Knorr-Held:

Dynamic Cumulative Probit Models for Ordinal Panel-Data; a Bayesian Analysis by Gibbs Sampling

Sonderforschungsbereich 386, Paper 2 (1995)

Dynamic Cumulative Probit Models for Ordinal Panel–Data; a Bayesian Analysis by Gibbs Sampling

Leonhard Knorr–Held

Seminar für Statistik

Ludwig–Maximilians–Universität München

Ludwigstr. 33

D–80539 München [1]

Abstract

This paper deals with a dynamic version of the cumulative probit model. A general multivariate autoregressive structure is proposed for modeling the temporal dynamic of both regression and threshold parameters. Conjugate and diffuse prior distributions are used for the variances of the (normally distributed) transition error terms. Introducing latent variables for each ordered categorical observation, statistical inference is done by means of the Gibbs sampler. The applicability is illustrated with two examples. The first analyzes monthly business panel data focusing on the effect of several covariates on a specific ordered response variable. In the second example results of the German soccer league 1993/94 are viewed as response from a dynamic ordered paired comparison system. Here unknown regression parameters corresponding to the underlying time–dependent abilities of the different teams are estimated based on the scores of each game (win–draw–loss).

---

[1] Email: `leo@stat.uni-muenchen.de` Typographical corrections 11 May 1995.

Keywords: Dynamic cumulative probit model; Gibbs sampler; Latent variables; Time–varying regression parameters; Structural time series; Ordered paired comparison system.

1 Introduction

A well–known model for the analysis of ordinal response Y_i, $i = 1, \ldots, n$ in $k = q + 1$ categories $1, \ldots, k$ is the cumulative probit model, given by

$$P(Y_i \leq j | x_i) = \Phi(\gamma_j + x_i'\beta) \qquad j = 1, \ldots, q.$$

The vector of covariates for observation Y_i is x_i, $\gamma = (\gamma_1, \ldots, \gamma_q)'$ is the unknown threshold vector with restriction $\gamma_1 < \gamma_2 < \ldots < \gamma_q$, β is the unknown regression parameter vector and Φ denotes the standard normal distribution function. This model can be derived by introducing latent variables $Z_i = -x_i'\beta + \epsilon_i$, connected to Y_i by

$$Y_i = j \Longleftrightarrow \gamma_{j-1} < Z_i \leq \gamma_j, \quad j = 1, \ldots, k$$

where ϵ_i has standard normal distribution, $\gamma_0 = -\infty$ and $\gamma_k = \infty$.

Recently Albert & Chib (1993) presented a Bayesian analysis of the cumulative probit model by Gibbs sampling. In this paper we will generalize this approach to ordinal panel data Y_{ti}, $t = 1, \ldots, T$, $i = 1, \ldots, n_t$. A time–dependent structure of the unknown parameters γ_t and β_t, $t = 1, \ldots, T$ can be modelled by a general multivariate autoregressive model, including all important members for trend or seasonal modeling like a random walk of any order, a local linear trend model and so on.

This autoregressive model is a generalization of the state space approach for time series given in Carlin et al. (1992) and for categorical time series given in Carlin & Polsen (1993). These authors propose a Gaussian transition equation

$$\beta_t = F\beta_{t-1} + u_t \quad u_t \sim N(0, \Sigma_u) \tag{1}$$

for modeling the dynamic of unknown time–dependent parameters β_t. However, this approach does not allow singular covariance matrices Σ_u of the normally distributed error random variable u_t. Therefore it does not contain important models like a random walk of second order or a simple flexible seasonal component. In contrast, our autoregressive model (AR model)

$$\sum_{l=0}^{z} F_l \beta_{t-l} = u_t \quad u_t \sim N(0, \Sigma_u)$$

covers all these models. For $z = 1$ we obtain an equivalent to model (1).

