Knorr-Held:

Markov Chain Monte Carlo Simulation in Dynamic Generalized Linear Mixed Models

Sonderforschungsbereich 386, Paper 8 (1995)

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

Projektpartner

Markov Chain Monte Carlo Simulation in Dynamic Generalized Linear Mixed Models

Leonhard Knorr–Held
Seminar für Statistik
Ludwig–Maximilians–Universität München
Ludwigstr. 33
D–80539 München [1]

Abstract

Dynamic generalized linear mixed models are proposed as a regression tool for nonnormal longitudinal data. This framework is an interesting combination of dynamic models, by other name state space models, and mixed models, also known as random effect models. The main feature is, that both time– and unit–specific parameters are allowed, which is especially attractive if a considerable number of units is observed over a longer period. Statistical inference is done by means of Markov chain Monte Carlo techniques in a full Bayesian setting. The algorithm is based on iterative updating using full conditionals. Due to the hierarchical structure of the model and the extensive use of Metropolis–Hastings steps for updating this algorithm mainly evaluates (log–)likelihoods in multivariate normal distributed proposals. It is derivative–free and covers a wide range of different models, including dynamic and mixed models, the latter with slight modifications. The methodology is illustrated through an analysis of artificial binary data and multicategorical business test data.

Some key words: Bayesian inference; Generalized linear model; Heterogeneity; Longitudinal data; Markov chain Monte Carlo; Metropolis–Hastings algorithm; Time–varying regression parameters.

[1] Email: `leo@stat.uni-muenchen.de`. Last corrections July 25, 1995.

1 Introduction

Generalized linear models provide a powerful regression tool for the analysis of nonnormal cross–sectional responses. To deal with nonnormal longitudinal data, several extensions are proposed. Generalized linear mixed models (Breslow & Clayton, 1993; Zeger & Karim, 1991) and dynamic generalized linear models (Fahrmeir, 1992a) are examples of hierarchical models, where independence assumptions are imposed conditioning on stochastic unit–specific or time–dependent parameters. These models try to meet the requirements of longitudinal data in different ways. While the first approach allows parameters to vary over units but not over time, the second does vice versa.

Consider the following situation: Longitudinal data (y_{ti}, x_{ti}) $(t = 1, \ldots, T,$ $i = 1, \ldots, n)$ is observed on n units over T time periods. For simplicity, it is assumed that response y_{ti} is univariate. An ordinary generalized linear model (GLM) assumes mutual independence of the y's. The linear predictor η_{ti}, which connects covariates x_{ti} with the mean $h(\eta_{ti})$ of responses y_{ti} via the response function h, is

$$\eta_{ti} = z_{ti}^T \beta,$$

where the design vector z_{ti} is formed out of x_{ti}. The unknown regression parameter β is independent of time t and unit i.

However, often there is heterogeneity among units arising from covariate effects varying from one unit to another or due to unobserved (unit–specific) covariates. Therefore unit–specific parameters, sometimes called random effects, $b_i \sim N(0, D)$ are introduced. The linear predictor is extended to

$$\eta_{ti} = z_{ti}^T \beta + w_{ti}^T b_i.$$

Often w_{ti} is a subset of z_{ti}. The parameter β represents the population average effect whereas the b_i's represent the unit–specific deviations from β.

This approach is especially attractive if n is large and T is small.

If n is small but T is large a dynamic model is more appropriate. The idea is to introduce time–varying parameters β_t in

$$\eta_{ti} = z_{ti}^T \beta_t.$$

The temporal variation of these parameters is described in an additional transition model. This approach allows for trend or seasonal components as well as for time–varying effects of covariates. However, possibly existing heterogeneity among the units is not taken into account.

In this paper we combine these two approaches. The linear predictor is extended to

$$\eta_{ti} = z_{ti}^T \beta_t + w_{ti}^T b_i,$$

so both time–dependent as well as unit–specific parameters are allowed. To include multivariate models such as cumulative or sequential models for multicategorical responses (a recent survey is given in Fahrmeir & Tutz, 1994), a more general form

$$\eta_{ti} = Z_{ti} \beta_t + W_{ti} b_i$$

is considered. Here η_{ti} is a vector of dimension q. Since dynamic and mixed models are combined this framework is called a dynamic generalized linear mixed model (DGLMM). Note that dynamic models ($D = 0$) as well as mixed models ($\beta_t \equiv \beta$) are special cases of a DGLMM.

For normal data models of this kind have been proposed already (e.g. Hsiao, 1986), whereas corresponding work for nonnormal data is quite rudimentary. This is mainly due to the fact that integrations necessary to compute functionals of the posterior distribution are very difficult using standard numerical integration techniques. Alternatively one may try to maximize the posterior to avoid integration. In fact, the algorithms proposed in Fahrmeir

(1992a,b) compute the posterior mode and curvature in dynamic generalized linear models. However, extensions of these methods to the model considered here suffer from the fact that maximization of both time– and unit–specific parameters has to be done iteratively using backfitting or Gauss–Seidel algorithms. Additional estimation of hyperparameters increases computing time enormously.

In contrast, the MCMC algorithm proposed here is surprisingly simple, covering a wide range of models with only minor changes and leading to sufficient exact results in a reasonable amount of time. It allows not only for posterior mean and covariance estimation but also for estimating the posterior density itself, computing simultaneously credible regions (Besag, Green, Higdon & Mengersen, 1995) and for selecting models based on Bayes factors (Raftery, 1995).

Like in Zeger & Karim (1991) and in Carlin, Polsen & Stoffer (1992) a Bayesian approach is adopted, treating all unknown parameters and hyperparameters as random with appropriate prior specification. The algorithm is based on iterative updating using full conditionals. The main difference to the well known Gibbs sampling algorithm proposed in a related context in the references above is the use of a Metropolis–Hastings step for updating, sometimes called "Metropolis within Gibbs". For a recent survey of MCMC methods see Tierney (1994), Besag, Green, Higdon & Mengersen (1995) or Smith & Roberts (1993).

Finally the transition model needs some further comments. We use a vector autoregressive model

$$\beta_t = -\sum_{l=1}^{z} F_l \beta_{t-l} + u_t, \quad u_t \sim N(0, Q) \quad (t = z+1, \ldots, T).$$

$F_1, \ldots, F_z$ are called transition matrices and are assumed to be known, Q has to be regular. This is roughly equivalent to the state space approach in

Fahrmeir (1992a), where Q is allowed to be singular. For more details on similarities and differences of both transition models see Knorr-Held (1995).

