An Autoregressive Ordered Probit Model With Application to High Frequency Financial Data

Gernot MÜLLER[1], Claudia CZADO[2]

Abstract

In this paper we introduce a model which can be considered as an autoregressive extension of the ordered probit model. For parameter estimation we first develop a standard Gibbs sampler which, however, exhibits bad convergence properties. Utilizing a special transformation group on the sample space we develop a Grouped Move Multigrid Monte Carlo (GM-MGMC) Gibbs sampler and illustrate its fundamental superiority in convergence compared to the standard sampler. To be able to compare the autoregressive ordered probit (AOP) model to other models we further provide an estimation procedure for the marginal likelihood which enables us to compute Bayes factors. We apply the new model to absolute price changes of the IBM stock traded on Dec 4, 2000 at the NYSE. To detect whether the data contains an autoregressive structure we then fit the AOP model as well as the common ordered probit (OP) model to the data. By estimating the corresponding Bayes factor we show that the AOP model fits the data decisively better than the common OP model.

Keywords: Bayes factor; Markov chain Monte Carlo; Multigrid Monte Carlo; Transformation group;

[1]Gernot Müller is Scientific Assistant, Center for Mathematics, Munich University of Technology, Boltzmannstraße 3, 85747 Garching/Munich, Germany (mueller@ma.tum.de).

[2]Claudia Czado is Professor, Center for Mathematics, Munich University of Technology, Boltzmannstraße 3, 85747 Garching/Munich, Germany (cczado@ma.tum.de).

1. INTRODUCTION

In many applications one has to deal with observations which take on values from a finite ordered set, where the order of the elements is important and usually has some natural interpretation in the application. If one observes also covariates, one can use for example the well-known ordered probit model for a regression analysis. In the ordered probit model, however, the observations are assumed to be independent of each other, so that a time dependence which may be present in the data cannot be modelled adequately. Therefore in this paper we introduce a model, which can be viewed as an extension of the common ordered probit (OP) model since it has the features of this model but also allows for an autoregressive time structure in the data. We call this model autoregressive ordered probit (AOP) model.

For fitting the AOP model, and the OP model, we follow the general strategy of Albert and Chib (1993) and introduce latent variables for the categorical responses and then apply Markov chain sampling methods to simulate the posterior distribution of these latent variables and parameters. In our implementation of this strategy we utilize the method of Liu and Sabatti (2000) to improve the efficiency of the sampling procedure. The resulting method is similar to the steps in the Albert and Chib (1993) algorithm for the OP model except that a randomly drawn element from a transformation group is used to scale the parameters and latent variables at the end of each iteration. We show that the use of this scaling helps to improve the mixing of the MCMC output.

To be able to decide whether the AOP model can fit a given data set better than other models, we developed an estimation procedure for the marginal likelihood. This, of course, allows for the computation of Bayes factors, which can be used for model comparison. Since the estimation of the marginal likelihood also requires a filtering procedure we also provide an auxiliary particle filter for the new model. Auxiliary particle filters were introduced by Pitt and Shephard (1999). Since the OP model is a submodel of the AOP model, the algorithms can be easily simplified and applied also to the common ordered probit model. Therefore, for given data we are able to compare in particular the fit of the AOP model to the fit of the OP model.

A possible application of the AOP model is to high-frequency data from finance. For a global overview about models for high-frequency financial data see for example Bauwens and Giot (2001) or Dacorogna, Gençay, Müller, Olsen, and Pictet (2001). In modeling the price process of a stock one is faced with

the feature that transaction price changes often occur in discrete increments, for example in integer multiples of one sixteenth of a dollar. Moreover, only few price differences occur, usually from $-3/16\$$ to $+3/16\$$. Hausman, Lo, and MacKinlay (1992) applied the common ordered probit model to such data, although this model does not take into account a possibly present autoregressive structure in the latent process. Another model approach was made by Rydberg and Shephard (2003). 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. Since we agree with Rydberg and Shephard (2003) that it may be more accurate to split the price process into components and to model these components seperately we apply the AOP model only to absolute price changes of the IBM stock on December 4, 2000, at the New York Stock Exchange (NYSE). We search for significant covariates and investigate using Bayes factors whether the AOP model fits the data better than the OP model.

2. MODEL AND STANDARD GIBBS SAMPLER

In this section we first introduce the autoregressive ordered probit model. Then, for a Bayesian inference, we specify prior distributions and finally develop a 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 $\boldsymbol{x}_t = (1, x_{t1}, \ldots, x_{tp})'$ 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 autoregressive real-valued time series $\{y_t^*, t = 1, \ldots, T\}$ which produces the discrete values y_t by thresholding. In particular, the following model is assumed:

$$y_t = k \iff y_t^* \in [c_{k-1}, c_k) \qquad k \in \{1, \ldots, K\} \tag{2.1}$$

$$y_t^* = \boldsymbol{x}_t'\boldsymbol{\beta} + \phi y_{t-1}^* + \varepsilon_t^* \qquad t \in \{1, \ldots, T\} \tag{2.2}$$

where $-\infty = c_0 < c_1 < \ldots < c_{K-1} < c_K = \infty$ 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

fix $c_1 = 0$ for reasons of identifiability. For the same reasons we also fix $\delta^2 = 1$. Since the model defined by Equations (2.1) and (2.2) differs from the common ordered probit model in the expression ϕy^*_{t-1} in Equation (2.2), we call this model autoregressive ordered probit (AOP) model.

We use the notations $\boldsymbol{\theta} := (\beta_0, \ldots, \beta_p, \phi)$, $\boldsymbol{y} := (y_1, \ldots, y_T)$, $\boldsymbol{y}^*_{-t} := (y^*_0, \ldots, y^*_{t-1}, y^*_{t+1}, \ldots, y^*_T)$, and $\boldsymbol{c}_{-k} := (c_2, \ldots, c_{k-1}, c_{k+1}, \ldots, c_{K-1})$. 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 $N_n(\boldsymbol{\mu}, \Sigma)$. Further we denote by $f(\cdot)$ and $f(\cdot|\cdot)$ distributions (densities) and conditional distributions (densities), respectively. For prior and posterior distributions we also use the notations $\pi(\cdot)$ and $\pi(\cdot|\cdot)$, respectively.

2.2 BAYESIAN INFERENCE USING A STANDARD GIBBS SAMPLER

Now we develop an MCMC algorithm that allows us to draw approximate samples from the posterior distribution $\pi(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{y}^*|\boldsymbol{y})$. For the Bayesian approach we have to specify prior distributions for $\boldsymbol{c}$, $\boldsymbol{\theta}$ and y^*_0. We choose

$$\pi(y^*_0, \boldsymbol{\theta}, \boldsymbol{c}) \quad \propto \quad \exp\left\{-\frac{1}{2}\left[\sigma^{-2}(y^*_0)^2 + \tau^{-2}\boldsymbol{\beta}'\boldsymbol{\beta} + \rho^{-2}\phi^2\right]\right\} \cdot \mathbb{1}_{\{0<c_2<\ldots<c_{K-1}<C\}} \tag{2.3}$$

where σ, τ, ρ and C are known hyperparameters. In particular, all parameters are a priori independent, except the vector $\boldsymbol{c}$ for which an order condition has to be fulfilled. We choose normal priors for y^*_0, $\boldsymbol{\beta}$ and ϕ, respectively, and a noninformative prior on the set $\{0 < c_2 < \ldots < c_{K-1} < C\}$ for the cutpoints. We can take large values for σ, τ and ρ, when there is little prior information about y^*_0 and $\boldsymbol{\theta}$. For notational convenience we now redefine $c_K := C$. Since the regression parameters and the autoregressive parameter can be updated in one block only the updates for the latent variables differ from the updates for the OP model considered in Albert and Chib (1993).

