



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Knorr-Held:

Conditional Prior Proposals in Dynamic Models

Sonderforschungsbereich 386, Paper 36 (1996)

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

Projektpartner



Conditional Prior Proposals in Dynamic Models

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

Abstract

Dynamic models extend state space models to non–normal observations. This paper suggests a specific hybrid Metropolis–Hastings algorithm as a simple, yet flexible and efficient tool for Bayesian inference via Markov chain Monte Carlo in dynamic models. Hastings proposals from the (conditional) prior distribution of the unknown, time–varying parameters are used to update the corresponding full conditional distributions. Several blocking strategies are discussed to ensure good mixing and convergence properties of the simulated Markov chain. It is also shown that the proposed method is easily extended to robust transition models using mixtures of normals. The applicability is illustrated with an analysis of a binomial and a binary time series, known in the literature.

Key words: Bayesian computing; Blocking; Conditional prior proposal; Discrete data; Dynamic models; Generalized linear model; Innovative outliers; Markov chain Monte Carlo.

¹Email: leo@stat.uni-muenchen.de. Last corrections July 23, 1996.

1 Introduction

Markov chain Monte Carlo (MCMC) simulation in dynamic models with non-normal observation model is an ongoing problem. Such dynamic models relate observations y_t , $t = 1, \dots, T$, to unobserved state parameters α_t with a so called observation model, typically a generalized linear model. Temporal dependence is modelled within a transition model, an autoregressive Gaussian prior for the latent parameters $\alpha = (\alpha'_1, \dots, \alpha'_T)'$. Hyperparameters are included in a third level of hierarchy and some conditional independence assumptions complete the model specification.

Such models are known as state space models if the observations y_t are Gaussian. MCMC simulation in state space models is discussed in several papers. Carlin, Polsen & Stoffer (1992) discuss Gibbs sampling and update α_t with a sample from the corresponding full conditional. However, Carter & Kohn (1994) and Fruehwirth–Schnatter (1994) observe bad mixing and convergence behavior in such a “single move” blocking strategy. They propose to update α all at once instead, again using a Gibbs step, i. e. a sample from the (now high dimensional) full conditional. Special properties of this Gaussian distribution ensure an efficient algorithm.

Corresponding work for the more general class of dynamic (generalized linear) models is rather rudimentary; the full conditionals are now fundamentally non-Gaussian due to the non-Gaussian observation model. Knorr–Held (1995) uses a specific Hastings proposal to update the α_t 's one at a time, when there is a large number of observations y_{ti} for a given time t . The resulting algorithm is appealing due to its simplicity and flexibility. However, it may show signs of slow mixing and convergence when the prior is tight relative to the information of the observation y_t , such as for binary time series.

Gamerman (1995) tries to counter this problem through a reparameterization of the model to a priori independent system disturbances. A Gaussian Hastings proposal, based on an approximation of the full conditional with additional Fisher scoring type steps, is used. Gamerman reports considerably improved mixing and convergence behavior. However, the simple structure of the full conditional is destroyed, leading to an algorithm of quadratic computa-

tional complexity in T . The algorithm also involves several evaluations of first and second order derivatives of the observation model for every updating step.

Also Shephard & Pitt (1995) use “2 to 5” Fisher scoring type steps and analytic Taylor expansions to calculate the moments of a Gaussian Hastings proposal that tries to approximate a specific full conditional. In contrast to Gamerman, they propose to divide α into several blocks (“block move”) as an intermediate strategy between updating α one at a time and all at once.

Both algorithms have in common proposals which try to approximate the full conditional, imitating a Gibbs step with acceptance probability close to 1. However, this is not necessary at all, as already mentioned in Besag, Green, Higdon & Mengersen (1995). For example, the widely used Metropolis updating step has optimal performance for average acceptance rates below 0.5 (Gelman, Roberts & Gilks, 1995).

For MCMC simulation in dynamic models we propose a specific proposal that reflects the autoregressive prior specification but is independent of the chosen observation model. The resulting algorithm is conceptually simple, since all proposals are Gaussian with known moments. Furthermore, it is derivative-free, which is a major advantage concerning both implementation and computation time. Updating is done within a certain blocking strategy to ensure good mixing and convergence of the simulated Markov chain. Tuning of the algorithm is done by choosing a block configuration, rather than the spread of the proposal as in the Metropolis case.

The next section reviews dynamic models as a useful framework for the analysis of categorical time series or panel-data. MCMC simulation by conditional prior proposals is discussed in Section 3. Some simulation results are given for a dataset, known to be problematic for the single move algorithm. Finally, extensions of the transition model to errors within the class of t -distributions are discussed in Section 4. Such models allow abrupt jumps in the transition model, also known as innovative outliers. As a final example, we analyze a binary time series with an additional hyperprior on the degrees of freedom of the t -distribution.

2 Dynamic Models

Let $y = (y_1, \dots, y_T)$ denote the sequence of observations and $\alpha = (\alpha'_1, \dots, \alpha'_T)'$ the sequence of state parameters. We assume that $p(\alpha_t | \alpha_{-t}, Q_t)$ ($t = z + 1, \dots, T$) has a Gaussian distribution with mean $-F_1\alpha_{t-1} - F_2\alpha_{t-2} - \dots - F_z\alpha_{t-z}$ and dispersion Q_t . Here α_{-t} denotes the sequence $(\alpha'_{t-z}, \dots, \alpha'_{t-1})'$, the matrices F_1, \dots, F_z are assumed to be known. So

$$p(\alpha | Q) = \prod_{t=z+1}^T p(\alpha_t | \alpha_{-t}, Q_t)$$

denotes a Gaussian (vector) autoregressive prior of lag z for α . It is often called the transition model. We write short Q for the sequence of dispersions Q_{z+1}, \dots, Q_T ; if Q is assumed to be time-constant we just write $Q \equiv Q_t$.