This article is organized as follows: Section 2 introduces dynamic generalized linear mixed models in a Bayesian formulation. The algorithm for simulating the (numerically intractable) posterior distribution is presented in Section 3. This section also gives a brief discussion of the available point and interval estimates using samples from the posterior. Section 4 illustrates the methodology through an analysis of artificial binary response data and an application to multicategorical business test data. Here several multivariate versions of DGLMM's with restrictions on the parameters are discussed. We summarize our findings and outline some further generalizations in Section 5.

2 Dynamic Generalized Linear Mixed Models

Let (y_{ti}, x_{ti}) $(t = 1, \ldots, T,\ i = 1, \ldots, n)$ denote the observation of unit i at time t, where $x_{ti} = (x_{ti1}, \ldots, x_{tim})^T$ is the vector of covariates and $y_{ti} = (y_{ti1}, \ldots, y_{tiq})^T$ is the q dimensional response vector. A DGLMM is based on

an observation model for the y_{ti}'s,
a (random effect) model for the b_i's,
and a transition model for β_t's.

The observation model consists of a *distributional assumption* for y_{ti} and a *structural assumption* for the mean of y_{ti} given parameters β_t and b_i:

$$E(y_{ti}|\beta_t, b_i, x_{ti}) = h(Z_{ti}\beta_t + W_{ti}b_i),$$

where $h : I\!R^q \rightarrow I\!R^q$ is a so–called response function, Z_{ti} is a $q \times p$–matrix and W_{ti} is a $q \times r$–matrix, both formed out of x_{ti}.

The unit–specific effects b_i of dimension r are assumed to follow a Gaussian distribution

$$b_i \sim N(0, D) \quad (i = 1, \ldots, n)$$

with mean zero and covariance matrix D.

The parameters β_t of dimension p are supposed to vary over time. The simplest model is a random walk of first order but to include important models like a local linear trend model or a seasonal component, a general multivariate autoregressive structure

$$\beta_t = -\sum_{l=1}^{z} F_l \beta_{t-l} + u_t, \quad u_t \sim N(0, Q) \quad (t = z+1, \ldots, T)$$

is proposed. Using the lag operator $L(\beta_t) = \beta_{t-1}$ and defining a matrix polynomial , $F(L) = I + F_1 L + \ldots + F_z L^z$, it may be written shorter as

$$F(L)\beta_t = u_t, \quad u_t \sim N(0, Q).$$

Diffuse priors on the initial values $\beta_t \propto$ const $(t = 1, \ldots, z)$ complete the specification. Note that this model definition is only reasonable, if all components of β_t have the same lag z. If different components have different lags, some formal problems arise for the initial values. In this situation it is useful to split up β_t in independent components and let every component follow a specific autoregressive model. However, these changes are obvious and omitted here to avoid a non–transparent setting.

A prior specification for the hyperparameters Q and D completes our model. For D we choose a conjugate prior, the inverted Wishart distribution. The same is possible for Q, but often the components of u_t are assumed to be independent, so Q is diagonal. Therefore inverse gamma priors, the univariate conjugate analogues, are assumed for the diagonal entries of Q.

The definition of this model is in spirit of a hierarchical model, where conditional independence is assumed among the following random variables:

$$\begin{aligned} y_{ti}|\beta_t, b_i, D, Q \qquad & (t = 1, \ldots, T, \quad i = 1, \ldots, n), \\ u_t|Q \qquad & (t = z+1, \ldots, T), \\ b_i|D \qquad & (i = 1, \ldots, n), \\ D \text{ and } Q. & \end{aligned}$$

Note that until now covariates are assumed to be non–stochastic. To include past observations as covariates or other stochastic covariates these independence assumptions have to be modified appropriately.

3 Simulating the Posterior

In this section a general algorithm is proposed for analyzing dynamic generalized linear mixed models. The MCMC sampling scheme is not affected by changes in the transition model, different distributional assumptions for the responses or different response functions. It is derivative–free and consists mainly of generating multivariate normal variates and evaluating (log-) likelihoods. However, due to this model flexibility, there may exist more efficient MCMC procedures for special versions of DGLMM's. For example, convergence of the simulated Markov chain may be better, if updating $\beta = (\beta_1, \ldots, \beta_T)$ is done componentwise rather than cross–sectional wise, but implementation is much more difficult. The reader should keep in mind that the major goal of this section is to present a unifying tool, which works in our limited experience pretty well for several kinds of models.

The sampling scheme is based on iterative updating using full conditionals. Full conditional densities are shortly denoted by $p(\beta_t|\quad)$, $p(b_i|\quad)$ and so on. Due to the hierarchical structure of the model, the full conditionals

to be considered are

$$\begin{aligned}
p(\beta_t|\beta_{s\neq t}, b, Q, D, y) &\sim p(\beta_t|\beta_{s\neq t}, b, Q, y) \quad (t = 1, \ldots, T), \\
p(b_i|b_{j\neq i}, \beta, Q, D, y) &\sim p(b_i|\beta, D, y) \quad (i = 1, \ldots, n), \\
p(Q|\beta, b, D, y) &\sim p(Q|\beta) \text{ and finally} \\
p(D|\beta, b, Q, y) &\sim p(D|b),
\end{aligned}$$

where b, β and y stands for all b_i's, all β_t's and all y_{ti}'s, respectively. A Metropolis–Hastings (M–H) step is used for updating the β_t's and the b_i's, whereas samples from $p(Q|\quad)$ and $p(D|\quad)$ are generated using a Gibbs step. The algorithm is a hybrid procedure as introduced in Tierney (1994) and further discussed in the context of updating full conditionals in Besag, Green, Higdon & Mengersen (1995).

