{2.9}$$

For all $\gamma > 0$ we have the equivalence

$$[0 < \gamma c_2 < \ldots < \gamma c_{K-1} < C \text{ and } c_{K-1} < C] \iff \left[0 < c_2 < \ldots < c_{K-1} < C \text{ and } \gamma^2 < C^2/c_{K-1}^2\right].$$

Since expression (2.9) is considered to be a density for $\gamma$ (up to a normalizing constant), and since during all updates the condition $0 < c_2 < \ldots < c_{K-1} < C$ is always fulfilled, this equivalence leads to the proportionality

$$\mathbf{1}_{\{0<\gamma c_2<\ldots<\gamma c_{K-1}<C\}} \;\propto\; \mathbf{1}_{\{0<\gamma c_2<\ldots<\gamma c_{K-1}<C\}}\mathbf{1}_{\{c_{K-1}<C\}} \;\propto\; \mathbf{1}_{\{\gamma^2<C^2/c_{K-1}^2\}}.$$

Therefore expression (2.9) simplifies to

$$\begin{aligned}
&\gamma^{T+K+p-2}\left[\prod_{t=1}^{T}\mathbf{1}_{[c_{y_t-1},c_{y_t})}(y_t^*)\right]\mathbf{1}_{\{\gamma^2<C^2/c_{K-1}^2\}}\exp\left\{-\frac{1}{2}\gamma^2\left[\sum_{t=1}^{T}\frac{(y_t^*-\boldsymbol{x}_t'\boldsymbol{\beta})^2}{\exp(h_t^*)}+\boldsymbol{\beta}'B_0^{-1}\boldsymbol{\beta}\right]\right\}\\
&\propto\quad \left(\gamma^2\right)^{\frac{T+K+p-2}{2}}\exp\left\{-\frac{1}{2}\gamma^2\left[\sum_{t=1}^{T}\frac{(y_t^*-\boldsymbol{x}_t'\boldsymbol{\beta})^2}{\exp(h_t^*)}+\boldsymbol{\beta}'B_0^{-1}\boldsymbol{\beta}\right]\right\}\mathbf{1}_{\{\gamma^2<C^2/c_{K-1}^2\}}.
\end{aligned}$$

If one chooses a prior for $\boldsymbol{c}$ with infinite support, i.e. $C = \infty$, this expression is proportional to a Gamma distribution $\Gamma(a, b)$ for $\gamma^2$ with parameters

$$a \;=\; \frac{T+K+p}{2} \qquad \text{and} \qquad b \;=\; \frac{1}{2}\left[\sum_{t=1}^{T}\frac{(y_t^*-\boldsymbol{x}_t'\boldsymbol{\beta})^2}{\exp(h_t^*)}+\boldsymbol{\beta}'B_0^{-1}\boldsymbol{\beta}\right], \tag{2.10}$$

where the $\Gamma(a, b)$ density is given by $f_{\Gamma(a,b)}(x) = b^a x^{a-1}e^{-bx}/\Gamma(a),\ x \geq 0$.

If a finite support for $\boldsymbol{c}$ is chosen, i.e. $C < \infty$, one gets a Gamma distribution for $\gamma^2$ with the same parameters as before, however truncated to $(0, C^2/c_{K-1}^2)$. Of course, one can easily sample also from this truncated Gamma distribution by rejection sampling.

If $\gamma^2$ is drawn from the (truncated) Gamma distribution with $a$ and $b$ given in (2.10), respectively, the Theorem by Liu and Sabatti (2000) guarantees that $\gamma\boldsymbol{w} = \sqrt{\gamma^2}\boldsymbol{w}$ can be considered as a sample from $\pi(\boldsymbol{w}) = f(\boldsymbol{w}|\boldsymbol{y}, \boldsymbol{R})$, if $\boldsymbol{w}$ itself is a sample from this conditional distribution. Such a sample is given directly after the updates of $\boldsymbol{\beta}$, $y_t^*$, $t = 1, \ldots, T$, and $c_k$, $k = 2, \ldots, K-1$. Therefore we insert the corresponding grouped move step exactly at this point in each iteration of the basic Hybrid sampler. Each iteration of the GM-MGMC sampler consists now of the following steps:

**Algorithm 1 One iteration of the GM-MGMC sampler for the OSV model**

1. **MCMC-Step (Part 1), more details in Appendix**

   - Draw $\boldsymbol{\beta}$ from $(p+1)$-variate normal.
   - Draw $y_t^*$, $t = 1, \ldots, T$, from truncated univariate normals.
   - Draw $c_k$, $k = 2, \ldots, K-1$, from Unif$(l_k, r_k)$ where

$$l_k \;=\; \max\{c_{k-1}, \max_{t=1,\ldots,T}\{y_t^*|y_t = k\}\}, \qquad r_k \;=\; \min\{c_{k+1}, \min_{t=1,\ldots,T}\{y_t^*|y_t = k+1\}\}.$$

Get $\boldsymbol{\beta}_{cur}, \boldsymbol{y}^*_{cur}, \boldsymbol{c}_{cur}$ as current values.

2. **GM-Step**

   Draw $\gamma^2$ from the (truncated) $\Gamma(a, b)$ distribution with $a$ and $b$ defined in (2.10), respectively, and update $\boldsymbol{\beta}^*_{cur}$, $\boldsymbol{y}^*_{cur}$, $\boldsymbol{c}_{cur}$ by multiplication with the group element $\gamma = \sqrt{\gamma^2}$,

$$\boldsymbol{\beta}_{new} \leftarrow \gamma\boldsymbol{\beta}_{cur}, \qquad \boldsymbol{y}^*_{new} \leftarrow \gamma\boldsymbol{y}^*_{cur}, \qquad \boldsymbol{c}_{new} \leftarrow \gamma\boldsymbol{c}_{cur}.$$

3. **MCMC-Step (Part 2), more details in Appendix**

   - Compute $\tilde{y}^*_t = \log(y^*_t - \boldsymbol{x}'_t\boldsymbol{\beta})^2$ for $t = 1, \ldots, T$.
   - Draw $s_t$, $t = 1, \ldots, T$, proportional to $\Pr(s_t)N(\tilde{y}^*_t | h^*_t + m_{s_t}, v^2_{s_t})$.
   - Draw $(\boldsymbol{\alpha}, \phi, \sigma)$ via Metropolis-Hastings step; use ML-estimates of $(\boldsymbol{\alpha}, \phi, \sigma)$ to find an adequate multivariate normal proposal.
   - Draw $\boldsymbol{h}^*$ in one block using the simulation smoother of De Jong and Shephard (1995).

## 2.3 Illustration and results of a simulation study

Now we illustrate the performance of the GM-MGMC sampler for a special parameter setting. Here and in the following we always use the hyperparameters $C = \infty$, $\boldsymbol{b}_0 = \mathbf{0}$, $B_0 = \text{diag}(10, \ldots, 10)$, $C_\alpha = 10^6$, $C_\sigma = 10$, so that the prior distributions are

$$\pi(\boldsymbol{c}) = \mathbf{1}_{\{0<c_2<\ldots<c_{K-1}<\infty\}}, \qquad \pi(\boldsymbol{\beta}) = N_{p+1}(\boldsymbol{\beta} \,|\, \mathbf{0}, \text{diag}(10, \ldots, 10)),$$
$$\pi(\alpha_j) = \mathbf{1}_{\{-10^6<\alpha_j<10^6\}}, \quad j = 1, \ldots, q, \qquad \pi(\phi) = \mathbf{1}_{\{-1<\phi<1\}}, \qquad \pi(\sigma) = \mathbf{1}_{\{0<\sigma<10\}}.$$

We simulate an OSV process of length $T = 22000$ where we allow for $K = 7$ response categories. For the log-volatility Equation (2.3) we use a two-dimensional covariate vector $\boldsymbol{z}_t$. The two components are exactly the covariates from the IBM data which will be used in Section 3. The simulation parameters in the log-volatility equation are set to $\alpha_1 = 0.25$, $\alpha_2 = 0.15$, $\phi = 0.90$ and $\sigma = 0.20$. Using these parameters we first simulate the log-volatility process $\{h^*_t \,|\, t = 1, \ldots, T\}$. The covariate vector $\boldsymbol{x}_t$ in the equation for the latent variables $y^*_t$ also has two components. The first corresponds to the intercept and is always 1, the second is the lagged response $y_{t-1}$. The simulation parameters are set to $\beta_0 = 3.50$ and $\beta_1 = -0.30$. To generate the response, we choose the cutpoints as $c_2 = 0.90$, $c_3 = 1.80$, $c_4 = 2.75$, $c_5 = 3.65$, and $c_6 = 4.50$. We note that these simulation parameters are chosen close to the estimated values for the IBM data investigated in Section 3. Therefore, by showing that the GM-MGMC sampler

works very well for these parameters and covariates, we get a first justification that the GM-MGMC algorithm works well for the IBM data.