Conditional independence of $y_t | \alpha_t$, $t = 1, \dots, T$, leads to the following posterior distribution:

$$p(\alpha, Q | y) \propto \prod_{t=1}^T p(y_t | \alpha_t) \times p(\alpha | Q) \times p(Q);$$

here $p(Q)$ is some hyperprior, independent of α and y .

Typical examples of such transition models (with time-constant variance Q) are first ($z = 1$) and second ($z = 2$) order random walks

$$\begin{aligned} \alpha_t | \alpha_{-t}, Q &\sim N(\alpha_{t-1}, Q) \\ \alpha_t | \alpha_{-t}, Q &\sim N(2\alpha_{t-1} - \alpha_{t-2}, Q) \end{aligned}$$

or seasonal models $\alpha_t | \alpha_{-t}, Q \sim N(-\alpha_{t-1} - \alpha_{t-2} - \dots - \alpha_{t-z}, Q)$ with period $z + 1$.

It is always possible to write Gaussian autoregressive priors within a penalty formulation

$$p(\alpha | Q) = \exp\left(-\frac{1}{2}\alpha'K\alpha\right).$$

Note, that $p(\alpha | Q)$ is improper due to implicitly assumed diffuse priors for the initial parameters $\alpha_1, \dots, \alpha_z$; therefore K^{-1} does not exist. For the random walks given above the

corresponding penalty matrices are

$$K = \frac{1}{Q} \begin{pmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix}$$

and

$$K = \frac{1}{Q} \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & 1 & & & \\ & 1 & -4 & 6 & -4 & 1 & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & 1 & -4 & 6 & -4 & 1 \\ & & & & 1 & -4 & 6 & -4 & 1 \\ & & & & & 1 & -4 & 5 & -2 \\ & & & & & & 1 & -2 & 1 \end{pmatrix},$$

respectively.

The penalty matrix K plays a key role in the derivation of the conditional distribution of a subvector of α . Defining

$$F = \begin{pmatrix} F_z & F_{z-1} & \dots & F_1 & F_0 & & & \\ & F_z & F_{z-1} & \dots & F_1 & F_0 & & \\ & & \ddots & & & & \ddots & \\ & & & F_z & F_{z-1} & \dots & F_1 & F_0 \end{pmatrix}$$

and the block-diagonal matrix

$$Q = \begin{pmatrix} Q_{z+1} & & & & \\ & Q_{z+2} & & & \\ & & \ddots & & \\ & & & & Q_T \end{pmatrix},$$

it follows that $K = F'Q^{-1}F$.

Since Q is symmetric, so is K . Furthermore, it can be shown that the elements of

$$K = \begin{pmatrix} k_{11} & k_{12} & \dots & k_{1T} \\ k_{21} & k_{22} & & \\ \vdots & & \ddots & \vdots \\ k_{T1} & \dots & k_{T,T-1} & k_{TT} \end{pmatrix}$$

are given by

$$k_{t,t+s} = \sum_{j=\max(0,s,1+z-t)}^{\min(z,z+s,T-t)} F_j'Q_{t+j}^{-1}F_{j-s}.$$

Note that K has zero blocks outside the z off diagonals.

Applications of dynamic models are widespread. Fahrmeir & Tutz (1994a) discuss smoothing of categorical time series, panel- and survival data. Fahrmeir & Tutz (1994b) introduce dynamic models for ordered paired comparison data. Duration data is covered in Fahrmeir & Knorr-Held (1996). Breslow & Clayton (1993) and Clayton (1996) discuss biostatistical applications with second order random walk priors in mixed models, which is somewhat related. Berzuini & Clayton (1994) propose second order random walk priors in survival models with multiple time scales. Also Besag, Green, Higdon & Mengersen (1995) use second order random walk priors in age-period-cohort models.

Most of the references above have bi- or multinomial logistic or log-linear Poisson models in the observation model. If several units $i = 1, \dots, n_t$ are observed, then $p(y_t|\alpha_t) = \prod_{i=1}^{n_t} p(y_{ti}|\alpha_t)$ is usually assumed by suitable additional conditional independence assumptions.

3 MCMC Simulation with Conditional Prior Proposals

Our MCMC implementation is based on updating using full conditionals as described in full detail in Besag, Green, Higdon & Mengersen (1995); we also use their terminology. We denote full conditionals by $p(\alpha_t | \cdot)$, for example. We start this Section with a technical note about

the conditional distribution of a subvector $\alpha_a, \dots, \alpha_b$, given $\alpha_1, \dots, \alpha_{a-1}$ and $\alpha_{b+1}, \dots, \alpha_T$. Then the single and the block move with conditional prior proposals is introduced. Blocking strategies, necessary for the implementation of the block move, are sketched. We close with some simulation results.