Latent variable update: Up to proportionality the joint density of the latent variables decomposes into

$$f(\boldsymbol{y}^*|\boldsymbol{y}, \boldsymbol{\beta}, \phi, \boldsymbol{c}) \propto f(\boldsymbol{y}^*, \boldsymbol{y}, \boldsymbol{\beta}, \phi, \boldsymbol{c}) = \left[\prod_{t=1}^{T} f(y_t|y^*_t, \boldsymbol{c}) f(y^*_t|y^*_{t-1}, \boldsymbol{\beta}, \phi)\right] \pi(y^*_0, \boldsymbol{\beta}, \phi, \boldsymbol{c}) \tag{2.4}$$

Together with the $N(0, \sigma^2)$-prior for y^*_0 it follows directly that

$$f(y^*_0 \mid \boldsymbol{y}, \boldsymbol{y}^*_{-0}, \boldsymbol{c}, \boldsymbol{\theta}) \sim N\left(\frac{\phi(y^*_1 - \boldsymbol{x}'_1\boldsymbol{\beta})}{\phi^2 + \sigma^{-2}}, \frac{1}{\phi^2 + \sigma^{-2}}\right)$$

Furthermore we can easily see from Equation (2.4) that the full conditionals $f(y_t^*|\boldsymbol{y}, \boldsymbol{y}_{-t}^*, \boldsymbol{\beta}, \phi, \boldsymbol{c})$, $t = 1, \ldots, T$ are a truncated normal distributions. In particular, we have

$$f(y_t^*|\boldsymbol{y}, \boldsymbol{y}_{-t}^*, \boldsymbol{\beta}, \phi, \boldsymbol{c}) \sim N_{[c_{y_t-1}, c_{y_t})} \left(\frac{\phi(y_{t+1}^* - \boldsymbol{x}_{t+1}'\boldsymbol{\beta}) + (\boldsymbol{x}_t'\boldsymbol{\beta} + \phi y_{t-1}^*)}{1+\phi^2}, \frac{1}{1+\phi^2} \right)$$

for $t = 1, \ldots, T-1$, and $f(y_T^*|\boldsymbol{y}, \boldsymbol{y}_{-T}^*, \boldsymbol{\beta}, \phi, \boldsymbol{c}) \sim N_{[c_{y_T-1}, c_{y_T})}\big(\boldsymbol{x}_T'\boldsymbol{\beta} + \phi y_{T-1}^*, 1\big)$.

Joint regression and autoregressive parameter update: The parameters β_j, $j = 0, \ldots, p$, and ϕ are updated in one block analogously to the ordered probit model (cf. Albert and Chib (1993)). We have $f(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{c}) \sim N_{p+2}(\Sigma Z' \boldsymbol{y}_{-0}^*, \Sigma)$, a $(p+2)$-dimensional normal distribution with covariance matrix $\Sigma := (Z'Z + \mathrm{diag}(\tau^{-2}, \ldots, \tau^{-2}, \rho^{-2}))^{-1}$ and $T \times (p+2)$-matrix $Z := (z_1, \ldots, z_T)'$, $z_i := (1, x_{i1}, \ldots, x_{ip}, y_{i-1}^*)'$.
Cutpoint parameter update: As in Albert and Chib (1993) it follows that for each $k \in \{2, \ldots, K-1\}$ the full conditional $f(c_k|\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{\beta}, \phi, \boldsymbol{c}_{-k})$ is uniformly distributed in the interval (l_k, r_k), where $l_k = \max\{c_{k-1}, \max_{t=1,\ldots,T}\{y_t^*|y_t = k\}\}$ and $r_k = \min\{c_{k+1}, \min_{t=1,\ldots,T}\{y_t^*|y_t = k+1\}\}$.

3. GM-MGMC SAMPLER

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 c_j and the regression intercept β_0 (cf. Subsection 3.2). This behavior was also observed by Cowles (1996) for the independent multinomial case, which can be explained as follows. The parameter c_j is drawn from a uniform distribution with boundaries l_j and r_j given above. If the observed dataset is large (e.g. $T = 2000$), the difference $r_j - l_j$ is very small, so c_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 truncated 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 c_k which seems to be most important. Therefore, we use now a method that was proposed by Liu and Sabatti (2000).

3.1 DEVELOPMENT OF AN APPROPRIATE GROUPED-MOVE-STEP

The idea proposed by Liu and Sabatti (2000) 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 $\boldsymbol{S}$, L its left-Haar measure (as defined in Rao (1987), p. 492), $\boldsymbol{w} \in \boldsymbol{S}$ follows a distribution with density π, 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 π, too (Liu and Sabatti (2000), Theorem 1). Typical examples of such transformation groups are the *translation group on $\boldsymbol{S}$ along an arbitrary direction*, $\Gamma = \left\{\gamma \in \mathbb{R}^1 : \gamma(\boldsymbol{w}) = \boldsymbol{w} + \gamma\boldsymbol{e} = (w_1 + \gamma e_1, \ldots, w_d + \gamma e_d)\right\}$ with Lebesque measure as left-Haar measure (here γ has to be drawn from $\pi(\boldsymbol{w} + \gamma\boldsymbol{e})$), or the *scale group on $\boldsymbol{S}$*, $\Gamma = \{\gamma > 0 : \gamma(\boldsymbol{w}) = (\gamma w_1, \ldots, \gamma w_d)\}$ with $\gamma^{-1}d\gamma$ as left-Haar measure (here γ has to be drawn from $\gamma^{d-1}\pi(\gamma\boldsymbol{w})$). 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 $\boldsymbol{w}$ is a vector with $T + p + K + 1$ components, namely

$$\boldsymbol{w} = (y_0^*, \ldots, y_T^*, \beta_0, \ldots, \beta_p, c_2, \ldots, c_{K-1}, \phi)\,,$$

and π is the posterior density of $\boldsymbol{w}$. 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

$$\Gamma_m := \{\gamma > 0 : \gamma(\boldsymbol{w}) = (\gamma w_1, \ldots, \gamma w_m, w_{m+1}, \ldots, w_d)\}\,,$$

which we call a *partial scale group on $\boldsymbol{S}$*. 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 $\det(\partial\gamma(\boldsymbol{w})/\partial\boldsymbol{w}) = \gamma^m$. Therefore

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

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

$$\gamma(\boldsymbol{w}) = (\gamma y_0^*, \ldots, \gamma y_T^*, \gamma\beta_0, \ldots, \gamma\beta_p, \gamma c_2, \ldots, \gamma c_{K-1}, \phi)\,.$$

The posterior distribution in our problem is given by

$$\begin{aligned}\pi(\boldsymbol{w}|\boldsymbol{y}) &= \pi(y_0^*,\ldots,y_T^*,\boldsymbol{\beta},\boldsymbol{c},\phi|y_1,\ldots,y_T)\\ &\propto \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T}(y_t^*-\boldsymbol{\beta}'\boldsymbol{x}_t-\phi y_{t-1}^*)^2+\sigma^{-2}(y_0^*)^2+\tau^{-2}\boldsymbol{\beta}'\boldsymbol{\beta}+\rho^{-2}\phi^2\right]\right\}\prod_{t=1}^{T}1_{[c_{y_t-1},c_{y_t})}(y_t^*)\end{aligned}$$

The density $\gamma^{m-1}\pi(\gamma\boldsymbol{w}|\boldsymbol{y})$ is therefore proportional to

$$\begin{aligned}&\gamma^{m-1}\exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T}(\gamma y_t^*-\gamma\boldsymbol{\beta}'\boldsymbol{x}_t-\phi\gamma y_{t-1}^*)^2+\sigma^{-2}(\gamma y_0^*)^2+\tau^{-2}\gamma^2\boldsymbol{\beta}'\boldsymbol{\beta}+\rho^{-2}\phi^2\right]\right\}\prod_{t=1}^{T}1_{[\gamma c_{y_t-1},\gamma c_{y_t})}(\gamma y_t^*)\\ &\quad\propto\ \gamma^{m-1}\exp\left\{-\frac{1}{2}\gamma^2\left[\sum_{t=1}^{T}(y_t^*-\boldsymbol{\beta}'\boldsymbol{x}_t-\phi y_{t-1}^*)^2+\sigma^{-2}(y_0^*)^2+\tau^{-2}\boldsymbol{\beta}'\boldsymbol{\beta}\right]\right\}\prod_{t=1}^{T}1_{[c_{y_t-1},c_{y_t})}(y_t^*)\\ &\quad\propto\ \left(\gamma^2\right)^{\frac{m-1}{2}}\exp\left\{-\frac{1}{2}\gamma^2\left[\sum_{t=1}^{T}(y_t^*-\boldsymbol{\beta}'\boldsymbol{x}_t-\phi y_{t-1}^*)^2+\sigma^{-2}(y_0^*)^2+\tau^{-2}\boldsymbol{\beta}'\boldsymbol{\beta}\right]\right\}\end{aligned}$$

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

$$a=\frac{T+K+p+1}{2}\qquad\text{and}\qquad b=\frac{\sum_{t=1}^{T}(y_t^*-\boldsymbol{x}_t'\boldsymbol{\beta}-\phi y_{t-1}^*)^2+\sigma^{-2}(y_0^*)^2+\tau^{-2}\boldsymbol{\beta}'\boldsymbol{\beta}}{2}\tag{3.1}$$

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

1. **MCMC-Step**: Generate an iteration from the standard Gibbs-sampler using the latent variable update, joint regression and autoregressive parameter update, and cutpoint parameter update to get $\boldsymbol{y}_{cur}^*,\boldsymbol{\beta}_{cur},\phi_{cur},\boldsymbol{c}_{cur}$ as current values.