Let us consider two simple examples of an univariate AR model: A random walk of second order is given by $z = 2$, $F_0 = 1$, $F_1 = -2$ and $F_2 = 1$ and a flexible seasonal component of period 12 is simply given by $z = 11$ and $F_0 = \ldots = F_{11} = 1$. There are two additional differences to the usual state space approach, originating in Kalman Filter methodology. First we have no initial state vector β_0. This is a small but natural change, since there is no data for $t = 0$, so estimation of β_0 is not of main interest. Secondly we put diffuse priors on our initial values $\beta_1, \ldots, \beta_z$. This way of modeling missing prior information is not easily incorporated in the Kalman filter algorithm, one approach is outlined in Ansley & Kohn (1990). It should be mentioned, that this transition model is also useful for a Bayesian analysis of other types of time–dependent data (like nominal or metrical panel data).

Fahrmeir (1992) already proposed a method for the analysis of ordinal panel–data, the extended Kalman filter. His approach leads to posterior mode and curvature estimation. Here the general tool is Gibbs sampling, allowing a fully Bayesian analysis including posterior density estimation (see Carlin et. al., 1992) and model selection by Bayes factors (see Raftery, 1994).

Because of the recent popularity of Markov chain Monte Carlo methods neither the methodology of Gibbs sampling nor these applications will be outlined here. The reader is referred to the references cited above as well as to the papers of Smith & Roberts (1993) and Tierney (1994).

The posterior mean is used as a point estimator of unknown regression and threshold parameters. Hyperparameters (e.g. Σ_u) are treated as random variables with appropriate prior specification. So inference is in contrast to Fahrmeir (1992), where posterior mode is used and hyperparameters are deterministic. The proposed algorithm is easily implemented, because it needs only samples from inverted gamma and (possibly truncated) normal distributions. It should be noted that the presented model is a quite general tool, since it is also suitable for ordinal time series ($n_t \equiv 1$) and survival data ($q = 1$, see Fahrmeir, 1994).

In Section 2 the proposed model is outlined. One part is the response or observation model which connects observations with both regression and threshold parameters using latent variables. The other part is the transition model which specifies the variation of regression and threshold parameters in time. Section 3 outlines the implementation by Gibbs sampling specifying the necessary complete conditional densities. Finally Section 4 gives two illustrative examples of the broad applicability of this approach to applied statistical problems. A result, concerning the complete conditional distribution of β_t in an AR model is proofed in the appendix. This result is used several times in Section 3.

2 Model Specification

Observation model

For given covariate vector x_{ti} ordinal response $Y_{ti}, t = 1, \ldots, T, i = 1, \ldots, n_t$ is given in $k = q+1$ categories $1, \ldots, k$. Assume that Y_{ti} is caused by a latent continuous variable

$$Z_{ti} = -x'_{ti}\beta_t + \epsilon_{ti}, \quad \epsilon_{ti} \sim N(0,1)$$

due to the threshold mechanism

$$Y_{ti} = j \Longleftrightarrow \gamma_{t,j-1} < Z_{ti} \leq \gamma_{tj}, \quad j = 1, \ldots, k.$$

From these assumptions it follows that the response variable Y_{ti} is determined by the cumulative probit model

$$P(Y_{ti} \leq j | x_{ti}) = \Phi(\gamma_{tj} + x'_{ti}\beta_t) \quad j = 1, \ldots, q. \tag{2}$$

The thresholds γ_{t0} and γ_{tk} are set to $-\infty$ and ∞ respectively for all t, $(\gamma_{t1}, \ldots, \gamma_{tq})'$ is called the threshold parameter vector γ_t and β_t the regression parameter vector.

Transition model

We split up γ_t into a trend and a seasonal component: $\gamma_t = \gamma_t^T + \gamma_t^S$ and assume that the parameter vector β_t as well as both components γ_t^T and γ_t^S follow a specific AR model

$$\sum_{l=0}^{z_1} F_l \beta_{t-l} = u_t \qquad u_t \sim N(0, \Sigma_u)$$
$$\sum_{l=0}^{z_2} G_l \gamma_{t-l}^T = w_t^T \qquad w_t^T \sim N(0, \Sigma_w^T)$$

$$\sum_{l=0}^{z_3} H_l \gamma_{t-l}^S = w_t^S \qquad w_t^S \sim N(0, \Sigma_w^S),$$

for example a local linear trend model for β_t, a random walk of second order for γ_t^T and a flexible seasonal component for γ_t^S. Without loss of generality F_0, G_0 and H_0 are assumed to be identity matrices. We put (independent) priors on the unknown initial values $\beta_1, \ldots, \beta_{z_1}$, $\gamma_1^T, \ldots, \gamma_{z_2}^T$ and $\gamma_1^S, \ldots, \gamma_{z_3}^S$. We consider these priors as diffuse, by that we mean a normal prior with infinite variance as outlined in the appendix. All random variables in the model, which are not obviously dependent are assumed to be mutually independent (i. e. ϵ_t, u_t, w_t^T and w_t^S).