We run the GM-MGMC sampler for 4000 iterations. As starting values for the cutpoints $c_2, \ldots, c_6$ we choose 2.0, 4.0, 6.0, 8.0, 10.0, respectively, 0.0 for each of the regression coefficients $\alpha_j$, 0.8 for $\phi$, and 0.3 for $\sigma$. Since each iteration starts with the $\boldsymbol{\beta}$-update, starting values for $\beta_j$, $j = 0, \ldots, p$, are not necessary.

Figure 1 demonstrates the very satisfying behaviour of the GM-MGMC sampler. The chains converge within about only 100 iterations for the cutpoints as well as for the regression coefficients and the parameters $\phi$ and $\sigma$. An interesting effect occurs at iterations 2 to 13 of the parameters $\alpha_1$, $\alpha_2$, $\phi$, and $\sigma$, which do not move during these 11 iterations. This effect is typical for the burn-in period, since the parameters $\alpha_1$, $\alpha_2$, $\phi$, and $\sigma$ are drawn by a Metropolis-Hastings step. During the iterations 2 to 13, the proposal values for these parameters were never accepted. However, when the chains of the other parameters get closer to the true values, the multivariate normal proposal for $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \phi, \sigma)$ approximates the target density very good, and the chains of the components of $\boldsymbol{\theta}$ start to converge. Beyond iteration 100, the average acceptance rate is about 90%. Since the proposal density for $\boldsymbol{\theta}$ is adapted to the target density very carefully, this leads to a fast mixing in the whole support of the target density.

We further investigate the empirical autocorrelations for the lags 0 to 200 in the observed chains after a burn-in period of 1000 iterations for both samplers. As can be seen from Figure 2, the empirical autocorrelations in the GM-MGMC chains decline very fast.

We conducted a simulation study to assess the accuracy of the posterior mean estimates by the GM-MGMC sampler. Since the behaviour of the GM-MGMC sampler was always similar to the previous illustration, we only summarize the results briefly. We considered two settings which differed in the choice of the covariates, the length of the data, and the values of the simulation parameters. For both settings the prior distributions from above were used, and 20 data sets were simulated with response categories $1, \ldots, 7$. We computed posterior mean estimates by running the GM-MGMC sampler for 4000 iterations each, where the first 1000 iterations were discarded for burn-in. In the first setting the data sets had length $T = 8000$. Running the GM-MGMC sampler for such data sets takes about 2.1 seconds per iteration on an UltraSPARC III Cu 900 Mhz processor. In the second setting the length of the data sets was $T = 22000$, which is close to the length of the IBM data set of Section 3. For both

settings the chains converged very rapidly and the autocorrelations in the chains were small. On average the posterior mean estimates agreed nearly with the true parameter values. The standard deviations for the posterior mean estimates were very small, so that the posterior mean estimates themselves had always been close to the true values.

# 3 APPLICATION TO IBM DATA

## 3.1 Data description

We investigate price changes of the IBM stock traded at the New York Stock Exchange (NYSE) from January 9, 2001 to January 25, 2001. In this period we removed data from Mondays and Fridays, and data before 09:50am and after 03:40pm to exclude data which might exhibit a special behavior. The minimum price in this period was 91.2500$, the maximum price 111.4375$. The prices and therefore also the price differences take on only values which are integer multiples of 1/16 US$. Since price changes of less than $-3/16$$ and more than $+3/16$$ hardly occur, we deal with them like price changes of $-3/16$$ and $+3/16$$, respectively. Therefore, we only observe seven different price changes. We associate these price changes to the categories $1, \ldots, 7$ in a natural way, as can be seen from Table 1. Adding up the frequencies in Table 1, we see that we have a total of 22689 observations.

The feature of the OSV model that the impact of exogenous variables can be captured allows for a simple strategy to deal with the irregularly spaced data: We use the covariate TIMEDIFF (the time which elapses between two subsequent transactions in seconds) to model the impact of different inter-transaction times. Furthermore we consider the covariate SIZE, i.e. the volume of the transaction. From the application of the AOP model to another IBM data set in Müller and Czado (2005) we know that these covariates have an impact on how large a price change is. We use log-transformations of TIMEDIFF and SIZE for the covariate vector $\boldsymbol{z}_t$ in the log-volatility equation. In addition, we center the covariate vector $\boldsymbol{z}_t$ at $\mathbf{0}$ for reasons of numeric stability. With $\text{TIMEDIFF}_t$ denoting the time which elapses between the transaction at time $t-1$ and the transaction at time $t$, and with $\text{SIZE}_t$ denoting the transaction volume at time $t$, we have

$$\begin{aligned} z_{t1} &:= \log(\text{TIMEDIFF}_t + 1) - \textstyle\sum_{k=1}^{22689} \left[\log(\text{TIMEDIFF}_k + 1)\right] / 22689, \\ z_{t2} &:= \log(\text{SIZE}_t) - \textstyle\sum_{k=1}^{22689} \left[\log(\text{SIZE}_k)\right] / 22689. \end{aligned}$$

Table 2 gives some summary information about these two covariates for the log-volatility equation.

Considering the response one can observe that often a positive price jump is followed by a negative one and vice versa. This can be taken into account by using the lagged response as covariate in the model Equation (2.2) since the covariates there have an impact on the mean of the latent variables $y_t^*$. Therefore we use $\boldsymbol{x}_t := (1, y_{t-1})'$ as covariate vector for Equation (2.2). We note that exploratory analyses for TIMEDIFF and SIZE show that often higher values of these covariates come along with higher price changes, but partly upwards and partly downwards. Therefore we do not expect an impact of these covariates on the mean of the latent variables $y_t^*$.

## 3.2 Parameter estimates, volatility estimates, and conclusions

With the range of the covariate values in mind, we decided to use hyperparameters which correspond to the prior distributions summarized in Table 3. We note explicitly that we must set $\boldsymbol{b}_0$ (mean of $\boldsymbol{\beta}$-prior) to $\mathbf{0}$ to be able to apply the grouped move step in the GM-MGMC algorithm (cf. Section 2.2.2). We checked out by using other hyperparameter values that the posterior estimates are not very prior-sensitive. This also can be expected because of the large number of observations.

We run the GM-MGMC sampler from Section 2.2 for 4000 iterations and discard the first 1000 iterations for burn-in. The results are summarized in Table 4. It shows the posterior mean estimates for the parameters of interest together with their corresponding estimated standard deviations and 90% credible intervals.

Since the 90% credible intervals for $\beta_0$, $\beta_1$, $\alpha_1$, and $\alpha_2$ are far away from zero, we conclude that the intercept, the lagged observation $y_{t-1}$, $\log(\text{TIMEDIFF}_t + 1)$, and $\log(\text{SIZE}_t)$ all have a significant impact on the new observation $y_t$. Also the credible intervals for the autoregressive parameter $\phi$ and the standard deviation $\sigma$ are far away from zero. The estimate 0.9061 for $\phi$ shows the high dependence of the log-volatility $h_t^*$ on the previous log-volatility $h_{t-1}^*$, it is, however, still away from the nonstationary case $\phi = 1$.

The estimated posterior marginal densities look all like densities from normal distributions and are therefore not shown here. The estimated autocorrelations in the chains produced by the GM-MGMC sampler after the burn-in period of 1000 iterations decline quite fast for all parameters. They look like the GM-MGMC-autocorrelations in Figure 2 and are therefore also not shown here. This justifies that

no subsampling is required to estimate the standard errors of the estimates.