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

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

Note that ϕ_{cur} does not need to be updated since it remains unchanged under the partial scale group.

Running the standard Gibbs sampler for simulated data sets one can see that, in contrast to all other parameters, the chain for ϕ converges quite fast to the region around the true value. Therefore it seems not necessary to include ϕ in the grouped move step. Moreover, including ϕ would also lead to a non-standard distribution for γ or γ^2 and therefore to a higher computational cost to draw γ, without further improving the convergence.

3.2 ILLUSTRATION: STANDARD SAMPLER AGAINST GM-MGMC

We now illustrate the improvement in convergence which is achieved by adding the GM-step presented in the previous subsection to the standard Gibbs sampler. For this purpose we first simulated two covariates x_{t1} and x_{t2} independently from $N(-1, 1)$ and $N(-0.25, 0.18^2)$, respectively, for $t = 1, \ldots, 2000$. Using these covariates we then simulated one data set from the autoregressive ordered probit model with parameters $c_2 = 1.2$, $c_3 = 2.2$, $c_4 = 3.1$, $c_5 = 4.1$, $c_6 = 5.3$, $\beta_0 = 2.9$, $\beta_1 = -0.6$, $\beta_2 = 9.0$ and $\phi = 0.5$.

We run both the standard Gibbs sampler and the GM-MGMC sampler for the simulated data set for 15000 iterations. The starting values for the cutpoints $c_2, \ldots, c_6$ are $2, 4, 6, 8, 10$, respectively, and 0.0 for each of the regression coefficients.

Figures 1 and 2 demonstrate the fundamental superiority of the GM-MGMC sampler. As can be seen from Figure 1 the chains for the cutpoints produced by the standard Gibbs sampler move very slowly to the regions around the true values which are indicated by the horizontal lines. The sharp drop in the chains of c_4, c_5, and c_6 from the standard Gibbs sampler at the beginning can be explained by the fact that the samples of $\{y_t^*|y_t \in \{4, 5, 6\}\}$ lie close to the lower limit of the corresponding intervals. Hence the cutpoints c_4, c_5, and c_6 have more room to move in the first iterations.

We can guess that it will take thousands of iterations until the chains for the cutpoints produced by the standard Gibbs sampler will have converged. The same holds for the chains for the regression coefficients, given in Figure 2. The chains produced by the GM-MGMC sampler are also given in the Figures 1 and 2. They converge within about only 20 iterations for both the cutpoints and the regression coefficients. Finally we note that the chain for the autoregressive parameter ϕ converges quite fast for both samplers and is therefore not shown here. We further investigated the autocorrelations in the chains after a burnin period of 5000 iterations for both samplers. The GM-MGMC sampler is better again since the autocorrelations in the GM-MGMC chains are much smaller than in the chains from the standard Gibbs sampler.

We conducted an extensive simulation study to test the GM-MGMC sampler in several situations. We used data sets with a for our practical concerns relative high number of categories, $K = 7$. Further we took $T = 2000$. In particular, we were 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

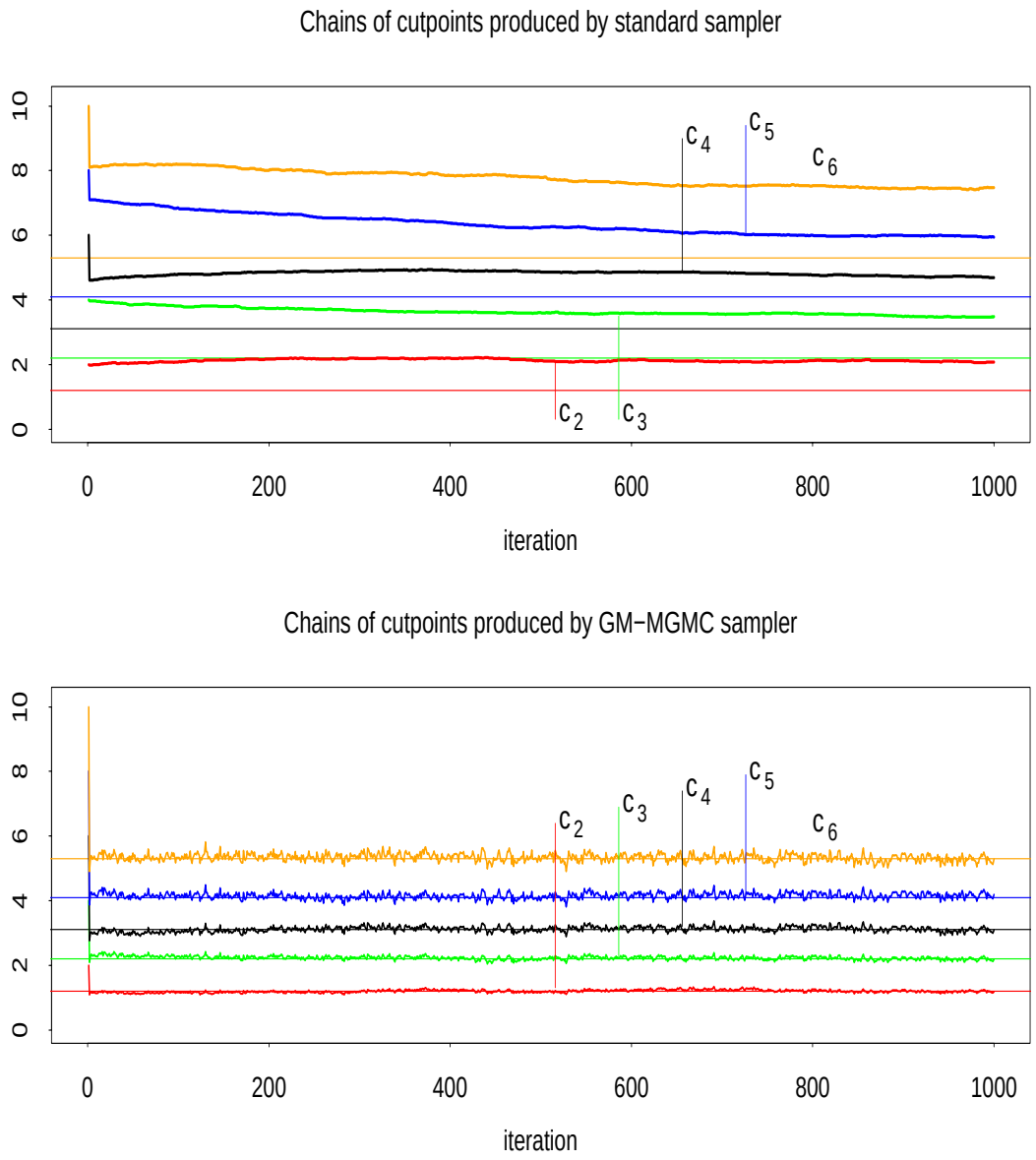

Figure 1: First 1000 iterations for the cutpoints produced by the standard Gibbs sampler (above) and the GM-MGMC sampler (below). The horizontal lines indicate the true values.