3.1 Conditional properties of autoregressive priors

The conditional distribution of a subvector of α , given the rest of α plays a key role in our algorithm. Let α_{ab} denote the subvector $(\alpha_a, \alpha_{a+1}, \dots, \alpha_b)$ and K_{ab} denote the submatrix out of K , given by the rows and columns a to b . Finally, let $K_{1,a-1}$ and $K_{b+1,T}$ denote the matrix left and right of K_{ab} , respectively:

$$K = \begin{pmatrix} & & K'_{1,a-1} \\ K_{1,a-1} & K_{ab} & K_{b+1,T} \\ & K'_{b+1,T} & \end{pmatrix}$$

Then the following result can be proved by simple matrix manipulations: The conditional distribution of α_{ab} , given $\alpha_{1,a-1}$ and $\alpha_{b+1,T}$ is normal $N(\mu_{ab}, \Sigma_{ab})$ with moments

$$\mu_{ab} = \begin{cases} -K_{ab}^{-1} K_{b+1,T} \alpha_{b+1,T} & a = 1 \\ -K_{ab}^{-1} K_{1,a-1} \alpha_{1,a-1} & b = T \\ -K_{ab}^{-1} (K_{1,a-1} \alpha_{1,a-1} + K_{b+1,T} \alpha_{b+1,T}) & \text{otherwise} \end{cases}$$

and $\Sigma_{ab} = K_{ab}^{-1}$.

Note that apart from hyperparameters, only $\alpha_{a-z}, \dots, \alpha_{a-1}$ and $\alpha_{b+1}, \dots, \alpha_{b+z}$ enter in μ_{ab} , since all blocks in K outside the z off-diagonals are zero.

3.2 Single move

The most natural blocking strategy for α is to update α_t one at a time. The main advantage is that the full conditional has a simple form, achieved by the hierarchical structure of the model:

$$p(\alpha_t | \cdot) \propto p(y_t | \alpha_t) p(\alpha_t | \alpha_{s \neq t}, Q).$$

One way to update α_t is to use a proposal α_t^* , distributed as $p(\alpha_t|\alpha_{s \neq t}, Q)$. Such a “conditional prior proposal” is independent of the current state of α_t but, in general, depends on the current states of all other parameters (here $\alpha_{s \neq t}$ and Q). Note, that “Gibbs proposals”, i.e. samples from the full conditional, have exactly the same “conditional independence” property.

It is illustrative to discuss differences between conditional and unconditional independence proposals (Tierney, 1994). It is often very difficult, at least for higher dimensions and non-normal models, to construct an (unconditional) independence proposal with good acceptance rates. In contrast, a conditional independence proposal depends on the current state of neighboring parameters, it is therefore far more constrained than the unconditional version, being already in the right part of the state space. On the other hand, its distribution changes in every iteration step (if neighboring parameters are updated and accepted), it is therefore still very flexible (Unconditional independence proposals are generated from exactly the same distribution in every iteration step).

The Hastings acceptance probability simplifies for the conditional prior proposal to

$$\min \left\{ 1, \frac{p(y_t|\alpha_t^*)}{p(y_t|\alpha_t)} \right\},$$

the likelihood ratio for observation y_t . Conditional prior proposals have a natural interpretation: α_t^* is drawn independently of the observation model and just reflects the specific autoregressive prior specification. If it produces improvement in the likelihood at time t , it will always be accepted, if not, then the acceptance probability is equal to the likelihood ratio. The α_t 's should be visited in random order to avoid an artificial drift of the simulated Markov chain.

Of course, a simple random walk proposal can be used instead, but it has to be tuned. Other single move updating schemes are more demanding in their proposals and therefore are likely to be slower in CPU time, slower also due to the fact that the computation of the acceptance probability gets more complicated. Gamerman (1995) and Shephard & Pitt (1995) construct proposals that try to approximate the full conditional using additional Fisher scoring steps and Taylor approximations. These procedures involve the evaluation of score functions and

information matrices at every update step. A real Gibbs step requires additional iterations via a rejection sampling procedure and is likely to be inefficient.

However, the single move blocking scheme might be very slow converging, especially if neighboring parameters are highly correlated. This is typically the case when the likelihood at time t is very flat in α_t and does not give much information relative to the autoregressive prior specification. Binary time series are a typical example. A simple modification of the single move conditional prior algorithm addresses this problem without losing its simplicity both in programming and computing time.

3.3 Block move

Instead of updating one parameter α_t at a time, the block move is based on updating one block $\alpha_{rs} = (\alpha_r, \dots, \alpha_s)$ at a time. The number of blocks may range from 2 up to T , which corresponds to the single move. Consider the breakpoints that divide α into blocks as fixed for the moment. The idea of this blocking strategy is to use blocks that are large enough, so that the corresponding likelihood provides enough information to ensure a good mixing and convergence behavior. So what kind of proposals are useful for the block move?

It is generally not clear how to choose the spread of a multivariate Metropolis proposal. But, in contrast, the generalization of the conditional prior proposal is straightforward: The simple structure of the full conditional is retained, since $p(\alpha_{rs} | \alpha_{1,r-1}, \alpha_{s+1,T}, Q)$ is still normal with known moments (see Section 3.1). Therefore a conditional prior proposal can be implemented similarly as in the previous section (using the Cholesky decomposition): Generate α_{rs}^* distributed as $p(\alpha_{rs} | \alpha_{1,r-1}, \alpha_{s+1,T}, Q)$ to update the full conditional

$$p(\alpha_{rs} | \cdot) \propto \prod_{t=r}^s p(y_t | \alpha_t) p(\alpha_{rs} | \alpha_{1,r-1}, \alpha_{s+1,T}, Q).$$