3 Implementation of the Gibbs sampler

A Gibbs sampling analysis (see for example Smith & Roberts, 1993) needs samples from the complete conditional densities of each unknown parameter, given all the others. For simplicity we first consider the case without a seasonal component γ_t^S, that is $\gamma_t = \gamma_t^T$. At the end we will sketch the general case.

Now we have the unknown parameters β_t, γ_t ($t = 1, \ldots, T$), the unknown latent variables Z_{ti} ($t = 1, \ldots, T$ $i = 1, \ldots, n_t$) and the unknown covariance matrices Σ_u and Σ_w, which also are treated as random variables. Let Z_t be the vector $(Z_{t1}, \ldots, Z_{tn_t})'$. Let Z^*, β^* and γ^* be abbreviations for $Z_1, \ldots, Z_T$, $\beta_1, \ldots, \beta_T$ and $\gamma_1, \ldots, \gamma_T$, respectively. Finally let Y^* denote the whole data. To implement the Gibbs sampler we require samples from the following complete conditional distributions:

$$\beta_t | \beta_{s \neq t}, \gamma^*, Z^*, \Sigma_u, \Sigma_w, Y^* \quad \sim \quad \beta_t | \beta_{s \neq t}, Z_t, \Sigma_u$$

$$
\begin{aligned}
\gamma_{tj}|\beta^*, \gamma_{tk\neq tj}, \gamma_{s\neq t}, Z^*, \Sigma_u, \Sigma_w, Y^* &\sim \gamma_{tj}|\gamma_{t,j-1}, \gamma_{t,j+1}, \gamma_{s\neq t}, Z_t, \Sigma_w, Y_t \\
Z_{ti}|\beta^*, \gamma^*, Z_{tk\neq ti}, Z_{s\neq t}, \Sigma_u, \Sigma_w, Y^* &\sim Z_{ti}|\beta_t, \gamma_t, Y_t \\
\Sigma_u|\beta^*, \gamma^*, Z^*, \Sigma_w, Y^* &\sim \Sigma_u|\beta^* \\
\Sigma_w|\beta^*, \gamma^*, Z^*, \Sigma_u, Y^* &\sim \Sigma_w|\gamma^*
\end{aligned}
$$

In the following we will derive all these complete conditional densities, shortly denoted by $p(\beta_t|\quad)$ and so on.

3.1 The Complete Conditional Densities

Using the results of the Appendix and Bayes's theorem, the complete conditional density of regression parameter β_t is given by

$$p(\beta_t|\quad) \propto \varphi(\beta_t, A_t a_t, A_t) p(Z_t|\beta_t),$$

where $\varphi(\beta_t, A_t a_t, A_t)$ denotes the multivariate normal density function with argument β_t, mean $A_t a_t$ and covariance matrix A_t. The specific values of A_t and a_t follow from the chosen AR model for β_t. $p(Z_t|\beta_t)$ can be written as $\prod_{i=1}^{n_t} p(Z_{ti}|\beta_t)$, due to the mutual conditional independence of Z_{ti}, $i = 1, \ldots, n_t$. Let X_t denote the matrix $(x'_{t1}, \ldots, x'_{tn_t})'$. Because of the normality of Z_{ti} it follows that

$$\beta_t| \quad \sim N(B_t b_t, B_t) \text{ with } B_t^{-1} = A_t^{-1} + X'_t X_t \text{ and } b_t = a_t - X'_t Z_t.$$

So we just have to sample from the multivariate normal distribution.