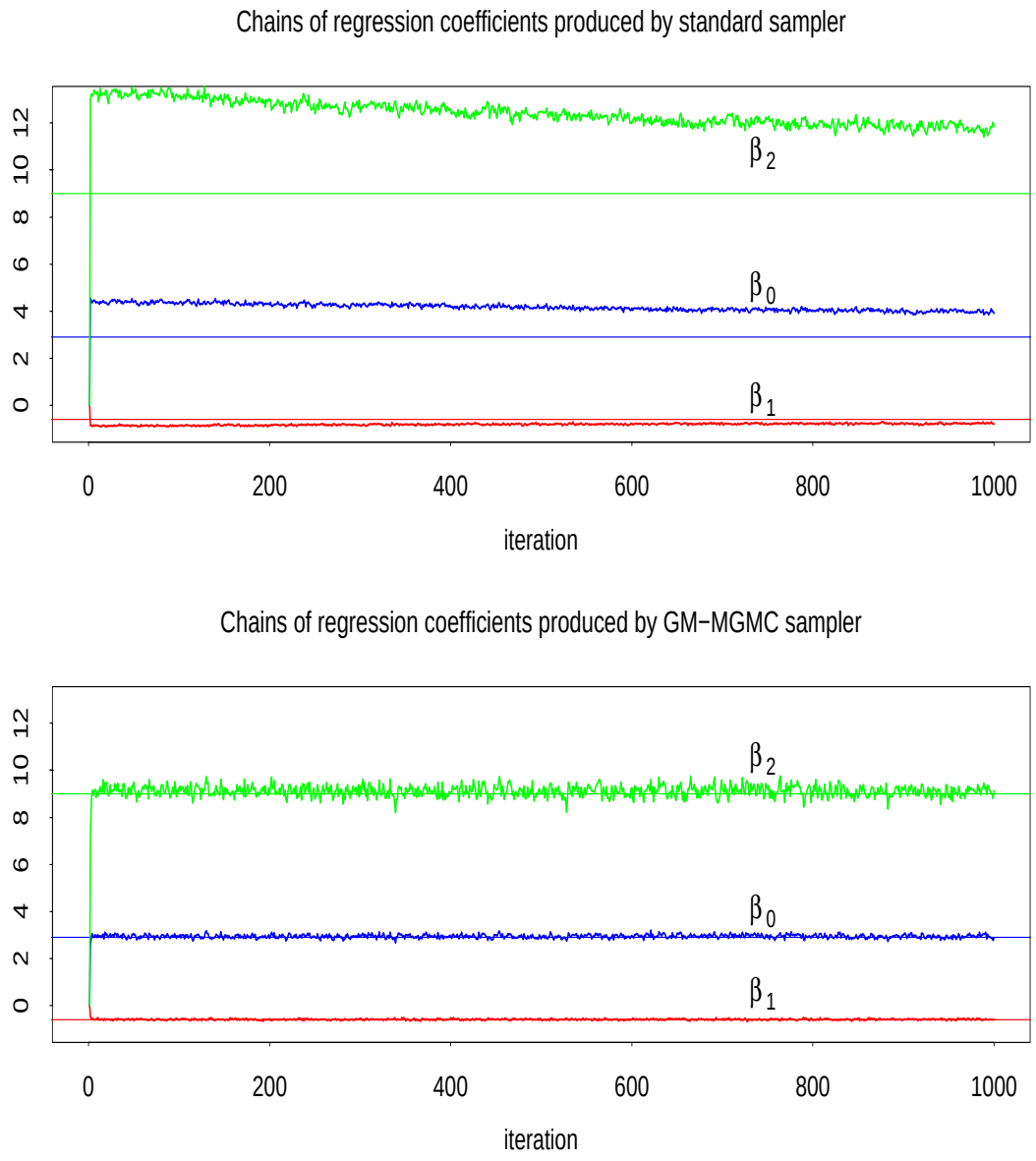

Figure 2: First 1000 iterations for the regression coefficients produced by the standard Gibbs sampler (above) and the GM-MGMC sampler (below). The horizontal lines indicate the true values.

observations lies in one or two categories and only few observations in the other categories as is the case in our application in Section 5. 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. Here we only summarize briefly the most important results. The autoregressive parameter ϕ is always estimated quite well. For the cutpoints as well as for the regression parameters the relative bias is in most cases between -1% and +1%. The relative MSE is in most cases less than 0.001. That means that on average the estimates are less than 3% away from the true values. 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.

We can 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. The autocorrelations in the chains produced by the GM-MGMC Gibbs sampler are explicitly lower than those in the chains produced by the standard Gibbs sampler.

4. ESTIMATION OF MARGINAL LIKELIHOOD

An interesting question for the application of a certain model to a given data set is always whether this model fits the data set better or worse than other ones. Since we followed a Bayesian approach we answer this question by estimating Bayes factors. The Bayes factor for comparing a model M_1 to a model M_2 is defined by the fraction of the corresponding marginal likelihoods, i.e. $m(\boldsymbol{y}|M_1)/m(\boldsymbol{y}|M_2)$. In this section we therefore provide an estimation procedure for the marginal likelihood for the autoregressive ordered probit model applying methods of Chib (1995) and Chib and Jeliazkov (2001).

We use the marginal likelihood identity in the form

$$m(\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{c}^\diamond, \boldsymbol{\theta}^\diamond)\pi(\boldsymbol{c}^\diamond, \boldsymbol{\theta}^\diamond)}{\pi(\boldsymbol{c}^\diamond, \boldsymbol{\theta}^\diamond|\boldsymbol{y})} \tag{4.1}$$

where $\boldsymbol{c}^\diamond$ and $\boldsymbol{\theta}^\diamond$ are corresponding posterior mean estimates. Note that this identity does not contain the latent variables so that the expressions are much more managable. In the following sections we first consider a filtering procedure and then the estimation and computation of the three factors which

appear on the right-hand side in Equation (4.1). $\mathcal{F}_t$ will always denote the vector of observations until time t, i.e. $\mathcal{F}_t := (y_1, \ldots, y_t)$.

4.1 FILTERING

In this section we present a fully adapted SIR-based auxiliary particle filter for the AOP model. This filtering procedure is important for the estimation of the likelihood ordinate $f(\boldsymbol{y}|\boldsymbol{c}^\diamond, \boldsymbol{\theta}^\diamond)$ and forms part of Algorithm 2. The general form of a SIR-based auxiliary particle filter is given in Pitt and Shephard (1999). To produce a sample $\{y_t^{*1}, \ldots, y_t^{*M}\}$ from $f(y_t^*|\mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta})$ when a sample $\{y_{t-1}^{*1}, \ldots, y_{t-1}^{*M}\}$ from $f(y_{t-1}^*|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta})$ is given, the following steps are necessary:

1. Make R proposals (y_t^{*r}, m^r) from $g(y_t^*, m|\mathcal{F}_t)$, where g is a proposal density.
2. For each $r \in \{1, \ldots, R\}$ evaluate the weights $w_r = f(y_t|y_t^{*r}) f(y_t^{*r}|y_{t-1}^{*m^r}) / g(y_t^{*r}, m^r|\mathcal{F}_t)$ and compute the probabilities $\pi_r = w_r / \sum_{j=1}^R w_j$.
3. Resample among $\{y_t^{*r}|r = 1, \ldots, R\}$ using the associated probabilities $\{\pi_r\}$ to produce a sample of size M from $f(y_t^*|\mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta})$.

In our case we choose

$$g(y_t^*, m|\mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta}) \;\propto\; f(y_t|y_t^*, \boldsymbol{c}, \boldsymbol{\theta}) f(y_t^*|y_{t-1}^{*m}, \boldsymbol{c}, \boldsymbol{\theta}) \;=\; \mathbb{1}_{\{y_t^* \in [c_{y_t-1}, c_{y_t})\}} N(y_t^*|\boldsymbol{x}_t'\boldsymbol{\beta} + \phi y_{t-1}^{*m}, 1)$$

as proposal density. This immediately implies that the numerator equals the denominator in Step 2 and therefore the weights in Step 2 all equal 1, i.e. the algorithm is fully adapted to the model (cf. Pitt and Shephard (1999)). Therefore the resampling Step 3 is not necessary and we just have to sample from the proposal density. Since

$$g(y_t^*, m|\mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta}) \;=\; g(y_t^*|m, \mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta}) g(m|\mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta})$$

this can be accomplished by first choosing $m \in \{1, \ldots, M\}$ with probability $g(m|\mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta})$ and then sampling from $g(y_t^*|m, \mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta})$. Since

$$g(m|\mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta}) \;= \int g(y_t^*, m|\mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta})\, dy_t^* \;\propto\; \int \mathbb{1}_{\{y_t^* \in [c_{y_t-1}, c_{y_t})\}} N(y_t^*|\boldsymbol{x}_t'\boldsymbol{\beta} + \phi y_{t-1}^{*m}, 1)\, dy_t^* \tag{4.2}$$

$$= \int_{c_{y_t-1}}^{c_{y_t}} N(y_t^*|\boldsymbol{x}_t'\boldsymbol{\beta} + \phi y_{t-1}^{*m}, 1)\, dy_t^* \;=\; \Phi(c_{y_t} - \boldsymbol{x}_t'\boldsymbol{\beta} - \phi y_{t-1}^{*m}) - \Phi(c_{y_t-1} - \boldsymbol{x}_t'\boldsymbol{\beta} - \phi y_{t-1}^{*m}) \;=:\; v_m \tag{4.3}$$

where $\Phi(\cdot)$ denotes the probability function of the standard normal distribution, we must first evaluate the weights v_m appearing in Equation (4.3) for $m = 1, \ldots, M$, and then compute the probabilities

$\psi_m := v_m / \sum_{j=1}^{M} v_j$ which are associated to the numbers $1, \ldots, M$. We note that the proportionality constant needed in Equation (4.2) is independent of y_t^* and m, so we can use the weights v_m directly without taking into account any normalizing constants. After drawing m with probability ψ_m we then have to sample from

$$g(y_t^*|m, \mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta}) \propto \mathbb{1}_{\{y_t^* \in [c_{y_t-1}, c_{y_t})\}} N(y_t^*|\boldsymbol{x}_t'\boldsymbol{\beta} + \phi y_{t-1}^{*m}, 1)$$

which is a normal distribution with mean $\boldsymbol{x}_t'\boldsymbol{\beta} + \phi y_{t-1}^{*m}$ and variance 1, truncated to $[c_{y_t-1}, c_{y_t})$.