Note, that both for the single and the block move, the conditional prior distribution $p(\alpha_{rs} | \alpha_{1,r-1}, \alpha_{s+1,T}, Q)$ depends on not more than $2 * z$ state parameters and the hyperparameter Q , a convenient fact for implementation of the conditional prior proposal. The

acceptance probability simplifies again to a likelihood ratio

$$\min \left\{ 1, \frac{\prod_{t=r}^s p(y_t | \alpha_t^*)}{\prod_{t=r}^s p(y_t | \alpha_t)} \right\}.$$

Typically a bigger block size coincidences with smaller acceptance rates, since the likelihood is more informative for an increasing number of units. Shephard & Pitt (1995) propose a different proposal in the block move. It is again based on a Taylor approximation of the full conditional like their version of the single move proposal. Furthermore they propose “2 to 5” additional Fisher scoring iterations to get a reasonable approximation and perform a pseudo rejection sampling step (Tierney, 1994). In contrast, conditional prior proposals benefit of block updating without spending too much effort in constructing appropriate proposals and calculating acceptance probabilities.

3.4 Blocking strategies

The block move provides a considerable improvement in situations where the single move has bad mixing behavior. However, fixed blocks still cause convergence and mixing problems for parameters close to a breakpoint. Changing the block configuration in every iteration cycle is a simple remedy. This can be done either by a deterministic or a random scheme. The random mechanism has to be independent from the MCMC output, though. In all following examples we used random blocking with fixed standard block size. The first block has uniform random block size between 1 and the standard block length. So, most of the updating involves blocks of a fixed block length, which has computational advantages, since the dispersion matrix of the standard block size full conditional can be computed in advance, at least for Gaussian transition models with constant variance. Block sizes proportional to the number of observations n_t per block may also be considered in situations where n_t is changing over time as in survival models (Fahrmeir & Knorr-Held, 1996).

3.5 An example: Tokyo rainfall data

To illustrate the gain of the block move, we analyze the Tokyo rainfall data (e.g. Fahrmeir & Tutz, 1994a), which is known to be problematic for single move algorithms. This data consists of a single binomial time series of length $T = 366$. We assume a binomial logit model

$$y_t | \alpha_t \sim \begin{cases} B(2, \pi_t) & t \neq 60 \\ B(1, \pi_t) & t = 60 \end{cases}, \quad \pi_t = 1/(1 + \exp(-\alpha_t)),$$

with a second order random walk prior for $\{\alpha_t\}$. A highly dispersed, but proper gamma prior was chosen for the random walk precision ($a = 1$, $b = 0.005$). This choice reflects sufficient prior ignorance about Q but avoids problems arising with improper posteriors. Figure 1 displays the data and some characteristics of the posterior distribution of $\{\pi_t\}$.

We separate our empirical analysis into two parts, speed of convergence and efficiency of estimation. First we focus on the empirical convergence behavior. For block size 1, 5, 20 and 40 we computed the average trajectories of 100 parallel chains after 10, 50, 100 and 500 iterations. For every chain, the state parameters were initialized to zero and the variance Q to 0.1. We also computed the average acceptance rate of the Hastings steps, averaged over all α_t 's. Figure 2 shows clear empirical evidence that the block move converges much faster for bigger block sizes, at least for this data set and model. The single move algorithm does not converge at all, at least for the first 500 iterations. The algorithm with blocksize 40 seems to have reached equilibrium after only 50 iterations. The corresponding empirical average acceptance rates have been 99.4 % (block size 1), 94.4 % (size 5), 65.5 % (size 20) and 35.3 % (block size 40).

We repeated the same analysis, assuming a random walk of first order instead. Convergence was a bit faster and, again, the block move algorithm exhibits superior convergence performance.

A measure of efficiency of estimation are the autocorrelations of parameters of the simulated Markov chain after reaching equilibrium. The larger these correlations are, the larger the variances of the estimate of the posterior mean. Intuitively it is clear that other posterior