The negative sign of the estimate $-0.3073$ for $\beta_1$ and the positive signs of the estimates 0.2599 for $\alpha_1$ and 0.1511 for $\alpha_2$ lead to the following qualitative conclusions:

- Positive price changes are often followed by negative ones and vice versa (this confirms what we can see directly from the data).
- The more time elapses between two subsequent transactions, the higher the (log-) volatility is, or, equivalently, the more time elapses, the higher the probability for a big price change is.
- The more stocks are traded, the higher the (log-) volatility is, or, equivalently, the more stocks are traded, the higher the probability for a big price change is.

These results agree with many publications about theoretical results for the price change process. Diamond and Verrecchia (1987) point out, that periods without transactions can be considered as a hint for the existence of bad news. Because of the prohibition of short-selling many investors cannot use bad information by selling. Therefore, longer periods between consecutive transactions usually lead to a higher volatility of the price change process. Following Easley and O'Hara (1987), well informed investors usually buy or sell large amounts of stocks in each transaction to take maximal advantage of their informations. Therefore, noninformed market participants associate large transaction volumes with existence of new information and trade themselves. Hence for large transaction volumes one can expect higher volatilities. The same dependence between market informations, transaction volumes, and expected volatility is derived by Tauchen and Pitts (1983).

Using the posterior mean estimates for $\alpha_1$ and $\alpha_2$ we can compare the impacts of TIMEDIFF and SIZE on the log-volatilities. From Table 2 we know that the time difference always lies between 0 and 116 seconds. For each value in this interval we compute the corresponding transformation $z_{\cdot 1}$ (cf. Section 3.1) and multiply the resulting value by the posterior mean estimate 0.2599 for $\alpha_1$. The same is done for the transaction volume with range 100 to 180000 stocks. Here the transformed values $z_{\cdot 2}$ are multiplied by the estimate 0.1511 for $\alpha_2$. The result can be seen in Figure 3. For the extreme values TIMEDIFF $= 0$ and TIMEDIFF $= 116$, the estimated impacts are about $-0.51$ and 0.73, respectively. The corresponding estimates for the covariate SIZE are $-0.32$ and 0.81. We conclude that the covariate TIMEDIFF affects the log-volatility slightly more than the covariate SIZE. Moreover, the impact of both covariates is quite large if one takes the posterior mean estimate 0.2230 for $\sigma$ into account.

From Section 2.1 we know that in the OSV model the log-volatility is not determined uniquely until the additive constant $\mu$ is fixed. For computational reasons we fixed this parameter to $= -0.6$, but in this context it may be more intuitive to consider log-volatilities with mean zero. Therefore we now define the **normalized volatility at time** $t$ by $v_t^n := \exp\{h_t^* - \mu\} = \exp\{h_t^* + 0.6\}$ since the covariates in the log-volatility equation were also centered at zero. In each iteration $i$ the GM-MGMC sampler produces estimates $\hat{h}_{t,i}^*$ of $h_t^*$, $t = 1, \ldots, 22689$, which can be used to get estimates $\hat{v}_t^n$ for the normalized volatilities $v_t$. Figure 4 shows the IBM stock prices and the estimated normalized volatilities on January 9, 09:50am to 03:40pm.

# 4 THE OSVt MODEL

Now we replace the normal distribution of the errors $\varepsilon_t^*$ in the OSV model by a Student-t distribution with $\nu$ degrees of freedom. Using the decomposition of a t-distributed random variable in a product of a normally and a Gamma-distributed random variable we define the **Ordinal-response Stochastic Volatility Model with t-distributed errors (OSVt Model)** by the following three equations:

$$y_t = k \quad \Leftrightarrow \quad y_t^* \in [c_{k-1}, c_k)\,, \tag{4.1}$$

$$y_t^* = \boldsymbol{x}_t'\boldsymbol{\beta} + \exp(h_t^*/2)\lambda_t^{*-1/2}\varepsilon_t^*, \tag{4.2}$$

$$h_t^* = \mu + \boldsymbol{z}_t'\boldsymbol{\alpha} + \phi(h_{t-1}^* - \mu - \boldsymbol{z}_{t-1}'\boldsymbol{\alpha}) + \sigma\eta_t^*, \tag{4.3}$$

where $\varepsilon_t^* \overset{\text{i.i.d.}}{\sim} N(0,1)$ independent of $\lambda_t^* \overset{\text{i.i.d.}}{\sim} \Gamma(\nu/2, \nu/2)$. In all other respects we assume the same conditions as for the OSV model. In addition to the parameters and variables to estimate in the OSV model, now the variables $\lambda_t^*$ and the parameter $\nu$ have to be estimated. For notational convenience we define $\boldsymbol{\lambda}^* := (\lambda_1^*, \ldots, \lambda_T^*)$ and $\boldsymbol{\lambda}_{-t}^* := (\lambda_1^*, \ldots, \lambda_{t-1}^*, \lambda_{t+1}^*, \ldots, \lambda_T^*)$.

## 4.1 GM-MCMC sampler for OSVt model

Since the OSVt model differs only slightly from the OSV model, the derivation of the updates is completely analogous to Section 2.2. Mostly, one has only to replace the term $\exp(h_t^*)$ by $\exp(h_t^*)\lambda_t^{*-1}$. Again one can use a state space approximation of Equations (4.2) and (4.3). However, in contrast to the OSV model, here one must compute $\tilde{y}_t^* := \log(y_t^* - \boldsymbol{x}_t'\boldsymbol{\beta})^2 + \log\lambda_t^*$, $t = 1, \ldots, T$. Using this definition of $\tilde{y}_t^*$ the updates of the mixture indices $s_t$, the complete Metropolis-Hastings step for the joint update of $\boldsymbol{\alpha}$, $\phi$, and $\sigma$, and the $\boldsymbol{h}^*$-update can be done exactly as in the OSV case.

In addition we now need an update for the degrees of freedom $\nu$ of the t-distribution, and updates for the variables $\lambda_t^*$, $t = 1, \ldots, T$. For $\nu$ we assume a uniform prior on the set $[1, 127]$. Since the t-distribution becomes more and more similar to the normal distribution as $\nu$ increases, the maximal value 127 for $\nu$ seems to be sufficiently large. Following the model definition, the variables $\lambda_t^*$ are assumed a priori independent of each other and $\Gamma(\nu/2, \nu/2)$-distributed. Moreover, we assume $\nu$ and $\lambda_t^*$, $t = 1, \ldots, T$, a priori independent of all other parameters and variables.

Again defining $\boldsymbol{\theta} := (\boldsymbol{\alpha}, \phi, \sigma)$, one can update $\nu$ and $\lambda_t^*$, $t = 1, \ldots, T$, by first drawing $\nu$ from $f(\nu \,|\, \boldsymbol{y}, \boldsymbol{c}, \boldsymbol{y}^*, \boldsymbol{\beta}, \boldsymbol{h}^*, \boldsymbol{\theta})$ and then sampling $\lambda_t^*$, $t = 1, \ldots, T$, from $f(\lambda_t^* \,|\, \boldsymbol{y}, \boldsymbol{c}, \boldsymbol{y}^*, \boldsymbol{\beta}, \boldsymbol{h}^*, \boldsymbol{\lambda}_{-t}, \nu, \boldsymbol{\theta})$. We draw $\nu$ by a Metropolis-Hastings step. As proposal distribution we use a truncated version of a Gamma-distribution. This choice has some computational advantages. Moreover, suitably chosen hyperparameters for this Gamma-distribution lead usually to acceptance probabilities of about 30 to 60 percent on average. As can be derived quite easily, the variables $\lambda_t^*$, $t = 1, \ldots, T$, have to be drawn from Gamma$(c, d_t)$-distributions, where $c = (\nu + 1)/2$ and $d_t = (\nu + (y_t^* - \boldsymbol{x}_t'\boldsymbol{\beta})^2 / \exp(h_t^*))/2$. The derivation of the GM step for the OSVt model is again analoguous to the OSV case. One uses the same vector $\boldsymbol{w} := (y_1^*, \ldots, y_T^*, c_2, \ldots, c_{K-1}, \beta_0, \ldots, \beta_p)$, but has now the following vector of remaining parameters: $\boldsymbol{R} := (\boldsymbol{h}^*, \boldsymbol{\lambda}^*, \boldsymbol{\alpha}, \phi, \sigma)$. Again one considers the conditional distribution $f(\boldsymbol{w}|\boldsymbol{y}, \boldsymbol{R})$. The scale group which was used as transformation group in the OSV case leads here also to a (truncated) Gamma-distribution for $\gamma^2$, however, with parameters

$$a \;=\; \frac{T + K + p}{2} \qquad \text{and} \qquad b \;=\; \frac{1}{2}\left[\sum_{t=1}^{T} \frac{(y_t^* - \boldsymbol{x}_t'\boldsymbol{\beta})^2 \lambda_t^*}{\exp(h_t)} + \boldsymbol{\beta}' B_0^{-1} \boldsymbol{\beta}\right]. \tag{4.4}$$

The GM-MGMC sampler for the OSVt model therefore consists of 3 MCMC parts and the GM step which must be inserted after the first MCMC part. Whereas for part 2 one switches to the state space approximation, parts 1 and 3 use the original model equations. The steps of one iteration of the GM-MGMC sampler are summarized in the following Algorithm 2.