The following Algorithm 1 summarizes the necessary steps to produce a sample $\{y_t^{*1}, \ldots, y_t^{*M}\}$ from $f(y_t^*|\mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta})$ when a sample $\{y_{t-1}^{*1}, \ldots, y_{t-1}^{*M}\}$ from $f(y_{t-1}^*|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta})$ is given. For a full filtering procedure this algorithm has to be applied successively for $t = 1, \ldots, T$. To get started one has to draw a sample $\{y_0^{*1}, \ldots, y_0^{*M}\}$ from $f(y_0^*|\mathcal{F}_0, \boldsymbol{c}, \boldsymbol{\theta}) = f(y_0^*|\boldsymbol{c}, \boldsymbol{\theta})$. This density is just the $N(0, \sigma^2)$-density, since the prior of y_0^* is chosen to be independent of $\boldsymbol{c}$ and $\boldsymbol{\theta}$.

Algorithm 1 Fully adapted SIR-based auxiliary particle filter

1. Given $\{y_{t-1}^{*1}, \ldots, y_{t-1}^{*M}\}$ from $f(y_{t-1}^*|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta})$, calculate the weights
$$v_m := \Phi(c_{y_t} - \boldsymbol{x}_t'\boldsymbol{\beta} - \phi y_{t-1}^{*m}) - \Phi(c_{y_t-1} - \boldsymbol{x}_t'\boldsymbol{\beta} - \phi y_{t-1}^{*m}), \qquad m \in \{1, \ldots, M\}$$
and the corresponding probabilities $\psi_m := v_m / \sum_{j=1}^{M} v_j$. Set $k = 1$.
2. Choose $m \in \{1, \ldots, M\}$ with probability ψ_m and sample y_t^{*k} from $N_{[c_{y_t-1}, c_{y_t})}(\boldsymbol{x}_t'\boldsymbol{\beta} + \phi y_{t-1}^{*m}, 1)$.
3. If $k < M$, increment k to $k+1$ and go to Step 2.
If $k = M$, consider $\{y_t^{*1}, \ldots, y_t^{*M}\}$ as a sample from $f(y_t^*|\mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta})$.

4.2 ESTIMATION OF THE LIKELIHOOD ORDINATE

Since $f(\boldsymbol{y}|\boldsymbol{c}, \boldsymbol{\theta})$ can be decomposed into $f(\boldsymbol{y}|\boldsymbol{c}, \boldsymbol{\theta}) = \prod_{t=1}^{T} f(y_t|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta})$, the likelihood ordinate $f(\boldsymbol{y}|\boldsymbol{c}, \boldsymbol{\theta})$ can be estimated by estimating all the one-step ahead densities $f(y_t|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta})$, $t = 1, \ldots, T$. The following procedure takes advantage of this decomposition. Although we want to evaluate $f(\boldsymbol{y}|\boldsymbol{c}^\diamond, \boldsymbol{\theta}^\diamond)$ we suppress the '$\diamond$' for notational convenience.

Algorithm 2 Estimation of the likelihood function

1. Set $t = 1$, initialize $\boldsymbol{c}, \boldsymbol{\theta}$ and obtain a sample $y_0^{*1}, \ldots, y_0^{*M}$ from $f(y_0^*|\boldsymbol{c}, \boldsymbol{\theta})$. Since the prior of y_0^* is chosen to be independent of $\boldsymbol{c}$ and $\boldsymbol{\theta}$, this density is just the $N(0, \sigma^2)$-density.

2. For each value y_{t-1}^{*m} from $f(y_{t-1}^*|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta})$ draw $y_t^{*(m)} \sim N(\boldsymbol{x}_t'\boldsymbol{\beta} + \phi y_{t-1}^{*m}, 1)$.

   Obviously, $\{y_t^{*(m)}|m = 1, \ldots, M\}$ can be considered to be a sample from $f(y_t^*|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta})$.

3. Estimate the one-step ahead density $f(y_t|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta})$ by

$$\hat{f}(y_t|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta}) := \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\left\{y_t^{*(m)} \in [c_{y_t-1}, c_{y_t})\right\}} \tag{4.4}$$

   This estimate can be derived from the equations

$$\begin{aligned} f(y_t|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta}) &= \int f(y_t|y_t^*, \mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta}) f(y_t^*|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta})\, dy_t^* \\ \text{and} \quad f(y_t|y_t^*, \mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta}) &= f(y_t|y_t^*, \boldsymbol{c}) = \mathbb{1}_{\left\{y_t^* \in [c_{y_t-1}, c_{y_t})\right\}} \end{aligned}$$

   since $\{y_t^{*(m)}|m = 1 \ldots, M\}$ represents a sample from $f(y_t^*|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta})$. Of course, the estimator in Equation (4.4) will be zero when all the $y_t^{*(m)}$'s lie outside the interval $[c_{y_t-1}, c_{y_t})$. This would nullify the complete likelihood ordinate. Therefore M should be chosen not too small (cf. Section 5.3) to avoid this problem and to achieve a sufficiently high accuracy. Instead of using one constant M, one also could choose K different constants $M_1, \ldots, M_K$ such that M depends on the value of the observation y_t. Since K will be small in our application we prefer to use $M := \max\{M_1, \ldots, M_K\}$.

4. Apply the filtering procedure in Algorithm 1 to obtain a sample $y_t^{*1}, \ldots, y_t^{*M}$ from $f(y_t^*|\mathcal{F}_t, \boldsymbol{c}, \boldsymbol{\theta})$.

5. If $t < T$, increment t to $t + 1$ and go to Step 2.

6. Return the estimated log likelihood ordinate $\log \hat{f}(\boldsymbol{y}|\boldsymbol{c}, \boldsymbol{\theta}) := \sum_{t=1}^{T} \log \hat{f}(y_t|\mathcal{F}_{t-1}, \boldsymbol{c}, \boldsymbol{\theta})$.

4.3 COMPUTATION OF THE PRIOR ORDINATE

Since we assumed

$$\pi(\boldsymbol{c}, \boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2}\left[\tau^{-2}\boldsymbol{\beta}'\boldsymbol{\beta} + \rho^{-2}\phi^2\right]\right) \cdot \mathbb{1}_{\{0 < c_2 < \ldots < c_{K-1} < C\}}$$

with known hyperparameters τ, ρ, C, we can evaluate the factor $\pi(\boldsymbol{c}^\diamond, \boldsymbol{\theta}^\diamond)$ by multiplying the three factors $\pi(\boldsymbol{\beta}^\diamond) = \prod_{j=0}^{p} N(\beta_j^\diamond|0, \tau^2)$, $\pi(\phi^\diamond) = N(\phi^\diamond|0, \rho^2)$, and $\pi(\boldsymbol{c}^\diamond) \propto \mathbb{1}_{\{0 < c_2^\diamond < \ldots < c_{K-1}^\diamond < C\}}$. Whereas the first and second factor can be evaluated directly with the density function of the normal distribution, the third factor demands the computation of the normalizing constant, which is the inverse of the volume of the $(K-2)$-dimensional subset $\{(c_2, \ldots, c_{K-1}) \mid 0 < c_2 < \ldots < c_{K-1} < C\}$ of $\mathbb{R}^{K-2}$. The volume of

this subset can easily be derived to equal

$$\int_0^C \int_{c_2}^C \cdots \int_{c_{K-2}}^C dc_{K-1} \ldots dc_2 = \frac{1}{(K-2)!} C^{K-2}$$

Therefore $\pi(\boldsymbol{c}^\diamond) = (K-2)!/C^{K-2}$.