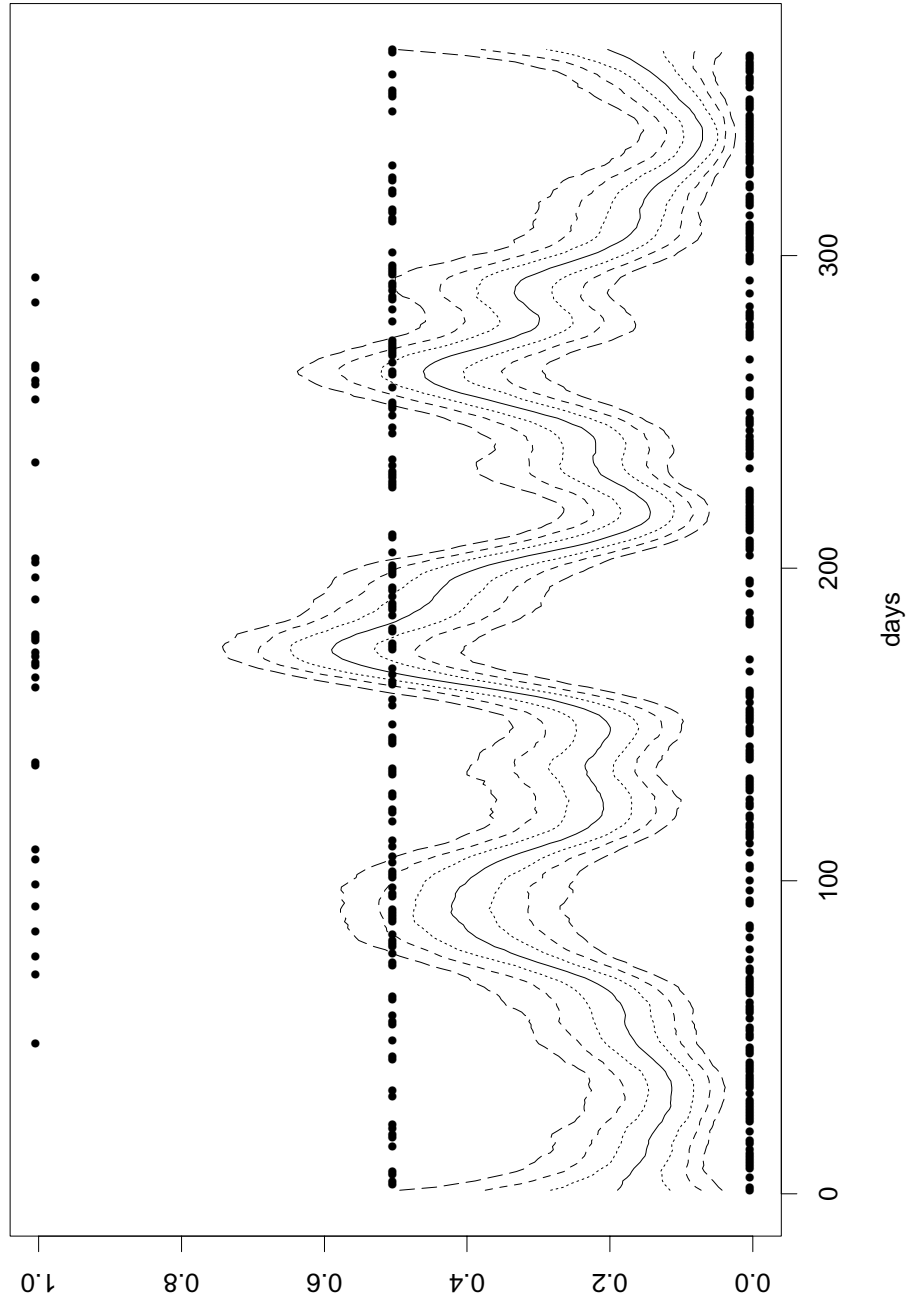


Figure 1: Tokyo rainfall data. Data and fitted probabilities (posterior median within 50, 80 and 95 % credible regions). The data is reproduced as relative frequencies with values 0, 0.5 and 1.