**Algorithm 2 One iteration of the GM-MGMC sampler for the OSVt model**

1. **MCMC-Step (Part 1)**

   Draw $\boldsymbol{\beta}$ from $(p+1)$-variate normal. Draw $y_t^*$, $t = 1, \ldots, T$, from truncated univariate normals. Draw $c_k$, $k = 2, \ldots, K-1$, from Unif$(l_k, r_k)$. Get $\boldsymbol{\beta}_{cur}, \boldsymbol{y}_{cur}^*, \boldsymbol{c}_{cur}$ as current values.

2. **GM-Step**

   Draw $\gamma^2$ from the (truncated) $\Gamma(a, b)$ distribution with $a$ and $b$ defined in (4.4), respectively, and update $\boldsymbol{\beta}^*_{cur}$, $\boldsymbol{y}^*_{cur}$, $\boldsymbol{c}_{cur}$ by multiplication with the group element $\gamma = \sqrt{\gamma^2}$,

$$\boldsymbol{\beta}_{new} \leftarrow \gamma\boldsymbol{\beta}_{cur}, \qquad \boldsymbol{y}^*_{new} \leftarrow \gamma\boldsymbol{y}^*_{cur}, \qquad \boldsymbol{c}_{new} \leftarrow \gamma\boldsymbol{c}_{cur}.$$

3. **MCMC-Step (Part 2)**

   Compute $\tilde{y}^*_t = \log(y^*_t - \boldsymbol{x}'_t\boldsymbol{\beta})^2 + \log\lambda^*_t$. Draw $s_t$ proportional to $\Pr(s_t)N(\tilde{y}^*_t|h^*_t + m_{s_t}, v^2_{s_t})$, for $t = 1, \ldots, T$,. Draw $(\boldsymbol{\alpha}, \phi, \sigma)$ via Metropolis-Hastings step; use ML-estimates of $(\boldsymbol{\alpha}, \phi, \sigma)$ to find an adequate multivariate normal proposal. Draw $\boldsymbol{h}^*$ in one block using the simulation smoother of De Jong and Shephard (1995).

4. **MCMC-Step (Part 3)**

   Draw $\nu$ by a Metropolis-Hastings step; use an ML-estimate of $\nu$ to find an adequate truncated Gamma proposal. Draw $\lambda^*_t$, $t = 1, \ldots, T$ from Gamma distributions.

We note that one can use also modified versions of this GM-MGMC sampler, since not all parameters need to be updated in each iteration. For example, the parameter $\nu$ is used only for modeling the tail-behavior of the error distribution for the latent variables $y^*_t$. Therefore one can omit the update of $\nu$ until the other chains have moved away from the starting values towards the area around the true values. Since $\nu$ remains unchanged under the GM-step one can use the same GM-step as in the original sampler.

## 4.2 Simulation study

Here we investigate the accuracy of the posterior mean estimates for the parameter $\nu$ in the OSVt model. We do this by two simulation settings where the simulation parameters for $\phi$, $\sigma$, $\beta_0$, $\beta_1$, $\alpha_1$, $\alpha_2$, and $c_2, \ldots, c_6$ are identical to that chosen for the illustration in Section 2.3. Also the used covariates from the IBM data set are the same as well as the prior distributions. For $\nu$ we use in both settings the starting value 10. In the first simulation setting we choose $\nu = 15$, in the second $\nu = 100$. We simulate 20 data sets for both parameter settings, each of length $T = 22000$. We compute the posterior mean estimates by running the GM-MGMC sampler for 4000 iterations each, discarding the first 1000 for burn-in.

Table 5 gives the means and standard deviations of the posterior mean estimates across the 20 samples

for the Settings 1 and 2, respectively. In Setting 1, where we chose the value 15 for $\nu$, the mean of the posterior mean estimates is about 14.9 with standard deviation 2.4. Hence $\nu$ was always estimated quite well, which is also true for the other parameters. In Setting 2, where the true value for $\nu$ was 100, the mean of the posterior mean estimates is 98.3. Therefore the GM-MGMC sampler estimates $\nu$ well on average. However, the standard deviation of about 8.3 is quite large. This may be a consequence of the fact that the t-distribution becomes more and more similar to the normal distribution when the degrees of freedom increase. Therefore one needs much more data to be able to distinguish clearly between t-distributions with high degrees of freedom. The other parameters are all estimated quite well again. We mention that for Setting 1 the chain for $\nu$ moves around more slowly than for Setting 2, however, fast enough to move several times around the whole support of the posterior distribution within 3000 iterations.

### 4.3 Application to IBM data

Now we answer the question whether our IBM high-frequency data set in fact requires modeling with the heavier tailed t-distributed errors. For this we run the GM-MGMC sampler for the OSVt model for 4000 iterations and discard again the first 1000 for burn-in. From the simulations in Section 4.2 we know that this leads to quite accurate estimates. The results are summarized in Table 6.

It shows the posterior mean estimates together with their corresponding estimated standard deviations and 90% credible intervals for all parameters of interest. Comparing these values to the results for the OSV model in Table 4, we see that the posterior mean estimates for the OSV model are nearly identical to the posterior mean estimates in the OSVt model. The posterior mean estimate for the additional parameter $\nu$ is about 107. Since a t-distribution with 107 degrees of freedom is already quite close to a normal distribution, we conclude that the usage of t-distributed errors is not really necessary for our IBM data. Therefore we prefer the OSV model. Figure 5 shows an histogram of the estimates for the parameter $\nu$ in iterations 1001 to 4000. It suggests that the marginal posterior for $\nu$ is unimodal, but not symmetric, since the chosen prior does not allow for values greater than 127. Furthermore, Figure 5 shows the estimated autocorrelations in the $\nu$-chain after iteration 1000. They decline very fast, therefore no subsampling is required to estimate the standard error of $\nu$.

# 5 SUMMARY AND DISCUSSION

In this paper we introduced two stochastic volatility models for time series with an ordinal response where covariates are involved on two different levels of the process. Whereas in the first model the errors in the latent variable equation are assumed to be normal, the second model allows here for the heavier-tailed Student-t distribution. We developed a GM-MGMC sampler, using a scale transformation group, whose elements operate on the random samples of a certain conditional distribution. This GM-MGMC sampler shows a very satisfying behavior. We used the GM-MGMC sampler to detect and to quantify significant covariates for the price changes of the IBM stock. The logarithms of the elapsed time between two subsequent transactions and of the transaction volume are important covariates for the volatility process. The results confirm theoretical results for the price change process, for example results by Diamond and Verrecchia (1987), Easley and O'Hara (1987), and Tauchen and Pitts (1983). The lagged observation plays a role for the means of the price changes. Further we investigated whether the price changes in the IBM data set require modeling using the heavy-tailed Student-t distribution. However, the analysis showed that at least for this data set the use of a normal distribution is sufficient. The data analysis can be extended in the direction that other covariates are investigated in addition. In particular, one can look for temporal effects, e.g. indicators for opening and closing periods, or some lagged covariates, e.g. the transaction volume of several transactions before. Another extension of this work could be to combine the OSV model with an ACD model for the durations in a similar way as it was done in Russell and Engle (2005) for the ACM and the ACD model.

# ACKNOWLEDGEMENT

This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 386 *Statistical Analysis of Discrete Structures*.