4.4 ESTIMATION OF THE POSTERIOR ORDINATE

The aim of this section is to find an estimate for $\pi(\boldsymbol{c}^\diamond, \boldsymbol{\theta}^\diamond | \boldsymbol{y})$. This expression can be decomposed into

$$\pi(\boldsymbol{c}^\diamond, \boldsymbol{\theta}^\diamond | \boldsymbol{y}) = \pi(\boldsymbol{\theta}^\diamond | \boldsymbol{c}^\diamond, \boldsymbol{y}) \pi(c_{K-1}^\diamond | c_2^\diamond, \ldots, c_{K-2}^\diamond, \boldsymbol{y}) \pi(c_{K-2}^\diamond | c_2^\diamond, \ldots, c_{K-3}^\diamond, \boldsymbol{y}) \cdots \pi(c_2^\diamond | \boldsymbol{y}) \tag{4.5}$$

We estimate the $K-1$ factors successively from right to left. We point out here that all full conditional distributions appearing in this section are known from Section 2.2, including the normalizing constants, so that the evaluation is straight-forward.

First we consider the estimation of $\pi(c_2^\diamond | \boldsymbol{y})$. Running the full Gibbs sampler, we get M samples

$$(c_2^{[m]}, \ldots, c_{K-1}^{[m]}, \boldsymbol{\theta}^{[m]}, \boldsymbol{y}^{*[m]}) \sim \pi(\boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{y}^* | \boldsymbol{y}) \qquad (m = 1, \ldots, M)$$

Therefore we can estimate $\pi(c_2^\diamond | \boldsymbol{y})$ by evaluating and averaging M full conditionals for c_2:

$$\hat{\pi}(c_2^\diamond | \boldsymbol{y}) := \frac{1}{M} \sum_{m=1}^{M} f(c_2^\diamond | c_3^{[m]}, \ldots, c_{K-1}^{[m]}, \boldsymbol{\theta}^{[m]}, \boldsymbol{y}^{*[m]}, \boldsymbol{y})$$

The other $K-2$ factors appearing in (4.5) are all estimated by reduced runs (cf. Chib (1995)) of the Gibbs sampler, which we discuss now. To estimate $\pi(c_3^\diamond | c_2^\diamond, \boldsymbol{y})$ we use a reduced run where c_2 is fixed at $c_2^\diamond$ and all components but c_2 are updated in each iteration. So we get J samples $(c_3^{[j]}, \ldots, c_{K-1}^{[j]}, \boldsymbol{\theta}^{[j]}, \boldsymbol{y}^{*[j]})$, $j = 1, \ldots, J$, from $\pi(c_3, \ldots, c_{K-1}, \boldsymbol{\theta}, \boldsymbol{y}^* | c_2^\diamond, \boldsymbol{y})$. Again we can now estimate $\pi(c_3^\diamond | c_2^\diamond, \boldsymbol{y})$ by evaluating and averaging J full conditionals for c_3:

$$\hat{\pi}(c_3^\diamond | c_2^\diamond, \boldsymbol{y}) := \frac{1}{J} \sum_{j=1}^{J} f(c_3^\diamond | c_2^\diamond, c_4^{[j]}, \ldots, c_{K-1}^{[j]}, \boldsymbol{\theta}^{[j]}, \boldsymbol{y}^{*[j]}, \boldsymbol{y})$$

This principle carries over to the remaining factors in (4.5). Finally, the last factor, $\pi(\boldsymbol{\theta}^\diamond | \boldsymbol{c}^\diamond, \boldsymbol{y})$, is estimated using a reduced run where the cutpoints c_k are fixed to $c_k^\diamond$, respectively, and only $\boldsymbol{\theta}$ and $\boldsymbol{y}^*$ are updated in each iteration. Here we get

$$\hat{\pi}(\boldsymbol{\theta}^\diamond | \boldsymbol{c}^\diamond, \boldsymbol{y}) := \frac{1}{N} \sum_{n=1}^{N} f(\boldsymbol{\theta}^\diamond | \boldsymbol{c}^\diamond, \boldsymbol{y}^{*[n]}, \boldsymbol{y})$$

We point out that it is very important to use only reduced runs from the standard Gibbs sampler without grouped moved steps. This is because all reduced runs require at least the fixing of the cutpoint $c_2 = c_2^{\diamond}$. Therefore it is not possible to multiply all cutpoints $c_2, \ldots, c_{K-1}$ by an element (unequal 1) from the partial scale group. This, however, would be done by a grouped move step. The requirement to use the standard Gibbs sampler without grouped move steps seems to be a major disadvantage for our estimating procedure, but this is not really the case. The grouped move steps were used mainly to speed up the convergence for the cutpoints. However, from the full run of the GM-MGMC sampler we have already good estimates for all parameters including the cutpoints, which can be used as starting values in the reduced runs, so that the convergence problem does not play a role any more.

Finally we note that an alternative approach to estimate the posterior ordinate was developed by Ritter and Tanner (1992). It is based on the invariance condition of the Gibbs chain and requires draws from the full Gibbs run, where all full conditionals must be known including their normalization. However, it is only precise when the posterior is low-dimensional and the model does not contain latent variables (cf. Chib and Jeliazkov (2001)). Therefore, for the AOP model we will get more accurate results with the decomposition used here.

5. APPLICATION TO FINANCIAL DATA

We now want to detect and to quantify the influence of covariates information on intraday price changes of stocks. The dependencies detected by modeling such high-frequency financial data are used for example by exchanges which may adapt the rules for the trading process. We choose data of the IBM stock traded on Dec 4, 2000 at the NYSE. We first present some results from an exploratory analysis of the covariates. Then we fit both the AOP and the OP model to the data and finally determine the Bayes factor to decide whether an autoregressive structure is needed in this data set.

5.1 DATA DESCRIPTION AND EXPLORATORY RESULTS

The IBM stock is a very frequently traded stock (about 400 transactions per hour), so that we have enough data even if we do not use data from the first minutes after opening and the last minutes before closing which might exhibit a different behavior. On the Dec 4, 2000, between 9:35:59 am and 2:42:40

price change	0\$	$\pm 1/16$\$	$\pm 2/16$\$	$\leq -3/16$\$ and $\geq 3/16$\$
category	1	2	3	4
frequency	859	782	275	84

Table 1: Absolute price changes: associated categories and observed frequencies.

pm there was a total of 2001 transactions of the IBM stock at the NYSE. 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

As response y_t we choose the absolute values of the price differences. The price differences take on only values which are integer multiples of 1/16 US\$, and 99.5% of them lie between $-3/16$ US\$ and $+3/16$ US\$. The absolute price differences are a reasonable quantity to consider, since a price difference can be decomposed into the product of the absolute price change and the direction of the price change. These two factors of the time series may not only depend on different covariates but may also demand for a different modeling. We note, that a decomposition of the price change into three factors (activity, direction of price change, and absolute price change) was considered in Rydberg and Shephard (2003). Because we 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 1.

We conducted an exploratory analysis to choose appropriate transformations of the covariates. Since the response variable is discrete and takes on only few values, ordinary scatter plots are not informative here, especially when the regressor is also discrete or categorial. Therefore we grouped the covariate data in intervals of the same length and then computed the average response for each interval. Considering the covariate TIMEDIFF it turns out to be useful to take the logarithm of TIMEDIFF plus 1 to get a nearly linear dependency of the response. Of course, by adding 1 we avoid the value log(0). For small (logarithmic) time differences we get an average response of 1.17, for big ones an average response of 1.90. The relatively high difference of $1.90 - 1.17 = 0.73$ is a first hint at the strong significance of this covariate. Note that only when the cutpoints are estimated to be equidistant to some extent, we have a validation for the chosen transformation of the covariate. Otherwise one should use other transformations that take into account the different distances between the cutpoints. Considering the covariate SIZE it is again useful to take the logarithm to get a nearly linear dependency of the response. However, the difference between the maximal and minimal average response is only $1.87 - 1.71 = 0.16$.

So we expect that log(SIZE) is not as significant as log(TIMEDIFF+1).