3.1 Updating using full conditionals

Let us start with the full conditional of parameter β_t. Applying Bayes's theorem we notice that

$$p(\beta_t|\quad) \propto \prod_{i=1}^{n} p(y_{ti}|\beta_t, b_i) p(\beta_t|\beta_{s\neq t}, Q).$$

Because the conditional distribution $p(\beta_t|\beta_{s\neq t}, Q)$ is Gaussian, the full conditional can be written as

$$p(\beta_t|\quad) \propto \prod_{i=1}^{n} p(y_{ti}|\beta_t, b_i) \varphi(\beta_t; \mu_t, \Sigma_t).$$

Here $\varphi(\beta_t; \mu_t, \Sigma_t) = \varphi(\beta_t)$ denotes the density function of the normal distribution $N(\mu_t, \Sigma_t)$ with mean μ_t and covariance matrix Σ_t. These parameters depend on the current values of Q and of neighboring parameters $\beta_{s\neq t}$. Different transition models result in different formulae for μ_t and Σ_t. Here we give two examples: A random walk of first order $\beta_t = \beta_{t-1} + u_t$ leads to

$$N(\mu_t, \Sigma_t) = \begin{cases} N(\beta_{t+1}, Q) & (t = 1) \\ N(\frac{1}{2}\beta_{t-1} + \frac{1}{2}\beta_{t+1}, Q/2) & (t = 2, \ldots, T-1) \\ N(\beta_{t-1}, Q) & (t = T) \end{cases} .$$

A seasonal model $\beta_{t-z} + \ldots + \beta_t = u_t$ results in $\Sigma_t = Q/(z+1)$ and $\mu_t = -\left(\beta_{t-z} + 2\beta_{t-z+1} + \ldots + z\beta_{t-1} + z\beta_{t+1} + \ldots + 2\beta_{t+z-1} + \beta_{t+z}\right)/(z+1)$ for $t = z+1, \ldots, T-z$ with slight modifications otherwise. In general the formulae are

$$\Sigma_t^{-1} = \sum_{j=\max(0,1+z-t)}^{\min(z,T-t)} F_j^T Q^{-1} F_j$$

and

$$\mu_t = -\Sigma_t \left\{ \sum_{j=\max(0,1+z-t)}^{\min(z,T-t)} F_j^T Q^{-1} \left(\sum_{i=0, i\neq j}^{z} F_i \beta_{t+j-i} \right) \right\}$$

for t=1, ...,T. A proof is given in Knorr–Held (1995).

The M–H update step uses $\varphi(\beta_t)$ as proposal density and evaluates the likelihood at time t

$$p(y_t|\beta_t, b_i) = \prod_{i=1}^{n} p(y_{ti}|\beta_t, b_i).$$

We call $\beta_t^* \sim N(\mu_t, \Sigma_t)$ a conditional independence proposal, since it does not depend on the current value of β_t but it does depend on current values of neighboring $\beta_{s\neq t}$ and of Q. This proposal density has some advantages. First the M–H acceptance probability

$$\alpha = \min\left\{1, \frac{p(\beta_t^*| \quad)\varphi(\beta_t^\square)}{p(\beta_t^\square| \quad)\varphi(\beta_t^*)}\right\}$$

simplifies to the ratio of $p(y_t|\beta_t, b_i)$, evaluated at the current value $\beta_t^\square$ ($\square$ stands for "current value") and at the proposed new value β_t^*. Secondly the algorithm shows good performance with an acceptance rate ranging from 0.3 to 0.9 for lots of different data and models. There is no need to tune the algorithm (choosing a different proposal) and the lack of a tuning parameter here is some kind of relief.

Updating β_t consists of two steps:

1. Sample $\beta_t^* \sim N(\mu_t, \Sigma_t)$

2. Accept β_t^* with probability

$$\alpha = \min\left\{1, \frac{p(y_t|\beta_t^*, b_i)}{p(y_t|\beta_t^{\square}, b_i)}\right\},$$

otherwise leave β_t unchanged.

In contrast to this very simple M–H update step a Gibbs step is much more demanding. For example, the use of $\varphi(\beta_t)$ as an envelope function in a rejection step as proposed in Carlin, Polsen & Stoffer (1992) in a related context often is unattractive due to very high rejection probabilities for nearly all kinds of models and data structures. It may work for time series ($n = 1$), but also in this case updating by M–H is much more effective. More sophisticated envelope functions often need the knowledge of the mean and curvature of the full conditional (see Zeger & Karim, 1991) or are applicable only to univariate densities as in the case of adaptive rejection sampling and its generalizations (Gilks & Wild, 1992; Gilks, 1992; Gilks, Best & Tan, 1994). Compared to these problems arising by use of a Gibbs step, the proposed M–H update step is astonishingly simple without any requirements on the full conditionals like log–concavity etc.

These considerations are completely transferable to the following case of updating the b_i's. Again, a M–H step is more effective and less demanding as a Gibbs step, proposed in Zeger & Karim (1991). Applying Bayes's theorem the full conditional of parameter b_i can be written as

$$p(b_i| \quad) \propto \prod_{t=1}^{T} p(y_{ti}|\beta_t, b_i)\varphi(b_i; 0, D).$$

In the following the likelihood of unit i is denoted by

$$p(y_i|\beta_t, b_i) = \prod_{t=1}^{T} p(y_{ti}|\beta_t, b_i).$$

Although this full conditional is very similar to $p(\beta_t| \quad)$, an independence proposal $b_i^* \sim N(0, D)$ often has low acceptance rate, thus performance is

poor. The main reason for that is, that $p(b_i|\quad)$ may differ substantially from $\varphi(b_i; 0, D)$, especially for large T. Therefore we use a random walk proposal $b_i^* \sim N(b_i^{\square}, E)$, which usually performs better in such situations (Tierney, 1994).

Updating the b_i's proceeds as follows:

1. Sample $b_i^* \sim N(b_i^{\square}, E)$
2. Accept b_i^* with probability

$$\alpha = \min\left\{1, \frac{p(y_i|\beta_t, b_i^*)\varphi(b_i^*; 0, D)}{p(y_i|\beta_t, b_i^{\square})\varphi(b_i^{\square}; 0, D)}\right\},$$

   otherwise leave b_i unchanged.

Note that α is not affected by the choice of E. Therefore E can be used as a tuning parameter to control the acceptance rate. We had good experience with $E = D/2$.

Sometimes restrictions are imposed on components of β_t, b_i or both. In a dynamic cumulative model without random effects, for example, those components of β_t, representing the thresholds $\theta_{t1}, \ldots, \theta_{tq}$ have to fulfill the order restriction $\theta_{t1} < \theta_{t2} < \ldots < \theta_{tq}$. Such constraints are easily incorporated in the sampling scheme given above by disregarding those proposals β_t^* or b_i^*, that do not obey the restriction.