# APPENDIX: GM-MGMC UPDATES

## A.1 Regression parameter, latent variable and cutpoint parameter update

Given $\boldsymbol{\beta}$ and $h_t^*$, the latent variable $y_t^*$ is normally distributed with mean $\boldsymbol{x}_t'\boldsymbol{\beta}$ and variance $\exp(h_t^*)$. Since $f(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{c},\boldsymbol{y}^*,\boldsymbol{h}^*,\boldsymbol{\alpha},\phi,\sigma) \propto \prod_{t=1}^T f(y_t^*|\boldsymbol{\beta},h_t^*)\cdot\pi(\boldsymbol{\beta})$ and since $\boldsymbol{\beta}$ has the prior $N_{p+1}(\boldsymbol{b}_0,B_0)$ we conclude that $\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{c},\boldsymbol{y}^*,\boldsymbol{h}^*,\boldsymbol{\alpha},\phi,\sigma \sim N_{p+1}(\boldsymbol{b},B)$ where

$$B := \left(\sum_{t=1}^T \frac{\boldsymbol{x}_t\boldsymbol{x}_t'}{\exp(h_t^*)} + B_0^{-1}\right)^{-1} \qquad \text{and} \qquad \boldsymbol{b} := B\left(\sum_{t=1}^T \frac{\boldsymbol{x}_t y_t^*}{\exp(h_t^*)} + B_0^{-1}\boldsymbol{b}_0\right).$$

Furthermore, it can be seen directly that the latent variable $y_t^*,\ t=1,\ldots,T$, must be drawn from the univariate truncated normal distribution $N_{\left[c_{y_t-1},c_{y_t}\right)}\left(y_t^* \,|\, \boldsymbol{x}_t'\boldsymbol{\beta},\exp(h_t^*)\right)$. Since, for $k\in\{2,\ldots,K-1\}$,

$$f(c_k|\boldsymbol{y},\boldsymbol{c}_{-k},\boldsymbol{y}^*,\boldsymbol{\beta},\boldsymbol{h}^*,\boldsymbol{\alpha},\phi,\sigma) \propto \left[\prod_{t=1}^T f(y_t|y_t^*,\boldsymbol{c})\right]\pi(\boldsymbol{c}) \propto \left[\prod_{t=1}^T \mathbf{1}_{\left[c_{y_t-1},c_{y_t}\right)}(y_t^*)\right]\mathbf{1}_{\{0<c_2<\ldots<c_{K-1}<C\}},$$

we conclude that the full conditional for $c_k$ is uniform on the interval $(l_k,r_k)$, where

$$l_k := \max\left\{c_{k-1}, \max_{t=1,\ldots,T}\{y_t^*|y_t=k\}\right\}, \tag{A.1}$$

$$r_k := \min\left\{c_{k+1}, \min_{t=1,\ldots,T}\{y_t^*|y_t=k+1\}\right\}. \tag{A.2}$$

## A.2 Mixture index update, $(\boldsymbol{\alpha},\phi,\sigma)$ joint update, and log-volatility update

Here we compute first $\tilde{y}_t^* = \log(y_t^* - \boldsymbol{x}_t'\boldsymbol{\beta})^2$, $t=1,\ldots,T$, and use the state space approximation (2.6) and (2.7) of the latent process. Since the mixture indices $\{s_t,\ t=1,\ldots,T\}$ are conditionally independent, we have on the one hand

$$f(s_t|\tilde{\boldsymbol{y}}^*,\boldsymbol{h}^*,\boldsymbol{s}_{-t},\boldsymbol{\alpha},\phi,\sigma) = f(s_t|\tilde{y}_t^*,h_t^*) \propto f(s_t,\tilde{y}_t^*,h_t^*) = f(\tilde{y}_t^*|h_t^*,s_t)\pi(s_t).$$

On the other hand, given $h_t^*$ and $s_t$, $\tilde{y}_t^*$ is normally distributed with mean $h_t^*+m_{s_t}$ and variance $v_{s_t}^2$. We conclude that $f(s_t|\tilde{\boldsymbol{y}}^*,\boldsymbol{h}^*,\boldsymbol{s}_{-t},\boldsymbol{\alpha},\phi,\sigma) \propto N(\tilde{y}_t^*|h_t^*+m_{s_t},v_{s_t}^2)\,\pi(s_t)$. For the update of $s_t$ we therefore first evaluate the seven densities $N(h_t^*+m_i,v_i^2)$, $i=1,\ldots,7$, each at the point $\tilde{y}_t^*$, resulting in the values $r_{t,i} := (2\pi v_i^2)^{-1/2}\exp\{-(\tilde{y}_t^*-h_t^*-m_i)^2/(2v_i^2)\}$ and then draw $s_t\in\{1,\ldots,7\}$ according to the probabilities

$$f(s_t=i\,|\,\tilde{\boldsymbol{y}}^*,\boldsymbol{h}^*,\boldsymbol{s}_{-t},\boldsymbol{\alpha},\phi,\sigma) = \frac{r_{t,i}\,q_i}{\sum_{k=1}^7 r_{t,k}\,q_k}, \qquad i=1,\ldots,7.$$

Now define $\boldsymbol{\theta} := (\boldsymbol{\alpha}',\phi,\sigma)'$. Since $f(\boldsymbol{h}^*,\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*,\boldsymbol{s}) = f(\boldsymbol{h}^*|\boldsymbol{\theta},\tilde{\boldsymbol{y}}^*,\boldsymbol{s})f(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*,\boldsymbol{s})$ we will draw a sample from $f(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*,\boldsymbol{s})$ and then use this sample for a block update of $\boldsymbol{h}^*$.

First we consider the sampling from $f(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s})$ which is done by a Metropolis-Hastings (MH) step. Let $\boldsymbol{\theta}^\bullet$ denote the current value of $\boldsymbol{\theta}$. The MH step requires the specification of an appropriate proposal density $q(\,\cdot\,)$ for $\boldsymbol{\theta}$, to sample a proposal $\boldsymbol{\theta}^\circ$ from this proposal density, the evaluation of the target density $f(\,\cdot\,|\tilde{\boldsymbol{y}}^*, \boldsymbol{s})$ and of the proposal density $q(\,\cdot\,)$ both at $\boldsymbol{\theta}^\circ$ and $\boldsymbol{\theta}^\bullet$ (at least up to a normalizing constant), finally accepting the proposal $\boldsymbol{\theta}^\circ$ with probability

$$\alpha(\boldsymbol{\theta}^\bullet, \boldsymbol{\theta}^\circ|\tilde{\boldsymbol{y}}^*, \boldsymbol{s}) = \min\left\{\frac{f(\boldsymbol{\theta}^\circ|\tilde{\boldsymbol{y}}^*, \boldsymbol{s})}{f(\boldsymbol{\theta}^\bullet|\tilde{\boldsymbol{y}}^*, \boldsymbol{s})}\frac{q(\boldsymbol{\theta}^\bullet)}{q(\boldsymbol{\theta}^\circ)}, 1\right\}. \tag{A.3}$$

First we show how $f(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s})$ can be evaluated up to a normalizing constant. Since

$$f(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s}) = d \cdot f(\tilde{y}_1^*|\boldsymbol{\theta}, \boldsymbol{s})\left[\prod_{t=2}^{T} f(\tilde{y}_t^*|\tilde{y}_1^*, \ldots, \tilde{y}_{t-1}^*, \boldsymbol{\theta}, \boldsymbol{s})\right]\pi(\boldsymbol{\theta}), \tag{A.4}$$

where $d \in \mathbb{R}$ is some unknown normalizing constant independent of $\boldsymbol{\theta}$, we can derive that $\log f(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s}) - \log d = \log g(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s}) + \log \pi(\boldsymbol{\theta})$, where

$$g(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s}) := f(\tilde{y}_1^*|\boldsymbol{\theta}, \boldsymbol{s})\left[\prod_{t=2}^{T} f(\tilde{y}_t^*|\tilde{y}_1^*, \ldots, \tilde{y}_{t-1}^*, \boldsymbol{\theta}, \boldsymbol{s})\right]. \tag{A.5}$$

To evaluate $g(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s})$ we apply the prediction error decomposition

$$\log g(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s}) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log f_{t|t-1} - \frac{1}{2}\sum_{t=1}^{T}\frac{e_t^2}{f_{t|t-1}}. \tag{A.6}$$