5.2 MODEL ESTIMATION

We now fit the AOP model defined in Equations (2.1) and (2.2) to the data. Since we are also interested in a comparison of this model to the OP model without an autoregressive component, we also fitted the corresponding OP model to the data. In particular, the latent structures of these two models are given by:

$$\text{AOP:} \qquad y_t^* = \beta_0 + \beta_1 \cdot \log(\text{TIMEDIFF+1})_t + \beta_2 \cdot \log(\text{SIZE})_t + \phi y_{t-1}^* + \varepsilon_t^*$$

$$\text{OP:} \qquad y_t^* = \beta_0 + \beta_1 \cdot \log(\text{TIMEDIFF+1})_t + \beta_2 \cdot \log(\text{SIZE})_t + \varepsilon_t^*$$

We first consider the specification of the hyperparameters τ (standard deviation of prior for the regression coefficients), σ (standard deviation of prior for y_0^*), ρ (standard deviation of prior for ϕ), and C (maximum for cutpoint c_3). Since the variance of the error term ε_t^* is fixed to 1 and the cutpoint c_1 is fixed to 0, we do not expect a very high value for the cutpoint c_3 and for $|y_0^*|$. Otherwise the error term ε_t^* would have hardly any impact on the time series, and therefore the data could be nearly deterministically explained by the covariates and the autoregression. Therefore we choose $C = 10$ and $\sigma^2 = 1.0$. For the same reason we do not expect an extreme intercept β_0. Since the logarithms of (TIMEDIFF+1) and SIZE have values between 0 and 10.8, we also do not expect extreme values for $\beta_j, j = 1, 2$. Therefore we choose $\tau^2 = 10$ which leads here to a sufficiently noninformative prior for the regression coefficients. The autoregressive component in the AOP model is expected to be present, but not too large, so the choice of $\rho^2 = 0.1$ seems to be adequate. By using other hyperparameter values we have seen that the posterior estimates are not very prior-sensitive. This also can be expected because of the large number of observations.

For the estimation of the OP model the corresponding GM-MGMC sampler of Liu and Sabatti (2000) is used. The GM-MGMC sampler there can be derived in a natural way from the GM-MGMC sampler for the AOP model since the OP model is a submodel of the AOP model. The partial scale group used for the sampler in the AOP case then simplifies to a (total) scale group since the parameter ϕ does not occur in the OP model.

We now run both GM-MGMC samplers for 15000 iterations and discard the first 5000 iterations for burnin. From the simulation study we know that this leads to very accurate estimates. The results are

	Autoregressive Ordered Probit			Ordered Probit		
	estimate	std.err.	90% cred.int.	estimate	std.err.	90% cred.int.
c_2	1.1231	0.0344	(1.0664,1.1798)	1.1172	0.0327	(1.0648,1.1710)
c_3	1.9593	0.0540	(1.8733,2.0496)	1.9381	0.0534	(1.8509,2.0278)
β_0	-0.5567	0.1437	(-0.7918,-0.3224)	-0.5179	0.1442	(-0.7527,-0.2813)
β_1	0.2350	0.0328	(0.1821,0.2888)	0.2278	0.0324	(0.1743,0.2805)
β_2	0.0368	0.0197	(0.0043,0.0688)	0.0368	0.0196	(0.0044,0.0695)
ϕ	0.1362	0.0277	(0.0913,0.1818)			

Table 2: Posterior mean estimates and corresponding estimated standard deviations and 90% posterior credible intervals for parameters in Models AOP and OP.

summarized in Table 2. It shows the posterior mean estimates for the cutpoints, the regression coefficients, and the autoregressive parameter using the iterations 5001 to 15000 of the GM-MGMC Gibbs sampler together with their corresponding estimated standard deviations and 90% credible intervals.

We conclude that the intercept, log(TIMEDIFF+1), and log(SIZE) are all significant in both models, as well as the autoregressive component in the AOP model with posterior mean estimate 0.1362. Because of the positive sign of the estimates for β_1 we conclude that the more time elapses since the last transaction, the higher the expected price change is. The same holds for the transaction volume: The more stocks are traded, the higher the expected price change is.

Figure 3 shows the estimated posterior marginal densities for the parameters in the AOP model. The densities are unimodal and quite symmetric, so that the posterior mean estimates represent a high density point. This property is used in the next section. The estimated marginal densities for the parameters in the OP model have the same shape and are therefore not shown. Figure 4 finally shows the estimated autocorrelations in the chains produced by the GM-MGMC samplers after the burnin period of 5000 iterations for both the AOP model (solid curves) and the OP model (dotted curves). The autocorrelations decline quite fast for all parameters and differ hardly between the two models.

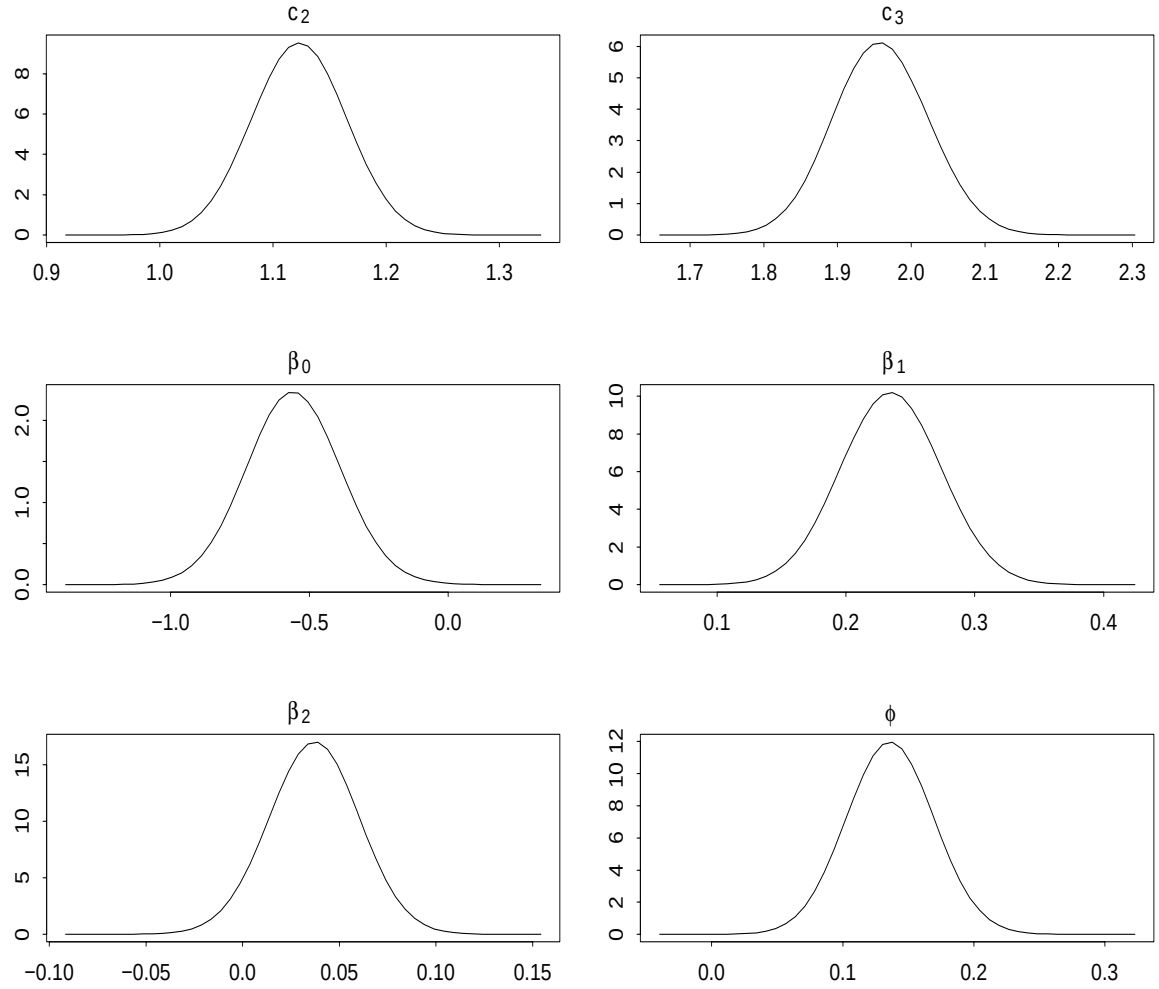

Figure 3: Estimated posterior marginal densities for parameters of AOP model.

5.3 BAYES FACTOR OF AOP AGAINST OP MODEL

Since the posterior mean estimate of ϕ in the AOP model is 0.1362 and the corresponding 90% credible interval is far away from 0, we can assume the presence of an autoregressive structure in the IBM data. From this point of view we prefer the AOP model to the OP model for this data set. However, so far we do not know, how big the benefit is of using the AOP model instead of the simpler OP model. Therefore we now estimate the marginal likelihoods for the data under the two models to get the Bayes factor of the AOP against the OP model.