Again using the results of the Appendix, the complete conditional density of the threshold parameter γ_{tj} is given by

$$p(\gamma_{tj}|\quad) \propto \varphi(\gamma_{tj}, C_t c_t, C_t) * \mathbf{1}_{D_{tj}}(\gamma_{tj}),$$

here $\mathbf{1}_{D_{tj}}$ denotes the indicator function for the interval D_{tj}. The values of C_t and c_t follow from the chosen AR model for γ_t. So $\gamma_{tj}|\quad$ has a truncated normal distribution restricted to

$$D_{tj} = [\max\{\max_i\{Z_{ti} : Y_{ti} = j\}, \gamma_{t,j-1}\}, \min\{\min_i\{Z_{ti} : Y_{ti} = j+1\}, \gamma_{t,j+1}\}]$$

($\max\{\emptyset\} = -\infty$, $\min\{\emptyset\} = \infty$). This means, that the samples γ_{tj} have to follow the restriction for γ_t ($\gamma_{t,j-1} < \gamma_{tj} < \gamma_{t,j+1}$) and in addition the restriction caused by all these Z_{ti} with $Y_{ti} = j$ ($Z_{ti} <= \gamma_{tj}$) and by all these Z_{ti} with $Y_{ti} = j+1$ ($\gamma_{tj} < Z_{ti}$). Sampling from the truncated normal distribution can be done by simple inversion (see Devroye, 1986) or a more sophisticated method, see the work of Geweke(1991).

Due to the definition of the model, $Z_{ti}|\quad$ has a truncated normal distribution with mean $-x'_{ti}\beta_t$ and unit variance. It is restricted to the interval $(\gamma_{t,j-1}, \gamma_{tj}]$, where j is given by the value of the corresponding response $Y_{ti} = j$.

In most cases the covariance matrix Σ_u (or Σ_w) in an AR model is a diagonal matrix. So it is sufficient to consider the independent components. Let σ_i^2 and u_{ti} be the i–th component of diag(Σ_u) and u_t respectively. By use of Jeffrey's improper prior specification $p(\sigma_i^2) \propto \sigma_i^{-2}$ the complete conditional distribution of σ_i^2 is given by

$$\sigma_i^2|\quad \sim IG\left((T - z_1)/2, \left(\frac{1}{2}\sum_{t=1+z_1}^{T} u_{ti}^2\right)^{-1}\right),$$

where IG denotes the inverse gamma distribution. By use of an informative conjugate prior specification $\sigma_i^2 \sim IG(a_i, b_i)$ the complete conditional distribution of σ_i^2 is given by

$$\sigma_i^2| \quad \sim IG\left(a_i + (T - z_1)/2, \left(\frac{1}{b_i} + \frac{1}{2}\sum_{t=1+z_1}^{T} u_{ti}^2\right)^{-1}\right).$$

It has turned out in our simulation studies, that we get good results with Jeffrey's prior. Nevertheless, when $\sigma_i^2 \approx 0$, the Gibbs sampling algorithm runs into numerical problems, which means, that it is not converging. Besag & Green (1992) call this phenomenon a "computational black hole" occurring in single component updating schemes like the Gibbs sampler. To avoid this (and of course in case of prior information) an informative prior is a good alternative.

3.2 How to Include a Seasonal Component

Now both components of $\gamma_t = \gamma_t^T + \gamma_t^S$ follow an (independent) AR model and a generalization to that case is straightforward. The obvious change we have to make is to sample from the complete conditional distributions of γ_{tj}^T, γ_{tj}^S, Σ_w^T and Σ_w^S, instead of sampling from γ_{tj} and Σ_w. The last two distributions are covered by Section (3.1) and the first two are given by

$$\begin{aligned} p(\gamma_{tj}^T| \quad) &\propto \varphi(\gamma_{tj}^T, C_t c_t, C_t) * \mathbf{1}_{D_{tj}}(\gamma_{tj}^T) \\ \text{and } p(\gamma_{tj}^S| \quad) &\propto \varphi(\gamma_{tj}^S, E_t e_t, E_t) * \mathbf{1}_{F_{tj}}(\gamma_{tj}^S) \end{aligned}$$

where

$$D_{tj} = [\max\{\max_i\{Z_{ti} : Y_{ti} = j\}, \gamma_{t,j-1}\} - \gamma_{tj}^S, \min\{\min_i\{Z_{ti} : Y_{ti} = j+1\}, \gamma_{t,j+1}\} - \gamma_{tj}^S]$$

and

$$F_{tj} = [\max\{\max_i\{Z_{ti} : Y_{ti} = j\}, \gamma_{t,j-1}\} - \gamma_{tj}^T, \min\{\min_i\{Z_{ti} : Y_{ti} = j+1\}, \gamma_{t,j+1}\} - \gamma_{tj}^T].$$