The values $e_t$ and $f_{t|t-1}$ are given by the Kalman recursions. Recall, that since $\pi(h_0^*) = \mathbb{1}_{\{h_0^*=\mu\}}$ and since $\boldsymbol{z}_0 = (0, \ldots, 0)'$ (cf. Equation (2.4)), $h_1^*$ given $h_0^*$ is normally distributed with mean $h_{1|0} := \mu + \boldsymbol{z}_1'\boldsymbol{\alpha}$ and variance $p_{1|0} := \sigma$. These quantities serve as initial values for the following recursions:

$$\begin{aligned}
y_{t|t-1} &:= m_{s_t} + h_{t|t-1}, & e_t &:= \tilde{y}_t^* - y_{t|t-1},\\
f_{t|t-1} &:= p_{t|t-1} + v_{s_t}^2, & n_t &:= p_{t|t-1}f_{t|t-1}^{-1},\\
h_{t|t} &:= h_{t|t-1} + n_t e_t, & p_{t|t} &:= (1-n_t)p_{t|t-1},\\
h_{t+1|t} &:= \mu + \boldsymbol{z}_{t+1}'\boldsymbol{\alpha} + \phi(h_{t|t} - \mu - \boldsymbol{z}_t'\boldsymbol{\alpha}), & p_{t+1|t} &:= \phi^2 p_{t|t} + \sigma^2.
\end{aligned} \tag{A.7}$$

Therefore, the target density can be evaluated at $\boldsymbol{\theta}^\circ$ and $\boldsymbol{\theta}^\bullet$ up to the normalizing constant $d$, which, of course, cancels down in the first fraction of expression (A.3).

As proposal density we take a $(q+2)$-dimensional normal distribution for $\boldsymbol{\theta}$ where the mean is determined by the maximum likelihood estimator $\boldsymbol{m}$ of the target density $f(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s})$. As covariance matrix $W$ we take the negative inverse of the Hessian matrix $V$ of $[\log g(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s}) + \log \pi(\boldsymbol{\theta})]$ at $\boldsymbol{m}$:

$$\boldsymbol{m} := \arg\max_{\boldsymbol{\theta}}\left[\log g(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s}) + \log \pi(\boldsymbol{\theta})\right], \qquad W := -V^{-1}.$$

$\boldsymbol{m}$ is found by numerical minimization of $-\left[\log g(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s}) + \log \pi(\boldsymbol{\theta})\right]$.

We now summarize the Metropolis-Hastings step to draw a sample from $f(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s})$:

1. Find the maximum likelihood estimator $\boldsymbol{m}$ of $[\log g(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s}) + \log \pi(\boldsymbol{\theta})]$ using the Nelder-Mead algorithm where $\log g(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}^*, \boldsymbol{s})$ is evaluated running the Kalman recursions (A.7) and applying the prediction error decomposition (A.6).

2. Approximate the Hessian matrix $V$ at $\boldsymbol{m}$ and calculate $W = -V^{-1}$.

3. Draw a proposal $\boldsymbol{\theta}^\circ$ from the $(q+2)$-dimensional normal distribution $q(\boldsymbol{\theta}) = N_{q+2}(\boldsymbol{\theta}|\boldsymbol{m}, W)$.

4. Accept the proposal $\boldsymbol{\theta}^\circ$ with probability

$$\alpha(\boldsymbol{\theta}^\bullet, \boldsymbol{\theta}^\circ|\tilde{\boldsymbol{y}}^*, \boldsymbol{s}) = \min\left\{\frac{g(\boldsymbol{\theta}^\circ|\tilde{\boldsymbol{y}}^*, \boldsymbol{s})}{g(\boldsymbol{\theta}^\bullet|\tilde{\boldsymbol{y}}^*, \boldsymbol{s})}\frac{\pi(\boldsymbol{\theta}^\circ)}{\pi(\boldsymbol{\theta}^\bullet)}\frac{q(\boldsymbol{\theta}^\bullet)}{q(\boldsymbol{\theta}^\circ)}, 1\right\}.$$

   If $\boldsymbol{\theta}^\circ$ is rejected, retain $\boldsymbol{\theta}^\bullet$ as the next draw.

Finally we consider the block update of $\boldsymbol{h}$ from $f(\boldsymbol{h}|\boldsymbol{\theta}, \tilde{\boldsymbol{y}}^*, \boldsymbol{s})$. Let $\boldsymbol{\theta}^\bullet = (\boldsymbol{\alpha}^{\bullet\prime}, \phi^\bullet, \sigma^\bullet)'$ denote the sample from the MH step before. Sampling can be done using the simulation smoother of De Jong and Shephard (1995). It requires running the Kalman recursions (A.7) with $\boldsymbol{\alpha} = \boldsymbol{\alpha}^\bullet$, $\phi = \phi^\bullet$, and $\sigma = \sigma^\bullet$, storing $e_t$, $f_{t|t-1}$ and $n_t$ for each $t = 1, \ldots, T$, and finally running the following backward recursions for $t = T, \ldots, 1$. Initially set $r_T = 0$ and $m_T = 0$.

$$\begin{aligned}
d_t &:= f_{t|t-1}^{-1} + \phi^2 n_t^2 m_t, & b_t &:= f_{t|t-1}^{-1} e_t - \phi n_t r_t,\\
q_t &:= v_{s_t}^2 - v_{s_t}^4 d_t, & \kappa_t &\sim N(0, q_t),\\
a_t &:= v_{s_t}^2 (d_t - \phi^2 n_t m_t), & r_{t-1} &:= f_{t|t-1}^{-1} e_t + (\phi - \phi n_t) r_t - a_t q_t^{-1} \kappa_t,\\
m_{t-1} &:= f_{t|t-1}^{-1} + (\phi - \phi n_t)^2 m_t + a_t^2 q_t^{-1}, & \xi_t &:= \tilde{y}_t^* - m_{s_t} - v_{s_t}^2 b_t - \kappa_t.
\end{aligned}$$

Now the vector $(\xi_1, \ldots, \xi_T)$ can be considered as a sample from the distribution $f(h_1^*, \ldots, h_T^*|\boldsymbol{\theta}, \tilde{\boldsymbol{y}}^*, \boldsymbol{s})$.

# References


Barndorff-Nielsen, O. E. and N. Shephard (2002). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society B 63*, 167–241.

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

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

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.

De Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–350.

Diamond, D. W. and R. E. Verrecchia (1987). Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics 18*, 277–311.

Easley, D. and M. O'Hara (1987). Price, trade size, and information in security markets. *Journal of Financial Economics 19*, 113–138.

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

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.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

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

Müller, G. and C. Czado (2005). An Autoregressive Ordered Probit Model with Application to High-Frequency Finance. *Journal of Computational and Graphical Statistics 14*(2), 320–338.

Russell, J. R. and R. F. Engle (2005). A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times: The Autoregressive Conditional Multinomial-Autoregressive Conditional Duration Model. *Journal of Business and Economic Statistics 23*(2), 166–180.

Rydberg, T. H. and N. Shephard (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics 1*, 2–25.

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.

Tauchen, G. E. and M. Pitts (1983). The price variability-volume relationship on speculative markets. *Econometrica 51*, 485–505.

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

| price diff. ($) | $\leq$ -3/16 | -2/16 | -1/16 | 0 | 1/16 | 2/16 | $\geq$ 3/16 |
|---|---|---|---|---|---|---|---|
| response $y_t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| frequency | 151 | 1053 | 4886 | 10333 | 5222 | 860 | 184 |

Table 1: Price differences and corresponding response categories together with observed frequencies.

| | TIMEDIFF (seconds) | $z_{\cdot 1}$ | SIZE (stocks) | $z_{\cdot 2}$ |
|---|---|---|---|---|
| min | 0 | -1.9563 | 100 | -2.1339 |
| avg | 8.3357 | 0.0000 | 2331 | 0.0000 |
| max | 116 | 2.8059 | 180000 | 5.3616 |

Table 2: Minimum, average, maximum for the two original covariates and their transformations $z_{\cdot 1}$, $z_{\cdot 2}$.

| Parameter | Prior distribution |
|---|---|
| $\boldsymbol{c}$ | Uniform on $\{(c_2, \ldots, c_6) \,\vert\, 0 < c_2 < \ldots < c_6 < \infty\}$ |
| $\boldsymbol{\beta}$ | $N_2(\mathbf{0}, \mathrm{diag}(10, 10))$ |
| $h_0^*$ | Dirac$(-0.6)$ |
| $\alpha_1$ | Uniform on $\{-100 < \alpha_1 < 100\}$ |
| $\alpha_2$ | Uniform on $\{-100 < \alpha_2 < 100\}$ |
| $\phi$ | Uniform on $\{-1 < \phi < 1\}$ |
| $\sigma$ | Uniform on $\{0 < \sigma < 10\}$ |