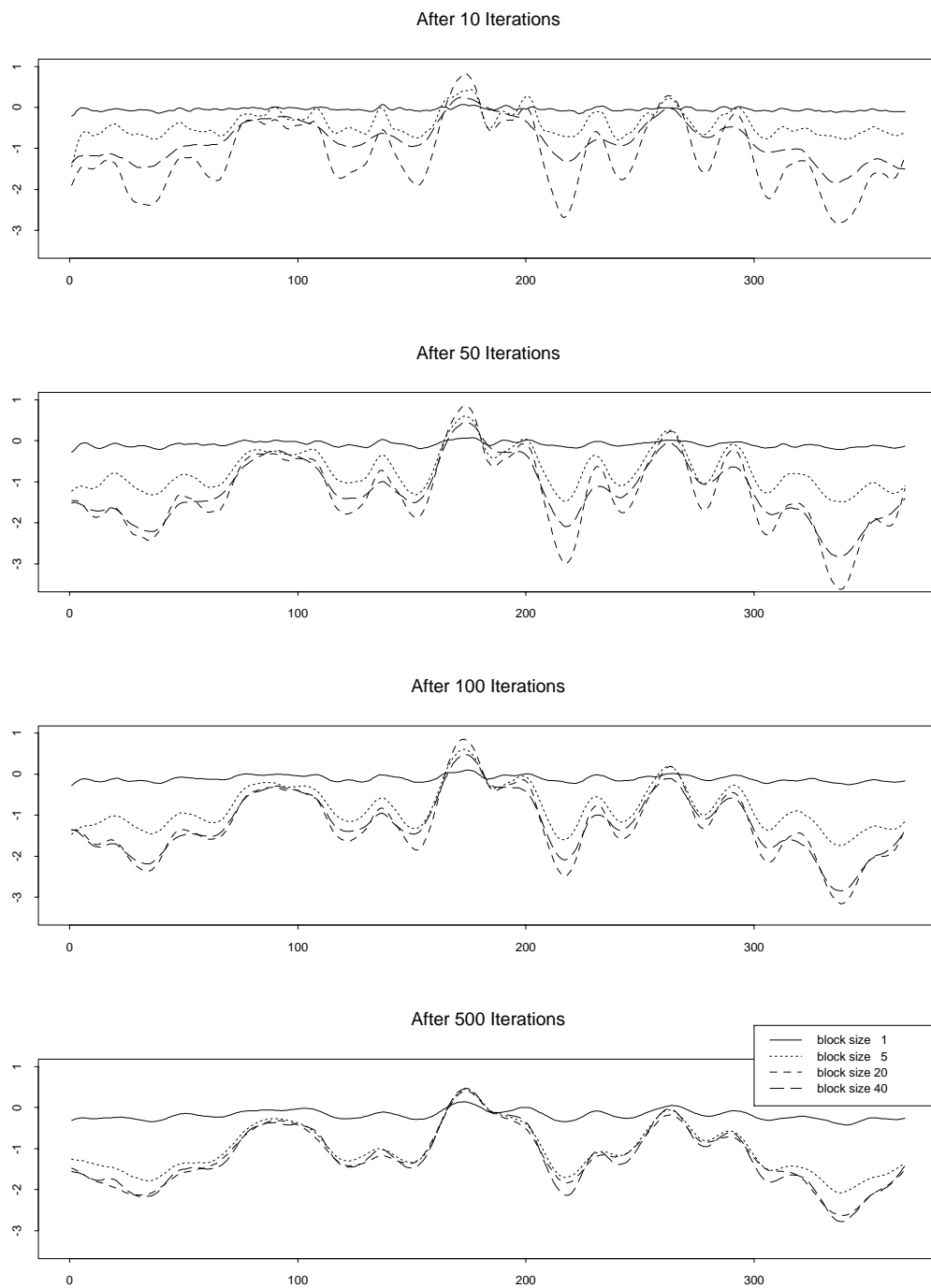


Figure 2: Speed of convergence of the block move algorithm for different block sizes.

characteristics are less efficiently estimated as well. We started the chain in equilibrium, ran it for 1000 iterations and stored every 10th sample until we had 10,000 samples. We calculated autocorrelations for 12 parameters, namely for $t = 1, 33, 67, 100, 133, 167, 200, 233, 267, 300, 333, 366$ and for the hyperparameter Q . We did this analysis twice, for blocksize 1 and blocksize 20, both assuming a second order random walk prior. The results can be summarized as follows: For block size 1, all autocorrelations up to lag 40 of these parameters and hyperparameters were all larger than 0.5. In contrast, for blocksize 20, the autocorrelations of all parameters considered were close to zero for lag 5 and bigger. Autocorrelations for the hyperparameter Q were somewhat bigger (around zero for lag 20 and more) but still much smaller than for blocksize 1.

Figure 3 shows trajectories of the last 2000 iterations for three representative parameters α_1 , α_{100} , α_{333} and the variance Q . Whereas the mixing behavior of the blocksize 1 algorithm is catastrophic, the blocksize 20 algorithm shows well-behaved mixing. The plots for the other parameters look very similar.

4 Incorporating model uncertainty

The temporal variation of underlying parameters may have jumps, so called innovative outliers. The Gaussian distributional assumption in the autoregressive prior, however, does not allow such abrupt movement. Distributions with heavier tails such as t -distributions are more adequate. In this section we will sketch how autoregressive priors can be extended via an hierarchical t -formulation with unknown degrees of freedom (Besag, Green, Higdon & Mengersen, 1995).