3.3 Parameter Estimation

Since Gibbs sampling leads to samples from the joint posterior distribution, posterior moments can be estimated by Monte Carlo integration. For example the estimator of the posterior mean is simply the arithmetic mean of these samples. It is not necessary to use independent samples, so we can implement one long Gibbs run and average all samples after a "burn–in". Estimated posterior mean ± posterior standard deviation for $t = 1, \ldots, T$ will be called "naive confidence band".

4 Real Data Applications

4.1 Business Data

We apply our approach to monthly business microdata, collected by the IFO institute in Munich. It is based on a questionnaire, answered by 55 firms of a specific branch from January 1980 to December 1990. This data was previously analyzed by Fahrmeir (1992) and Fahrmeir & Nase (1994) with Kalman filter methods leading to posterior mode estimation. The response variable "production plan" is given in three categories, "decrease" (-), "no change" (=) and "increase" (+). Its conditional distribution is assumed to depend on the covariates "orders in hand", "expected development of the state of business for the next 6 month" as well as on the production plans of the previous month. These three covariates are all trichotomous, which leads to 3*2 dummy variables, denoted by A^+, $A^=$ (orders in hand), G^+, $G^=$ (expected state of business development) and PE^+, $PE^=$ (production plans of the previous month) with "decrease" as the reference category. For more details on this data see Fahrmeir & Nase (1994).

In our analysis we assume a random walk of first order for β_t and γ_t^T and a simple flexible seasonal component with period 12 for γ_t^S. To obtain smooth estimates for β_t and γ_t^T, we choose an informative prior with expectation and standard deviation equal to 0.0001 for the components of Σ_u and Σ_w^T. For the components of Σ_w^S we set expectation and standard deviation equal to 1. The resulting estimates of γ_t^S show a strong seasonal pattern (Figure 1), whereas the estimates of γ_t^T are nearly time constant (Figure 2). The estimates of the components of Σ_w^T are 0.0075 and 0.0039 for lower and upper threshold–trend respectively. The estimates of β_t are very similar to those obtained in Fahrmeir & Nase (1994): Only the variable G^+ for expected increase of "development of the state of business for the next 6 month" shows a significant temporal variation (Figure 3 and 4) with a low around 1982, when a new government was established in Germany.