Table 3: Prior distributions of parameters.

| | estimate | std.err. | 90% cred.int. | | estimate | std.err. | 90% cred.int. |
|---|---|---|---|---|---|---|---|
| $\phi$ | 0.9061 | 0.0119 | (0.8853,0.9248) | $c_2$ | 0.9332 | 0.0137 | (0.9102,0.9554) |
| $\sigma$ | 0.2230 | 0.0194 | (0.1922,0.2570) | $c_3$ | 1.8248 | 0.0208 | (1.7894,1.8587) |
| $\beta_0$ | 3.5152 | 0.0463 | (3.4402,3.5908) | $c_4$ | 2.7609 | 0.0310 | (2.7087,2.8113) |
| $\beta_1$ | -0.3073 | 0.0070 | (-0.3188,-0.2962) | $c_5$ | 3.6893 | 0.0443 | (3.6140,3.7609) |
| $\alpha_1$ | 0.2599 | 0.0182 | (0.2298,0.2912) | $c_6$ | 4.5562 | 0.0636 | (4.4500,4.6623) |
| $\alpha_2$ | 0.1511 | 0.0090 | (0.1363,0.1665) | | | | |

Table 4: Posterior mean estimates and corresponding estimated standard deviations and 90% posterior credible intervals for parameters in OSV model.

| | true | mean | std. dev. | | true | mean | std. dev. |
|---|---|---|---|---|---|---|---|
| $\phi$ | 0.90 | 0.8919 | 0.0171 | $c_2$ | 0.90 | 0.9129 | 0.0178 |
| $\sigma$ | 0.20 | 0.2152 | 0.0304 | $c_3$ | 1.80 | 1.8169 | 0.0360 |
| $\beta_0$ | 3.50 | 3.5181 | 0.0391 | $c_4$ | 2.75 | 2.7788 | 0.0343 |
| $\beta_1$ | -0.30 | -0.3010 | 0.0069 | $c_5$ | 3.65 | 3.6740 | 0.0468 |
| $\alpha_1$ | 0.25 | 0.2540 | 0.0138 | $c_6$ | 4.50 | 4.5303 | 0.0552 |
| $\alpha_2$ | 0.15 | 0.1549 | 0.0100 | $\nu$ | **15** | **14.8781** | **2.4203** |
| $\phi$ | 0.90 | 0.8928 | 0.0164 | $c_2$ | 0.90 | 0.9167 | 0.0192 |
| $\sigma$ | 0.20 | 0.2131 | 0.0297 | $c_3$ | 1.80 | 1.8276 | 0.0384 |
| $\beta_0$ | 3.50 | 3.5214 | 0.0524 | $c_4$ | 2.75 | 2.7741 | 0.0395 |
| $\beta_1$ | -0.30 | -0.3022 | 0.0075 | $c_5$ | 3.65 | 3.6735 | 0.0525 |
| $\alpha_1$ | 0.25 | 0.2517 | 0.0136 | $c_6$ | 4.50 | 4.5405 | 0.0670 |
| $\alpha_2$ | 0.15 | 0.1545 | 0.0098 | $\nu$ | **100** | **98.2844** | **8.3142** |

Table 5: Means and standard deviations of posterior mean estimates across the 20 samples in Setting 1 (above) and Setting 2 (below).

| | estimate | std.err. | 90% cred.int. | | estimate | std.err. | 90% cred.int. |
|---|---|---|---|---|---|---|---|
| $\phi$ | 0.9133 | 0.0087 | (0.8984,0.9275) | $c_2$ | 0.9489 | 0.0095 | (0.9333,0.9641) |
| $\sigma$ | 0.2066 | 0.0136 | (0.1851,0.2294) | $c_3$ | 1.8548 | 0.0163 | (1.8280,1.8812) |
| $\beta_0$ | 3.5388 | 0.0385 | (3.4749,3.6016) | $c_4$ | 2.7794 | 0.0243 | (2.7381,2.8183) |
| $\beta_1$ | -0.3082 | 0.0064 | (-0.3181,-0.2976) | $c_5$ | 3.7217 | 0.0329 | (3.6679,3.7759) |
| $\alpha_1$ | 0.2584 | 0.0172 | (0.2301,0.2861) | $c_6$ | 4.6116 | 0.0476 | (4.5351,4.6868) |
| $\alpha_2$ | 0.1482 | 0.0078 | (0.1362,0.1615) | $\nu$ | **106.81** | **15.4530** | **(77,126)** |

Table 6: Posterior mean estimates and corresponding estimated standard deviations and 90% posterior credible intervals for parameters in OSVt model.

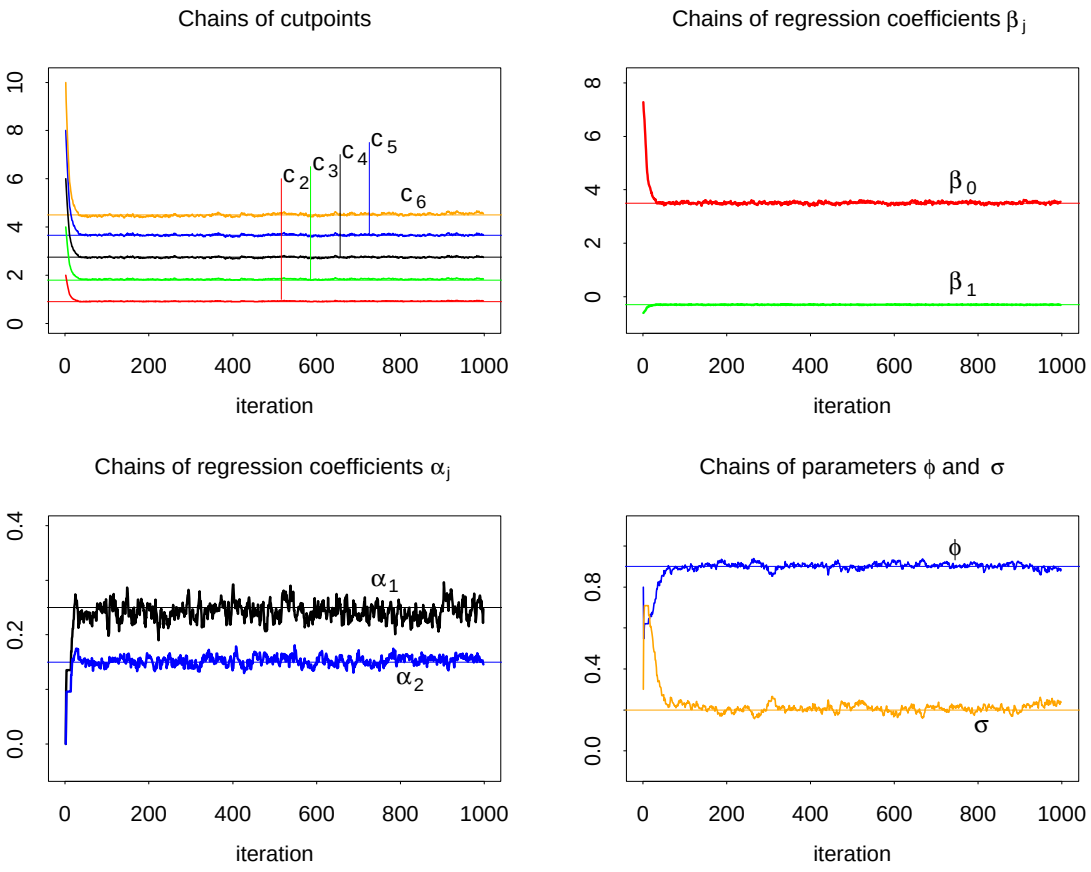


Figure 1: First 1000 iterations of chains produced by the GM-MGMC sampler. The horizontal thin lines indicate the true values.