Sampling from $p(Q|\quad)$ and from $p(D|\quad)$ is straightforward due to conjugate settings. Assuming an inverted Wishart prior $D \sim IW_r(\zeta, B)$, the full conditional is given by

$$D|\quad \sim IW_r\left(\zeta + n, \left(B^{-1} + \sum_{i=1}^{n} b_i b_i^T\right)^{-1}\right).$$

Similar results hold for Q. If Q is assumed to be diagonal and all components of β_t have the same lag z, an inverse gamma prior $Q_{jj} \sim IG(a, b)$ results in

the full conditional

$$Q_{jj}| \quad \sim IG\left(a + (T-z)/2, \left(1/b + \sum_{t=z+1}^{T} u_{tj}^2\right)^{-1}\right).$$

3.2 Visiting Schedule

Some considerations have to be made concerning the order in which parameters and hyperparameters are updated. It is a natural choice to visit the blocks $[\beta_1, \ldots, \beta_T]$, $[b_1, \ldots, b_n]$, $[Q_{11}, \ldots, Q_{pp}]$ and $[D]$ in a deterministic order. To avoid an artificial "drift" as discussed in Besag, Green, Higdon & Mengersen (1995) the components of the first block are visited in random order. Implementation is easy using the ranks of T uniformly distributed random variates. The second and the third block are so–called coding sets (Besag, 1974; Besag, Green, Higdon & Mengersen, 1995), thus deterministic updating within the blocks is the obvious choice. Of course computation time can be improved considerably using a parallel implementation.

Concerning the speed of convergence it is helpful to start the iterations without the b_i's (that is $b_i = 0$) for let say half of the burn in. Then after the β_t's reached the population average, the b_i's are added to the sampling scheme.

3.3 Estimation Based on Posterior Samples

From a computational point of view, parameter estimates may be divided into two groups, those, which can be evaluated recursively and those for which the whole sample has to be stored. This distinction is especially important in multi–parameter models like the one considered here, since storing all samples often requires an enormous storage size. Therefore recursively evaluated estimates, such as the mean and the variance, should be the standard output of the algorithm, although they might give just a crude characterization of the posterior. The statistician has to decide, which parameters require more

sophisticated estimation procedures ranging from several quantile estimates over simultaneous credible regions up to (marginal) density estimation itself. Then the samples from the parameter of particular interest should be stored to make the evaluation of those estimates possible.

4 Applications

4.1 Artificial data

We generated binary response data $Y_{ti} \sim B(1, \pi_{ti})$ according to the logistic model

$$\text{logit}(\pi_{ti}) = (1, x_{ti}) * \beta_t + b_i.$$

The b_i's were generated from a standard normal distribution, the group indicator $x_{ti} = x_i$ was set to 1 for half of the $n = 50$ units and zero for the remainders. The β_t's were generated following a random walk of first order with initial value $\beta_1 = (0, 1)^T$, $Q = \text{diag}(0.05, 0.05)$ and $T = 50$. The second component of the β_t's may be interpreted as a time–dependent group effect. This simple model was chosen to investigate, if the procedure is able to separate unit–specific and time–dependent parameters.

Parameter estimates are the result of a single run of length 52,500 cycles, discarding the first 2,500 and using every 5th sample thereafter. Expectation and standard deviation of the priors for all hyperparameters have been set to the real values. Figure 1 shows the generated β_t's, posterior mean estimates and pointwise one standard deviation confidence bands. Figure 2 shows the true b_i's again with posterior mean $\pm$ one standard deviation confidence interval. The figures indicate that the MCMC procedure gives reasonable results for this data set. Finally Figure 3 and 4 give examples of more sophisticated estimation methods, see Section 3.3. Figure 3 presents the estimated marginal posterior distribution of the group effect, obtained by applying a kernel estimate to the posterior realizations. 50 %, 80 % and 95 % simultaneous credible regions for the group effect are shown in Figure 4.

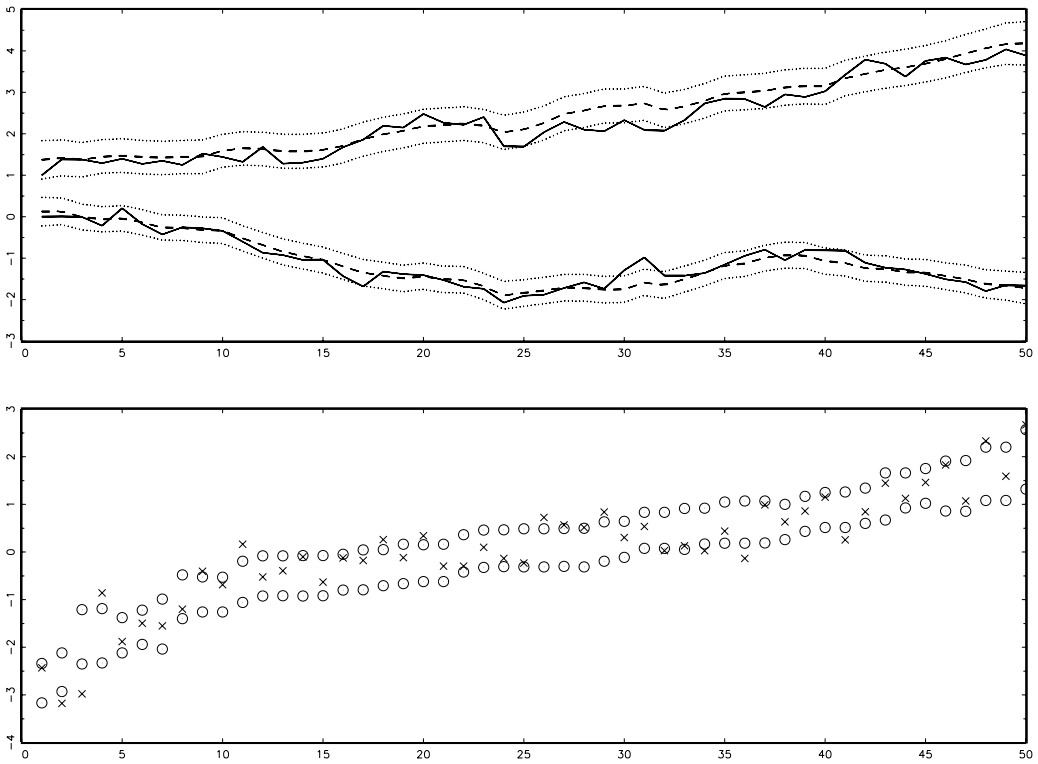

Figure 1 (top): True β_t's (solid line), posterior mean estimates (dashed line) and pointwise one posterior standard deviation confidence bands (dotted lines).