FIGURE 1a: Seasonal Component of First Threshold Parameter

Estimates and Naive Confidence Bands

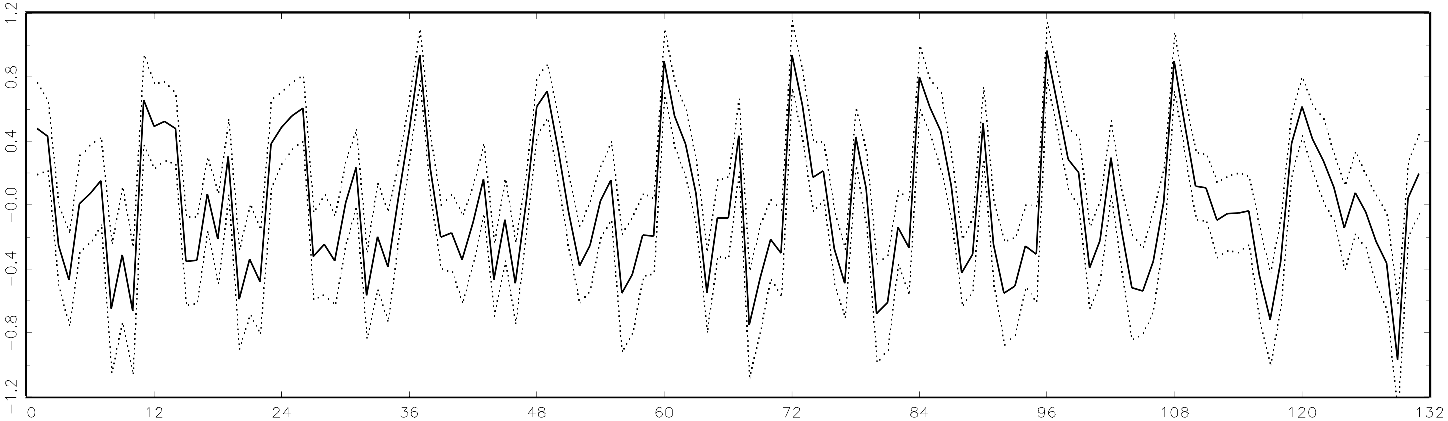


FIGURE 1b: Seasonal Component of Second Threshold Parameter

Estimates and Naive Confidence Bands

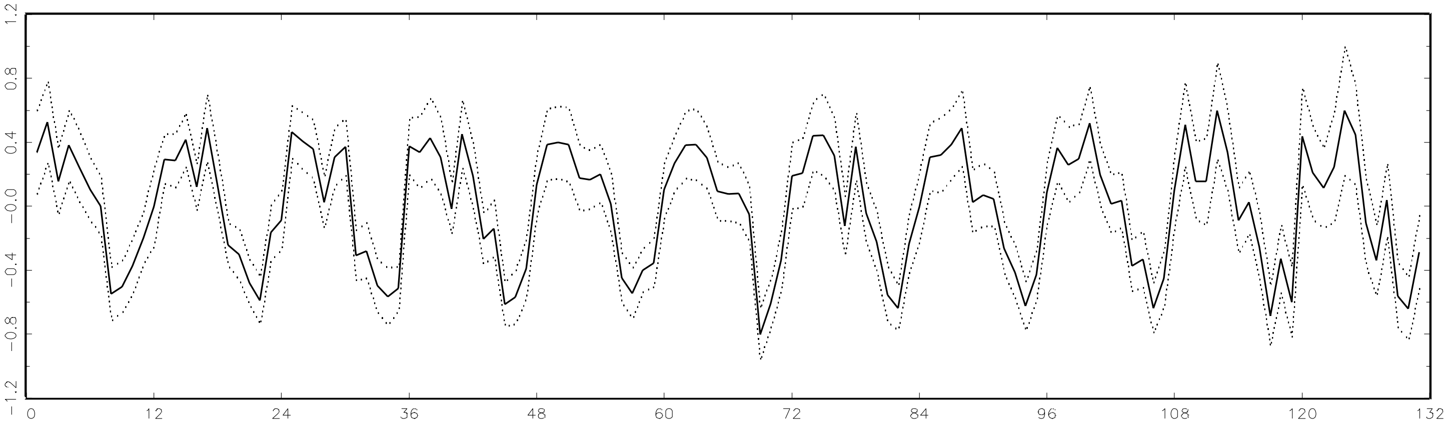


FIGURE 2a: Trend Components of Both Threshold Parameters

Estimates and Naive Confidence Bands

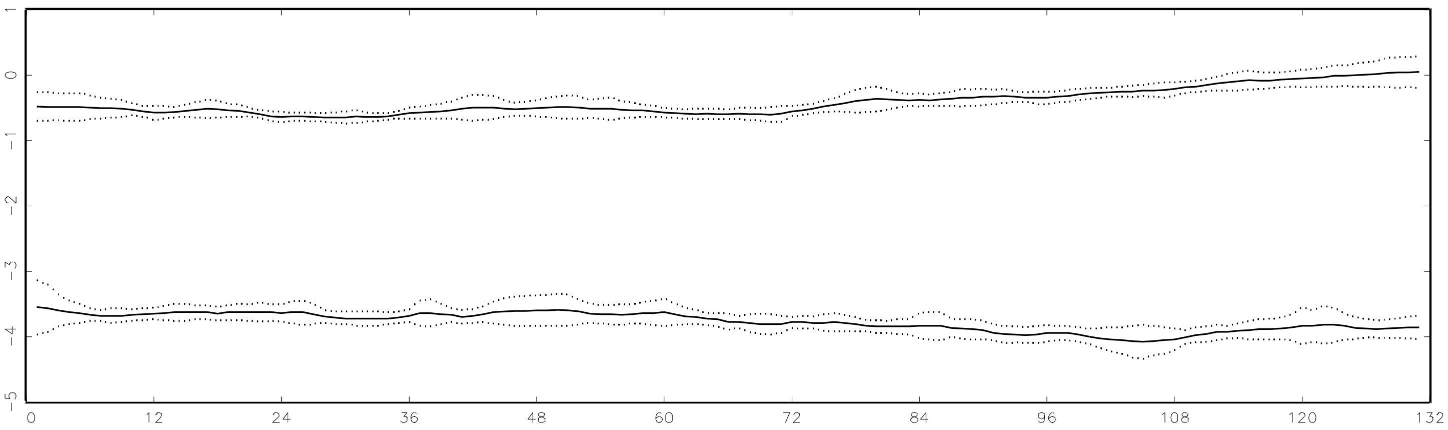


FIGURE 2b: Both Threshold Parameters (Trend plus Season)