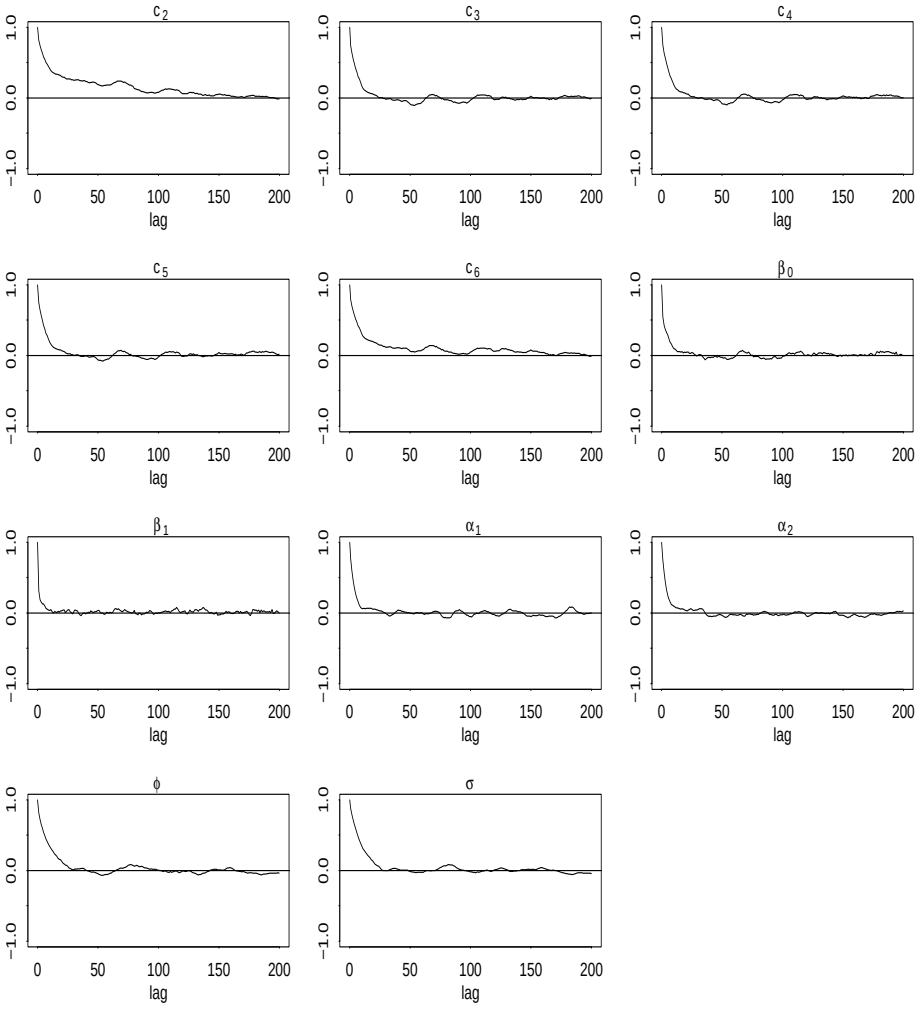


Figure 2: Autocorrelations of chains produced by the GM-MGMC sampler.

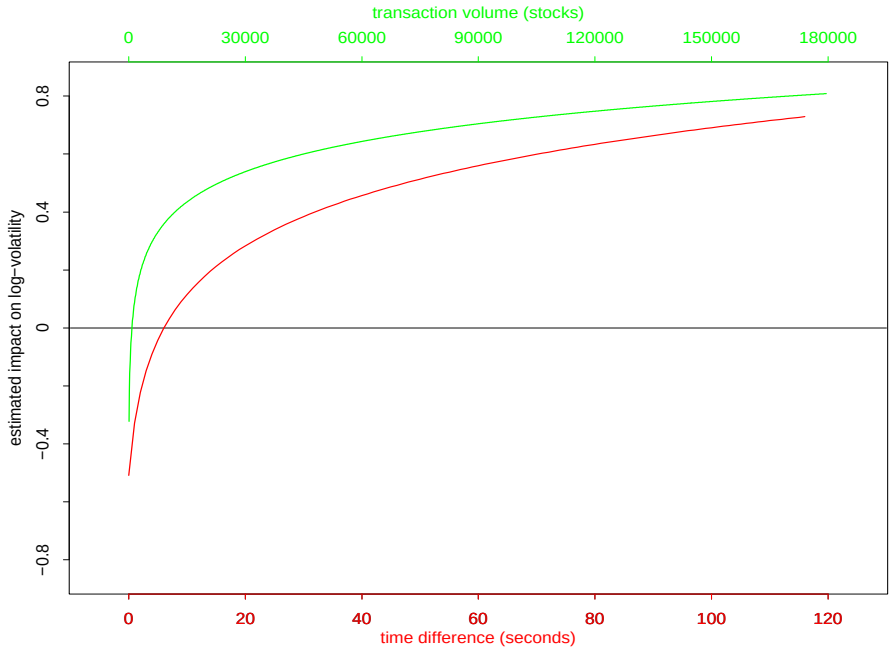


Figure 3: Estimated impacts of TIMEDIFF and SIZE on the log-volatilities $h_t^*$. The estimated joint impact of both covariates is given by adding these individual impacts.

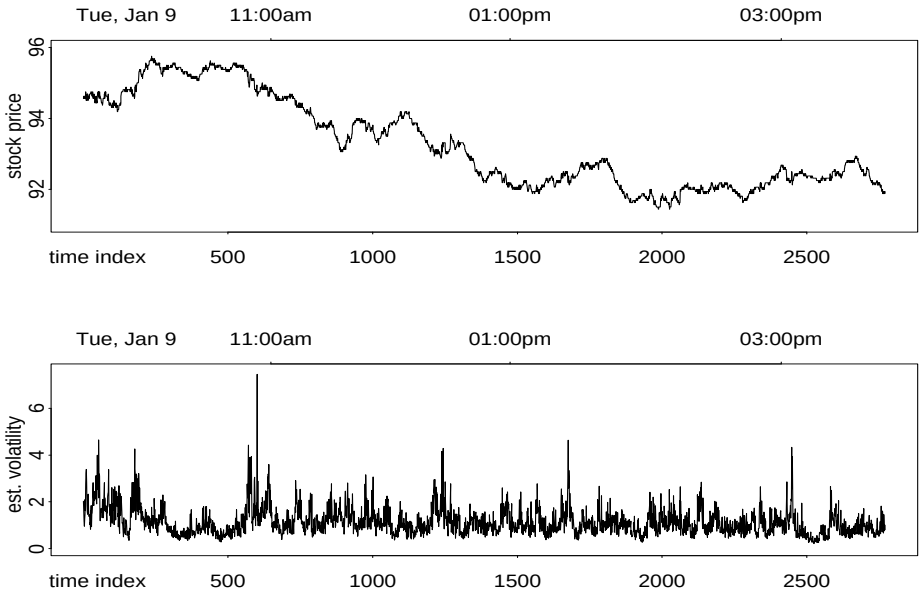


Figure 4: IBM stock prices and estimated normalized volatilities on January 9, 2001.

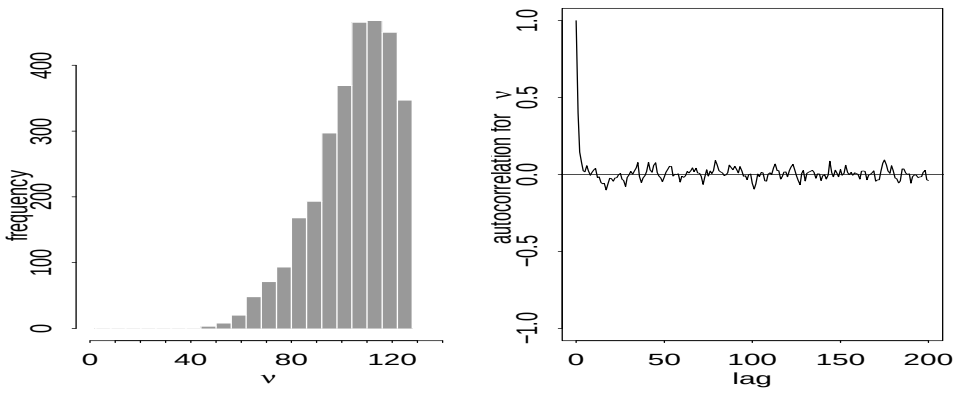


Figure 5: Histogramm for estimates of $\nu$ for iterations 1001 to 4000 (left). Autocorrelation of chain for $\nu$ after iteration 1000 (right).