Figure 2 (bottom): True b_i's ($\times$) and posterior mean estimates $\pm$ one posterior standard deviation ($\odot$). The units are ordered respective to the mean estimates.

4.2 Business test data

Fahrmeir (1992a), Fahrmeir & Nase (1994) and Knorr–Held (1995) analyzed data from the IFO business test applying a dynamic cumulative model. This monthly data is based on a questionnaire, answered by $n = 55$ firms of a specific industrial branch for the years 1980 to 1990. The response variable "short range production plans" is given in three ordered categories, "decrease", "no change" and "increase". Its conditional distribution is assumed to depend on answers concerning "orders in hand", "expected development of the state of business for the next 6 months" as well as on the production plans of the previous month. These three questions are also trichotomous

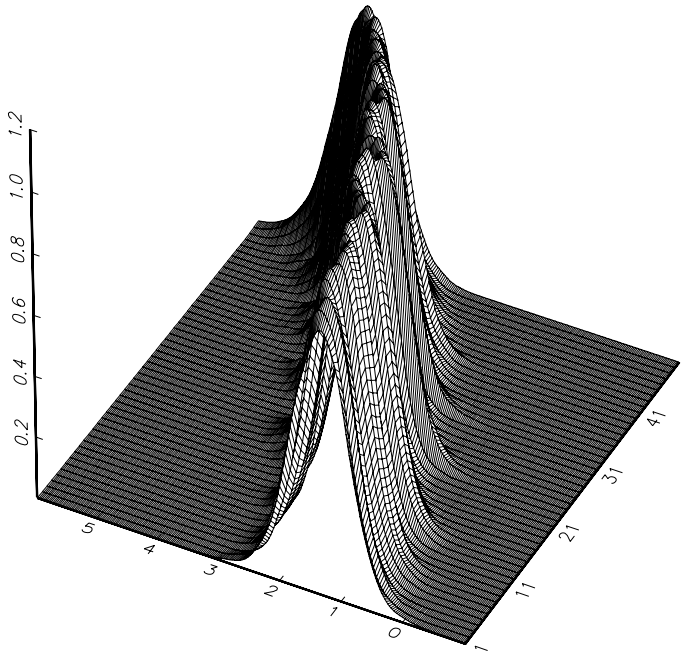

Figure 3: Estimated marginal posterior densities of the group effect $p(\beta_{t2}|y)$ versus time t.

leading to 3*2 dummy variables, denoted by A^+, $A^=$ (orders in hand), G^+, $G^=$ (expected development of the state of business for the next 6 months) and PE^+, $PE^=$ (production plans of the previous month) with "decrease" as the reference category. The covariate vector x_{ti} consists of these six dummy variables.

To illustrate the flexibility of the model, we will discuss several model approaches with or without unit–specific parameters. They all use the fact, that a dynamic cumulative model for ordered response in $q + 1 = 3$ categories can be embedded in a multivariate DGLM (see Fahrmeir & Tutz, 1994, for more details) through the following specification: Response y_{ti} is multinomially distributed

$$y_{ti} \sim M_2(1, \pi_{ti}),$$

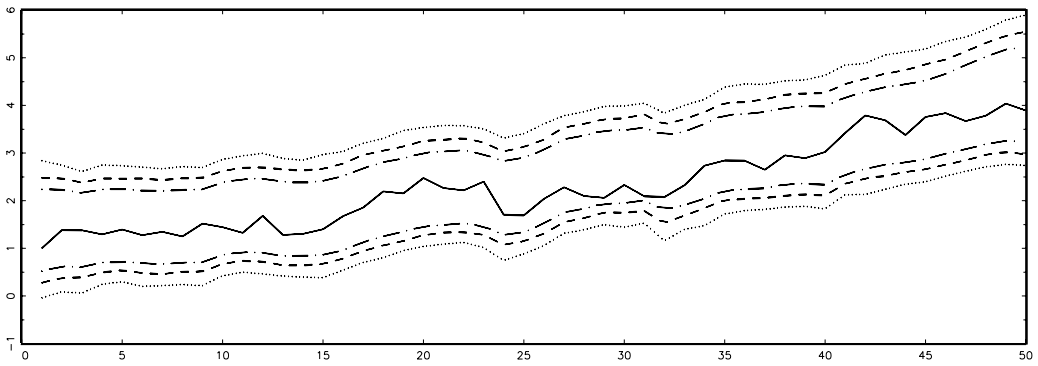

Figure 4: The group effect (solid line), 50 %, 80 % and 95 % simultaneous credible regions.

where $y_{ti} = (1,0)^T$, $(0,1)^T$ or $(0,0)^T$, if the first, second or third category is observed, respectively. The response function is given by

$$h(\eta_{ti}) = \begin{pmatrix} F(\eta_{ti1}) \\ F(\eta_{ti2}) - F(\eta_{ti1}) \end{pmatrix},$$

where F must have all properties of a distribution function. Here $F(x) = 1/\{1 + \exp(-x)\}$ is used leading to dynamic versions of the cumulative logit model.

Assuming no existence of unit–specific heterogeneity and a random walk of first order for all components of β_t, we obtain the model

$$\pi_{ti} = h(\eta_{ti}) = h(Z_{ti}\beta_t), \quad \beta_t = \beta_{t-1} + u_t,$$

where the design matrix Z_{ti} is given by

$$Z_{ti} = \begin{pmatrix} 1 & 0 & x_{ti}^T \\ 0 & 1 & x_{ti}^T \end{pmatrix}.$$

Note that the first two components of β_t represent the threshold parameters θ_{t1} and θ_{t2}. They have to follow the restriction $\theta_{t1} < \theta_{t2}$ for all t.

Going one step further, we introduce a category unspecific random effect b_i. Then the model above is extended to

$$\pi_{ti} = h(Z_{ti}\beta_t + W_{ti}b_i),$$

where $W_{ti} = (1,1)^T$. The threshold restriction $\theta_{t1} + b_i < \theta_{t2} + b_i$ for all t and i boils down to the simpler form given above.