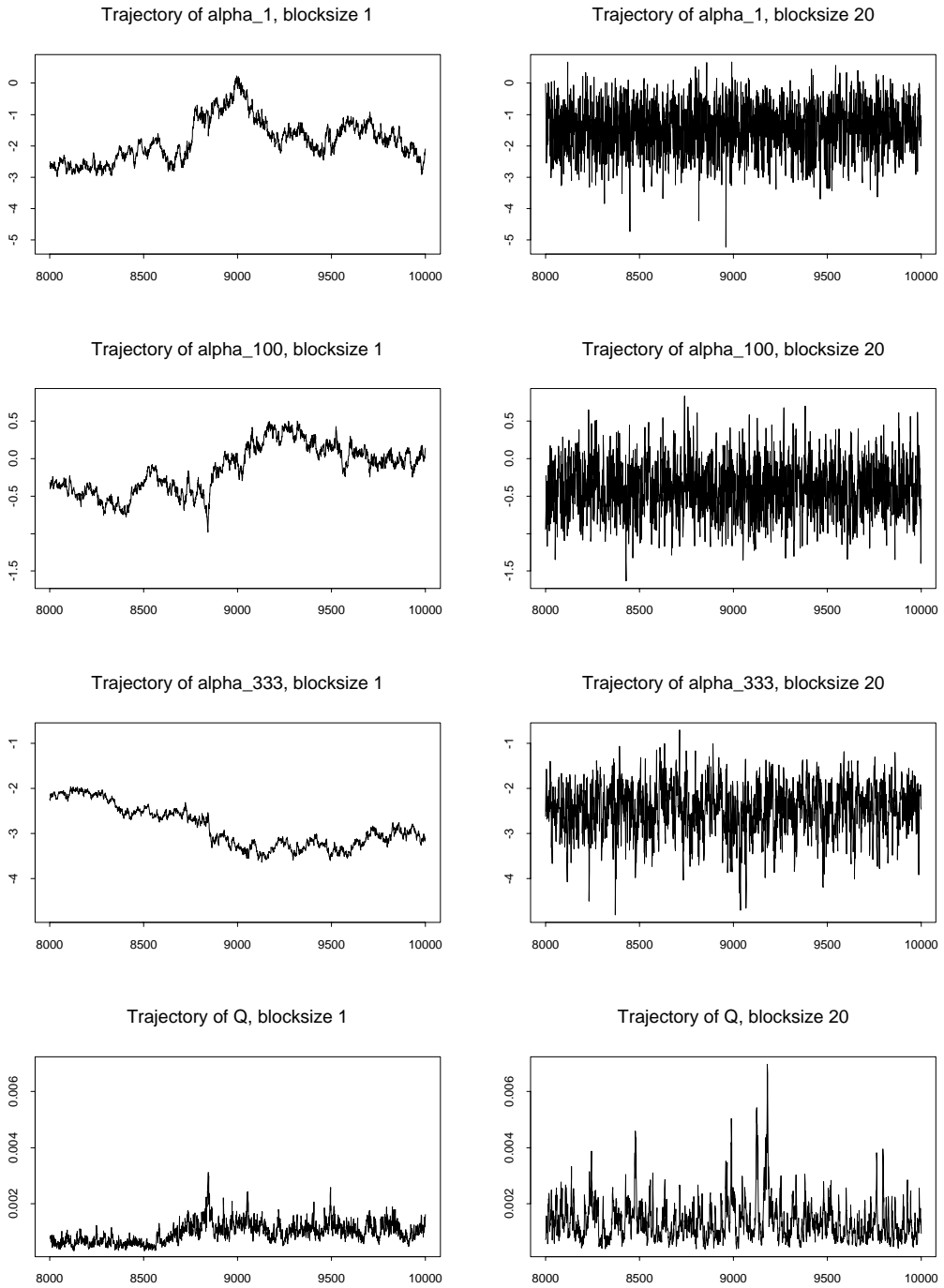


Figure 3: Trajectories of α_1 , α_{100} , α_{333} and Q (from up to down).

4.2 A second example: Sleep data

Carlin & Polson (1992) present an analysis of a binary time series of length $T = 120$ minutes. The outcome variable y_t corresponds to the sleep status (REM ($y_t = 1$) or non-REM) of a specific child. We reanalyze this data to illustrate the hierarchical t -formulation. The response variable is assumed to depend on a latent “sleep status” α_t via a dynamic logistic model. We place an equally weighted hyperprior $p(\nu)$ on the values $\{2^k, k = -1, -0.9, -0.8, \dots, 6.9, 7.0\}$ and assume α_t to follow a hierarchical t random walk. For updating ν , we used a discrete random walk-type proposal which gave equally weight to the two neighbors of the current value. Note that for the limit cases $\nu = 0.5$ and $\nu = 128$, this proposal becomes deterministic, proposing the only neighbor. The acceptance probability has to be modified adequately for proposed jumps to or away from these limit values.

The following analysis is based on a run of length 505000, discarding the first 5000 values and storing every 100th thereafter. The chosen block length was 10 which resulted in an average acceptance rate of 68.6 %. Starting values were zero for all α_t 's. Since the posterior might be multimodal the chain might stay in one part of the posterior for a long time. To account for that we started several chains with different values for ν over the whole range of the prior: 0.5 to 128. However, all of these chains moved after not more than 1000 iterations into the region around $\nu = 1$.

Figure 4 shows the data and estimates. Note that our model formulation gives a significant better fit to the data than the analysis by Carlin & Polson (1992, Figure 1, p. 583). The resulting posterior for the hyperparameter ν has its mode at $\nu = 2^{k=-0.3} \approx 0.81$. The 90 and 95 % credible regions for ν are $[0.66, 3.3]$ and $[0.54, 13.0]$, respectively, showing strong evidence for highly non-normal system disturbances. The estimates of the sequence $\{\alpha_t\}$, the latent sleep status, exhibit some huge abrupt jumps, e.g. around $t = 53$ and $t = 62$. Note that the posterior for some values of t is highly skewed.