We follow Section 4 and estimate the likelihood ordinate, the posterior ordinate, and the prior ordinate for both models. Since AOP and OP are nested models and, in particular, OP is a submodel of AOP,

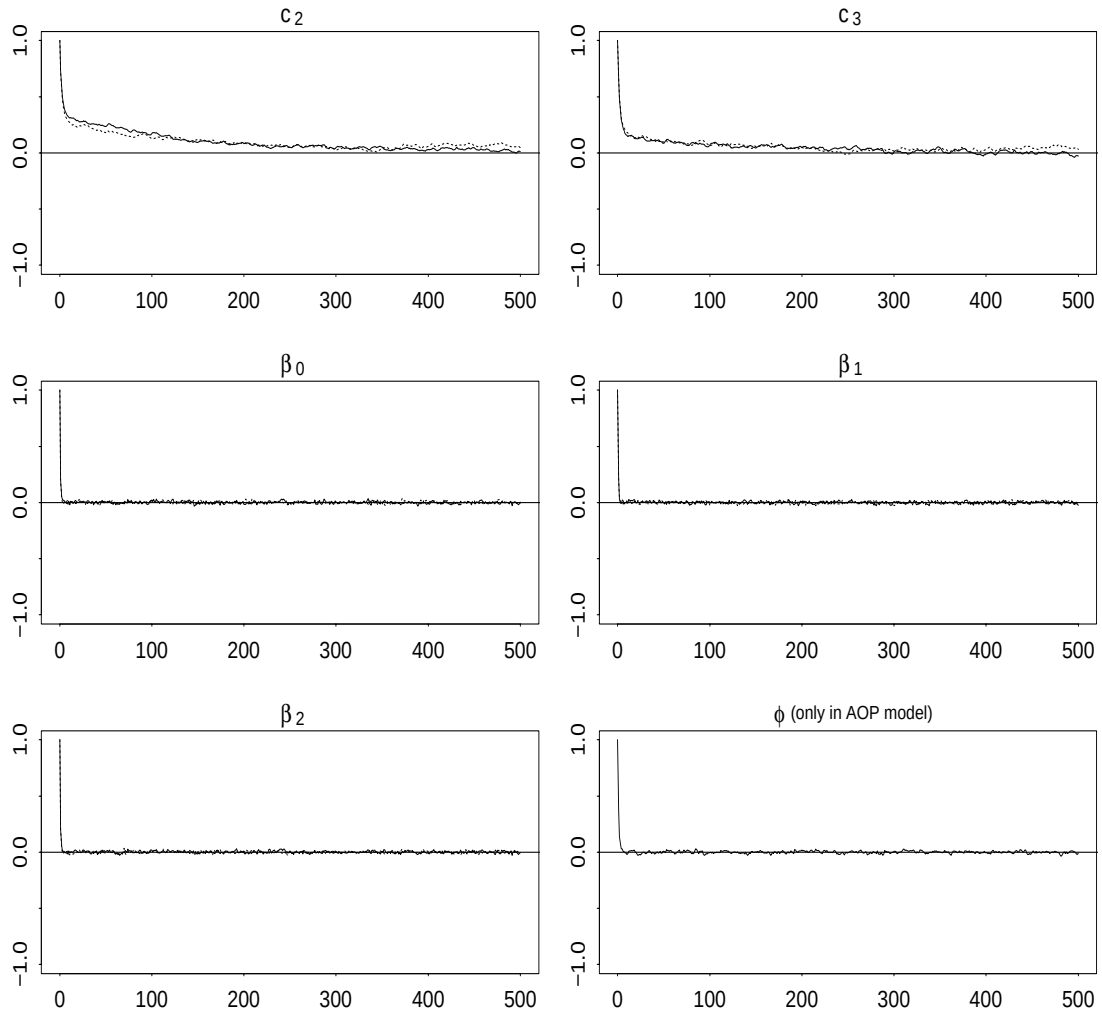

Figure 4: Autocorrelations of chains produced by the GM-MGMC samplers for the AOP model (solid) and the OP model (dotted). Visible difference only for cutpoints.

we can use the estimation methods for the AOP model also for the OP model. Of course, the algorithms presented in Section 4 simplify in this case. For example, since y_t^* is independent of y_{t-1}^* in the OP model, a filtering procedure as for the AOP model is not necessary here and Algorithm 1 reduces to sampling M times from $N_{[c_{y_t-1},c_{y_t})}(\boldsymbol{x}_t'\boldsymbol{\beta},1)$. Furthermore, since the OP model does not contain the autoregressive component of the AOP model, $\boldsymbol{\theta}=\boldsymbol{\beta}$ and the prior ordinate $\pi(\boldsymbol{c}^\diamond,\boldsymbol{\theta}^\diamond)$ equals $\pi(\boldsymbol{c}^\diamond)\pi(\boldsymbol{\beta}^\diamond)$.

For the fixed parameters $\boldsymbol{c}^\diamond$ and $\boldsymbol{\theta}^\diamond$ appearing in Equation (4.1) one should use high density points. Hence we take the corresponding posterior mean estimates given in Table 2. We run the Algorithms 1 and 2 with $M=30000$ particles. This takes about 130 minutes for the AOP model on an UltraSPARC

B_{ij}	Evidence against model j
$1 - 3.2$	Not worth more than a bare mention
$3.2 - 10$	Substantial
$10 - 100$	Strong
> 100	Decisive

Table 3: Jeffrey's Bayes factor scale with B_{ij} denoting the Bayes factor for model i versus model j.

	Autoregressive Ordered Probit	Ordered Probit
$\log f(\boldsymbol{y}\|\boldsymbol{c}^\diamond, \boldsymbol{\theta}^\diamond)$	-2233.3235	-2244.9539
$\log \pi(\boldsymbol{c}^\diamond, \boldsymbol{\theta}^\diamond)$	-10.0014	-10.1388
$\log \pi(\boldsymbol{c}^\diamond, \boldsymbol{\theta}^\diamond\|\boldsymbol{y})$	15.4477	13.0259
$\log m(\boldsymbol{y})$	-2258.7726	-2268.1186
Bayes factor of AOP against OP	exp(-2258.7726+2268.1186) = **11452.92**	

Table 4: Estimated log likelihood ordinate, log prior ordinate, log posterior ordinate and marginal log likelihood for Models AOP and OP. Bayes Factor of AOP against OP.

III Cu 900 Mhz processor. The choice of M seems to be sufficiently high since we got nearly the same results for $M = 10000$ and $M = 20000$. The computation of the prior ordinate can be done exactly. Here we only mention that since the hyperparameter C was chosen to be 10 and we have 4 response categories, we get $\pi(\boldsymbol{c}^\diamond) = 2/C^2 = 0.02$. For the means which have to be computed for estimating the posterior ordinate, we use 10000 samples from the full and the reduced runs each. Table 4 summarizes the results.

We get a Bayes factor of 11452.92 for the AOP against the OP model. Following Jeffrey's Bayes factor scale in Table 3 which was first proposed by Jeffreys (1961), we conclude that the AOP model fits the IBM data decisively better than the OP model.

6. SUMMARY AND DISCUSSION

In this paper we introduced an autoregressive ordered probit (AOP) model where the latent process includes both covariates and 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 GM-MGMC Gibbs sampler shows a very satisfying behavior. Furthermore we provided a filtering procedure as well as an estimation procedure for the marginal likelihood required to calculate Bayes factors. Finally we used the GM-MGMC Gibbs sampler to detect and to quantify significant covariates for the price development of the IBM stock. For absolute values of the price changes the logarithm of the elapsed time between two following transactions and the transaction volume are important covariates. The AOP model fits the IBM data decisively better than the OP model. Of course, other applications of the model are possible. For example, one could think about pain patients where patients assess the severity of their pain on an ordinal scale, and where besides an autoregressive structure some covariates are supposed to influence the pain severity. 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. In addition, we would like to cover other time dependence structures such as stochastic volatility models.

ACKNOWLEDGEMENT

This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 386 *Statistical Analysis of Discrete Structures*. We like to thank the referees for their comments which led to an improved presentation.

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 (2001). *Econometric Modelling of Stock Market Intraday Activity*. Boston:

Kluwer Academic Publishers.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association 90*, 1313–1321.

Chib, S. and I. Jeliazkov (2001). Marginal Likelihood from the Metropolis-Hastings Output. *Journal of the American Statistical Association 96*, 270–281.

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.

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.

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.

Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). Oxford: Clarendon Press.

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

Pitt, M. K. and N. Shephard (1999). Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association 94*, 590–599.

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

Ritter, C. and M. A. Tanner (1992). Facilitating the Gibbs sampler: The Gibbs Stopper and the Griddy-Gibbs sampler. *Journal of the American Statistical Association 87*, 861–868.

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

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