Category–specific random effects may be more flexible and are easily integrated by choosing

$$W_{ti} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The two components of b_i represent the unit–specific deviations from the two threshold parameters θ_{t1} and θ_{t2}. Now the restriction $\theta_{t1} + b_{i1} < \theta_{t2} + b_{i2}$ for all t and i cannot be simplified.

A realistic data analysis must assume, that this monthly data shows strong seasonality with period 12. Although a random walk for both thresholds θ_{t1} and θ_{t2} will somehow reflect a seasonal pattern, a decomposition into trend and season $\theta = \tau + \gamma$ is more appropriate. Here we assume a random walk of first order for both trend components and a flexible seasonal model $\gamma_{t-11} + \ldots + \gamma_t = u_t$, u_t white noise, for both season components. The design matrix Z_{ti} now changes to

$$Z_{ti} = \begin{pmatrix} 1 & 1 & 0 & 0 & x_{ti}^T \\ 0 & 0 & 1 & 1 & x_{ti}^T \end{pmatrix},$$

the linear predictor is

$$\eta_{ti} = \begin{pmatrix} \tau_{t1} + \gamma_{t1} + b_{i1} + x_{ti}^T\widetilde{\beta}_t \\ \tau_{t2} + \gamma_{t2} + b_{i2} + x_{ti}^T\widetilde{\beta}_t \end{pmatrix},$$

where $\widetilde{\beta}_t = (\beta_{t5}, \ldots, \beta_{tp})^T$. Formally the transition model is given by

$$\beta_t = (\gamma_{t1}, \tau_{t1}, \gamma_{t2}, \tau_{t2}, \beta_{t5}, \ldots, \beta_{tp})^T = -\sum_{i=1}^{11} F_i\beta_{t-i} + u_t,$$

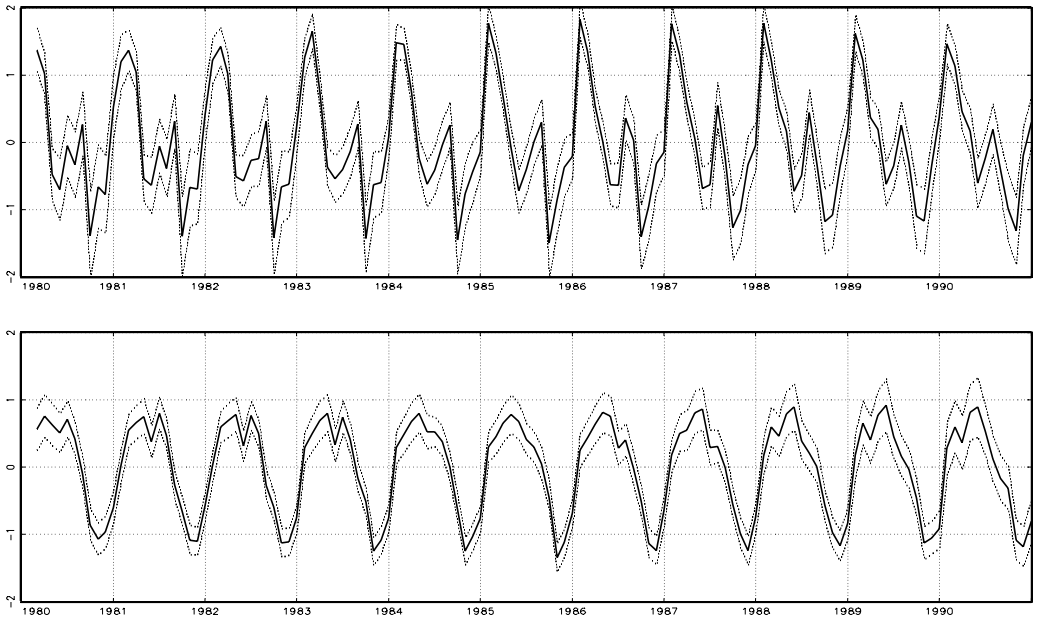

Figure 5: Posterior mean estimates (solid line) and pointwise one posterior standard deviation confidence band (dotted lines) of the seasonal components of the first (above) and second threshold parameter.

where $F_1 = \text{diag}(-1, +1, -1, +1, -1, -1, \ldots, -1)$ and $F_2 = \ldots = F_{11} = \text{diag}(0, +1, 0, +1, 0, 0, \ldots, 0)$.

We run the procedure for the latter model with a single run of length 52,500 cycles, discarding the first 2,500 and using every 5th sample thereafter. We specified the priors for the hyperparameters as follows: expectation of the inverted Wishart prior for D was set to $\text{diag}(0.5, 0.5)$ with standard deviation equal to 0.5 for the diagonal elements. The inverse gamma priors for the elements of Q had expectation and standard deviation 0.1 for the seasonal components and 0.01 for the others.

Figure 5 shows the estimated seasonal components within pointwise 1 STD confidence intervals. A strong seasonal pattern can be seen with highs in spring and in August, whereas the estimates of the two trend parame-