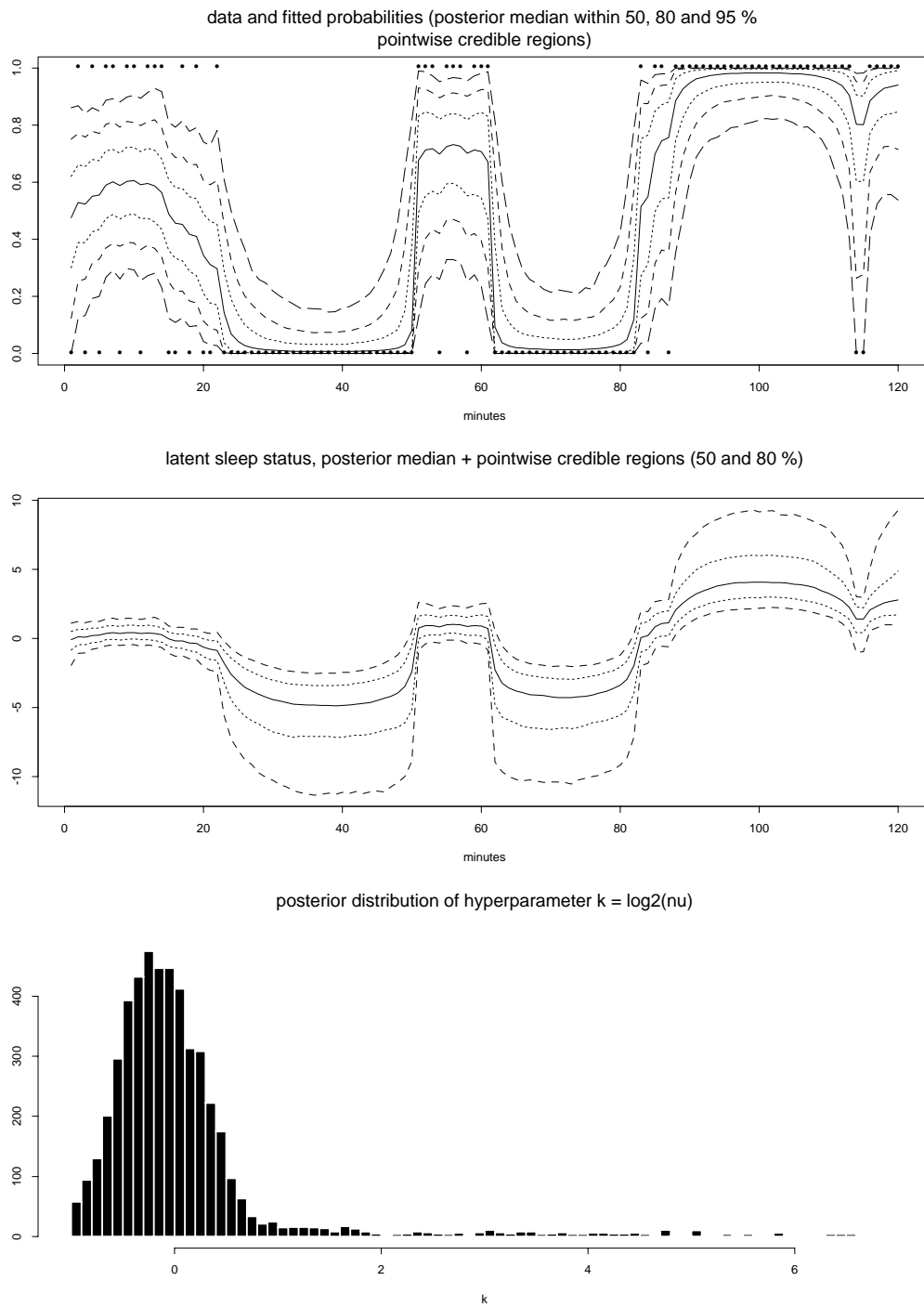


Figure 4: Sleep data. Data and estimates.

5 Discussion

Conditional prior proposals reflect the dependence of underlying parameters and therefore provide a useful tool for highly dependent parameters in dynamic models. The resulting algorithm is appealing since all proposals are easy to generate and all acceptance probabilities are easy to calculate. The choice of a blocking strategy serves as a tuning device.

We also experimented with conditional prior proposals in dynamic models, where $p(\alpha)$ is a product of several autoregressive prior specifications. For example, each component of α_t may correspond to a certain covariate effect (plus intercept) and independent random walk priors are assigned to all components. Here two generalizations are possible: either updating each component within its own blocking strategy or updating all components within one blocking strategy. The former approach provides more flexibility in tuning the algorithm and has been successfully implemented for duration time data. However, the latter is faster, especially for large dimension of α_t and is usually sufficient accurate.

There might also be a wide field of applications in models for non-normal spatial data. Here intrinsic (or undirected) autoregressions replace directed autoregressions. Conditional prior proposals can be implemented in similar lines, since intrinsic autoregressions can be written in a penalty formulation as well, see Besag & Kooperberg (1995).

Acknowledgements

This research was performed during a visit to the Department of Statistics, University of Washington, Seattle, U.S.A., whose hospitality is gratefully acknowledged. The visit was supported by a grant of the German Academic Exchange Service (DAAD). The author likes to thank Julian Besag for frequent discussions. Parts of this work was presented at the *Séminaire Européen de Statistique*, Toulouse May 1996.

References

- Besag, J. E., Green, P. J., Higdon, D. & Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science*, **10**, 3–66.
- Besag, J. E. & Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, **82**, 733–746.
- Berzuini, C. & Clayton, D. (1994). Bayesian Analysis of Survival on Multiple Time Scales. *Statistics in Medicine*, **13**, 823–838.
- Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.
- Carlin, B. P. & Polsen, N. G. (1992). Monte Carlo Bayesian Methods for Discrete Regression Models and Categorical Time Series. In *Bayesian Statistics 4*, Ed. J. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith. Oxford: Oxford University Press, Oxford, 577–586.
- Carlin, B. P., Polsen, N. G. & Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space-modeling. *Journal of the American Statistical Association*, **87**, 493–500.
- Carter, C. K. & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, **81**, 541–553.
- Clayton, D. G. (1996). Generalized Linear Mixed Models. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson & D. J. Spiegelhalter, Eds., London: Chapman & Hall, 275–301.
- Fahrmeir, L. & Knorr-Held, L. (1996). Dynamic discrete time duration models. Discussion paper Nr. 14, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München. ²

²Available under <http://www.stat.uni-muenchen.de/sfb386/publikation.html>

- Fahrmeir, L. & Tutz, G. (1994a). *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer-Verlag.
- Fahrmeir, L. & Tutz, G. (1994b). Dynamic Stochastic Models for Time-dependent Ordered Paired Comparison Systems. *Journal of the American Statistical Association*, **89**, 1438-1449.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**, 183-202.
- Gamerman, D. (1995). Monte Carlo Markov Chains for Dynamic Generalized Linear Models. Unpublished manuscript.
- Gelman, A., Roberts, G. O. & Gilks, W. R. (1995). Efficient Metropolis Jumping Rules. In *Bayesian Statistics 5*, Ed. J. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith. Oxford: Oxford University Press (in press).
- Knorr-Held, L. (1995). Markov Chain Monte Carlo Simulation in Dynamic Generalized Linear Mixed Models. Discussion paper Nr. 8, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München. ³
- Shephard, N. & Pitt, M. K. (1995). Parameter-driven exponential family models. Unpublished manuscript.
- Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, **22**, 1701-1762.

³Available under <http://www.stat.uni-muenchen.de/sfb386/publikation.html>