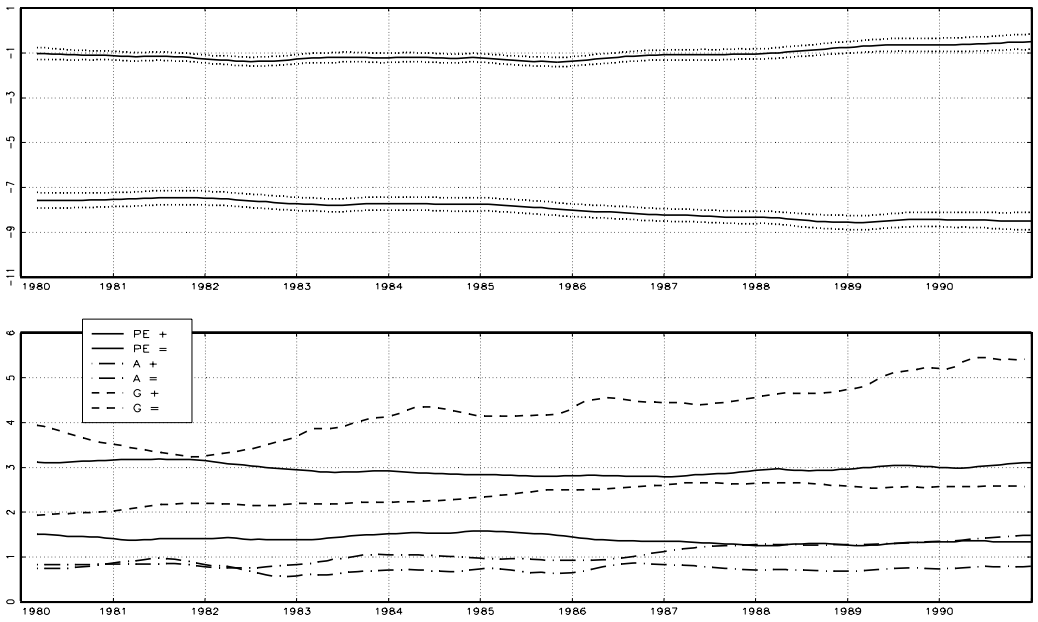

Figure 6 (top): Posterior mean estimates (solid line) and pointwise one posterior standard deviation confidence band (dotted lines) of the trend components of the first and second threshold parameter.

Figure 7 (bottom): Posterior mean estimates of covariate effects.

ters (Figure 6) are nearly time–constant. Posterior mean estimates of time–dependent parameters (Figure 7) correspond to the results in Fahrmeir & Nase (1994) and Knorr–Held (1995), obtained with different methods (and without unit–specific parameters). Figure 8 shows 50 %, 80 % and 95 % simultaneous credible regions for those parameters. Only the dummy for expected increase of "expected development of the state of business for the next 6 months" shows a significant temporal variation with a low around 1982, when a new government was established in Germany. From that time on the effect is increasing and may be interpreted as a growing trust in the government.

Posterior mean estimates of the unit–specific parameters are shown in Figure 9, where estimates of the first and second unit–specific effect, cor-

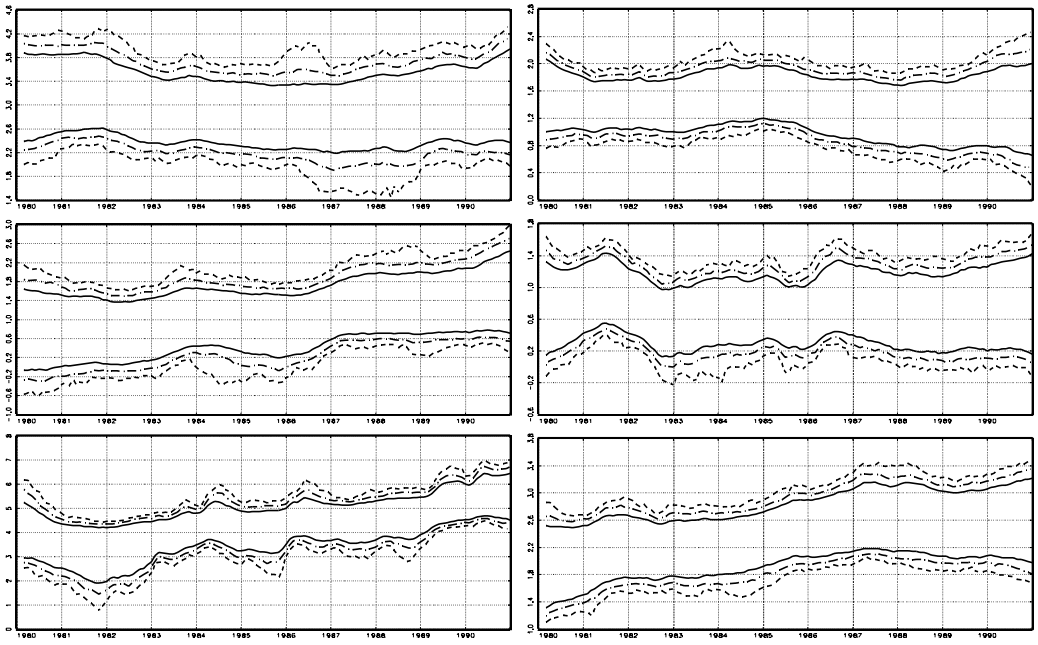

Figure 8: 50 %, 80 % and 95 % simultaneous credible regions for covariate effects of PE+, PE=, A+, A=, G+ and G=.

responding to the first and second threshold parameter, are plotted against each other for all 55 firms. Interestingly, these two effects are highly correlated (estimated correlation in Q is –0.50) and the following interpretation seems to be plausible: Some firms are more conservative in their answers, often choosing "no change" for the response variable while others often answer with the categories "decrease" or "increase". Finally Figure 10 gives the estimates plotted separately against corresponding estimated standard deviations. We observe the (reasonable) result, that the more the estimates tend "to the middle" (positive for the first and negative for the second), the more precise they are (measured in posterior standard deviation).

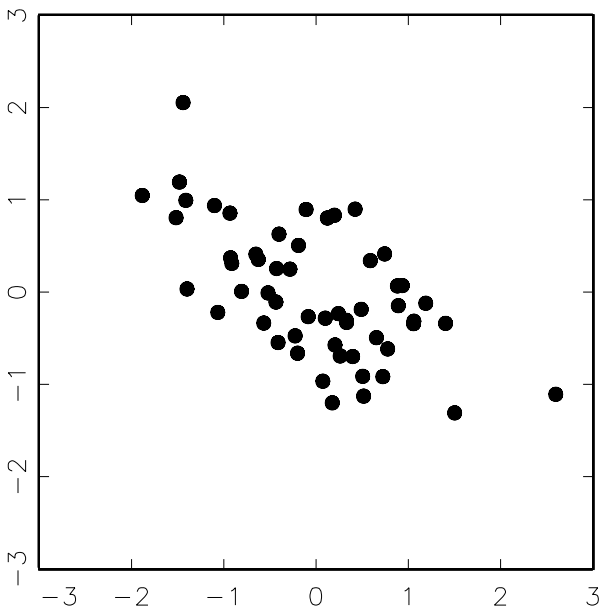

Figure 9: Posterior mean estimates of the unit–specific parameters $\hat{b}_{i1}$ and $\hat{b}_{i2}$, plotted against each other for every unit.

5 Concluding Remarks

The major advantage of MCMC as a statistical inference technique is its provided model flexibility together with implementation simplicity. The approach in this article is a convincing example for this duality. However, the flexibility of MCMC is not yet exhausted; possible extensions of dynamic generalized linear mixed models include

- the introduction of time–constant components of β_t,
- the use of robust mixtures of normals for the Gaussian error terms, see Carlin, Polsen & Stoffer (1992),
- nonlinear models,
- extensions to spatial data, where the dependence of the (now space–dependent) parameters β_t is modelled through Markov random fields

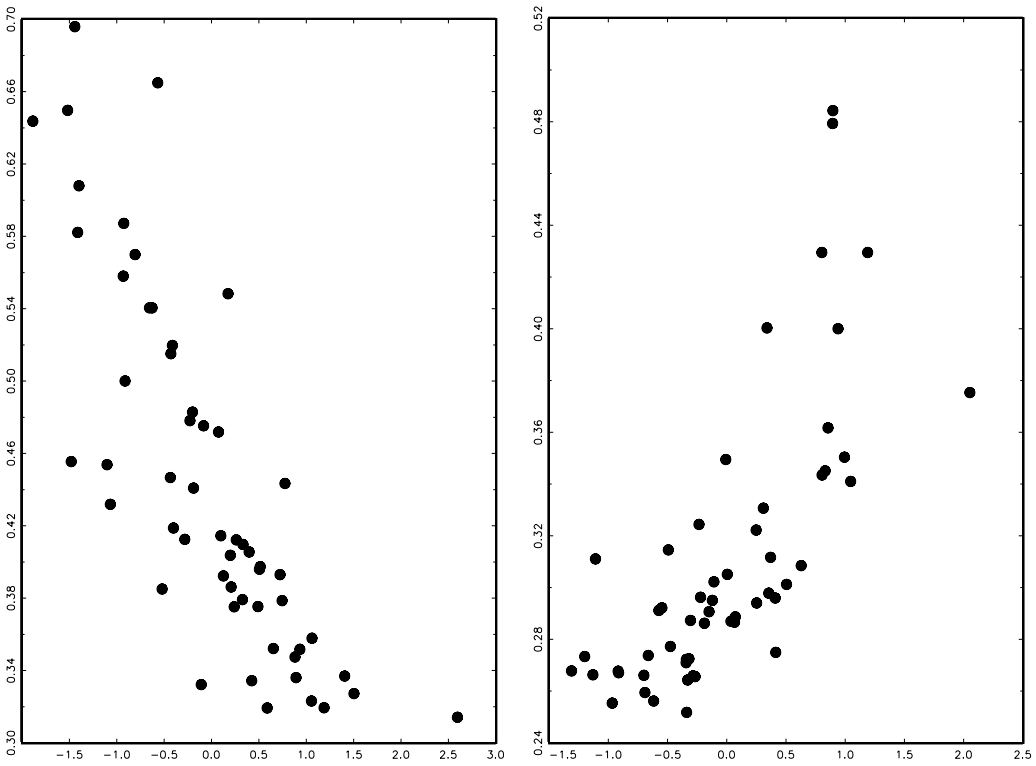

Figure 10: Posterior mean estimate (horizontal) versus posterior standard deviation estimate (vertical) for first (left) and second unit–specific parameter.

(Besag, York & Mollie, 1991; Besag, Green, Higdon & Mengersen, 1995).

It seems that nowadays the complexity of statistical models is no longer limited by the ability of inference techniques but (more natural) by the amount of information given in the data, since too complex models often lead to serious identification problems. It lies in the responsibility of the statistician to find a compromise between parsimony and complexity. However, sensitivity analysis and model selection by MCMC may help to derive an appropriate model. For further details see Raftery (1995), Chib (1995), Green (1995) and Besag, Green, Higdon & Mengersen (1995).

References

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc.* B **36**, 192–236.

Besag, J. E., Green, P. J., Higdon, D. & Mengersen, K. (1995) Bayesian computation and stochastic systems (with discussion). *Statist. Sci.* to appear.

Besag, J. E., York, J. & Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Ann. Inst. Statist. Math.* **43**, 1–59.

Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Am. Statist. Assoc.* **88**, 9–25.

Carlin, B. P., Polsen, N. G. & Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space-modeling. *J. Am. Statist. Assoc.* **87**, 493-500.

Chib, S. (1995). Marginal likelihood from the Gibbs output. Preprint, Washington University.

Fahrmeir, L. (1992a). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *J. Am. Statist. Assoc.* **87**, 501-509.

Fahrmeir, L. (1992b). State space modeling and conditional mode estimation for categorical time series. In *New Directions in Time Series Analysis*, Ed. D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M. S. Taqqu, 87–109. New York: Springer–Verlag.

Fahrmeir, L. & Nase, H. (1994). Dynamische Modellierung und Analyse von Mikrodaten des Konjunkturtests. *ifo Studien*, **40**, 1-22.

Fahrmeir, L. & Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models.* New York: Springer–Verlag.

Gilks, W. R. (1992). Derivative–free adaptive rejection sampling for Gibbs sampling. In *Bayesian Statistics 4*, Ed. J. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith, 641–649. Oxford University Press.

Gilks, W. R., Best, N. G. & Tan, K. K. C. (1994). Adaptive rejection Metropolis sampling. to appear.

Gilks, W. R. & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* **41**, 337–348.

Green, P. J. (1995). Reversible jump MCMC computation and Bayesian model determination. Preprint, University of Bristol.

Hsiao, C. (1986). *Analysis of Panel–Data.* Cambridge: Cambridge University Press.

Knorr–Held, L. (1995). Dynamic cumulative probit models for ordinal panel–data; A Bayesian analysis by Gibbs sampling. Discussion paper Nr. 2, Sonderforschungsbereich 386 der Ludwig–Maximilians–Universität München.

Raftery, A.E. (1995). Hypothesis testing and model selection via posterior simulation. In *Practical Markov Chain Monte Carlo* , Ed. W. R. Gilks, D. J. Spiegelhalter and S. Richardson. London: Chapman and Hall, to appear.

Smith, A. F. M. & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *J. R. Statist. Soc.* B **55**, 3-23.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **21**, to appear.

Zeger, S. L. & Karim, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *J. Am. Statist. Assoc.* **86**, 79-86.