Estimates and Naive Confidence Bands

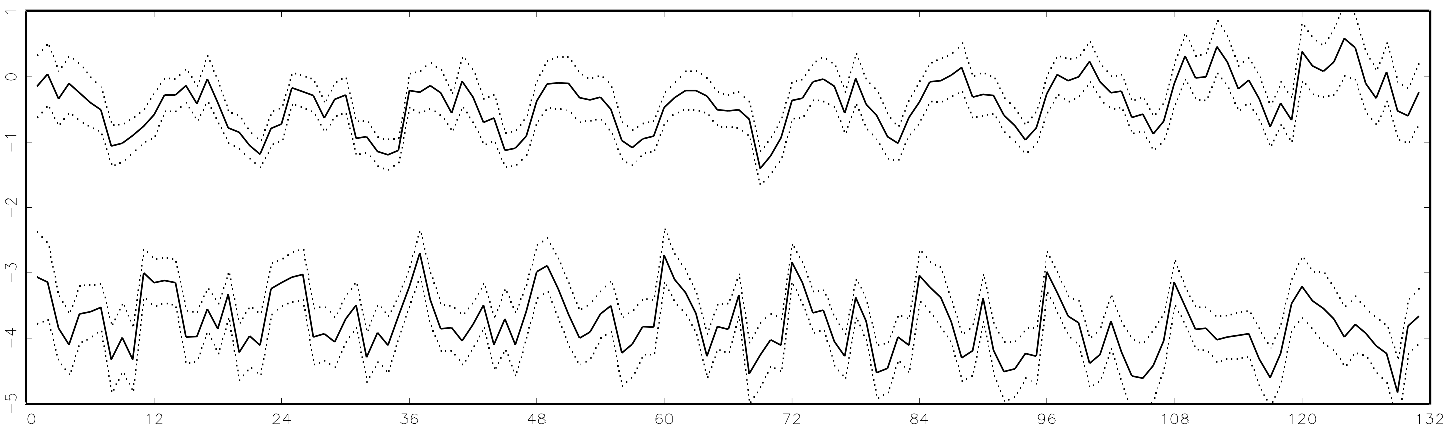


FIGURE 3: Parameter Estimates

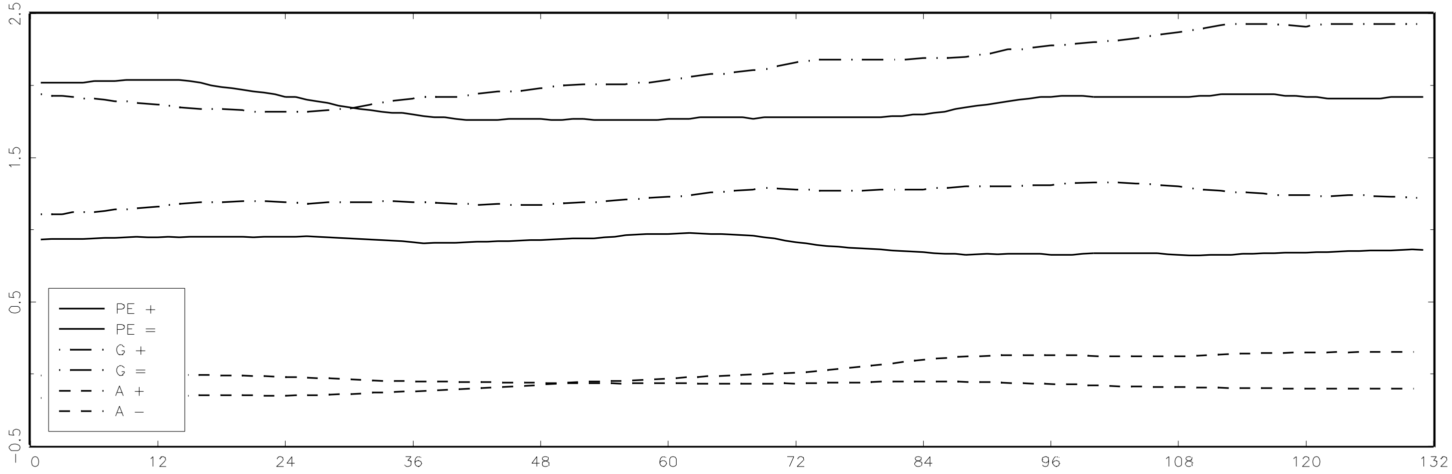
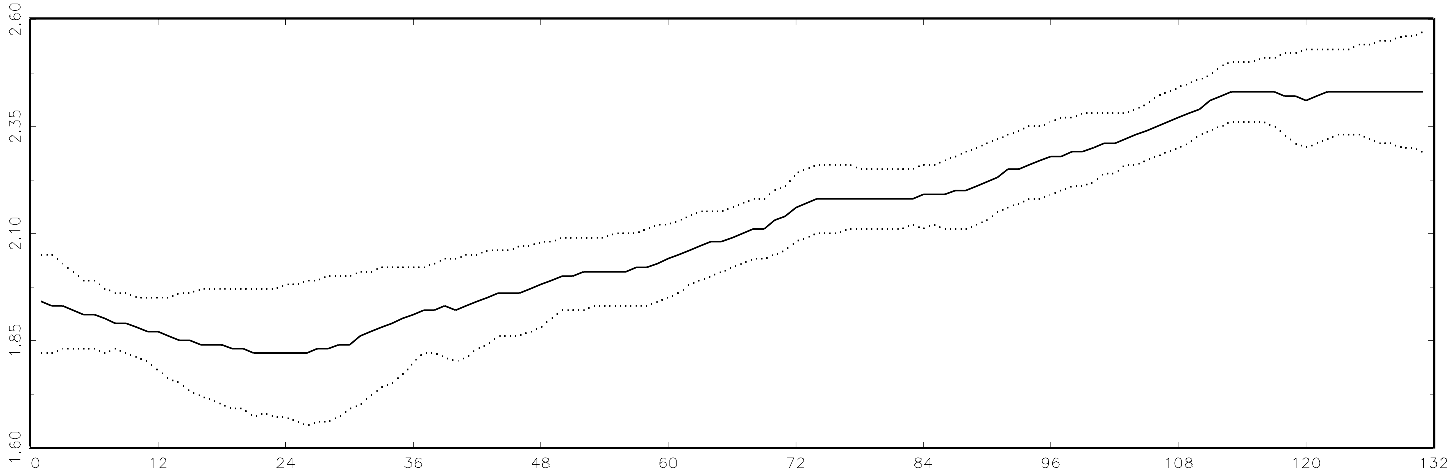


FIGURE 4: Parameter Estimate of G+ with Naive Confidence Bands



4.2 German Soccer Data

In the first division of the German soccer league 18 teams meet twice within each season giving each team the home court advantage once. We analyze data from the full season 1993/1994 with $T = 34$ days of games. The result Y_{ti} of one game is given in three ordered categories: win (1), draw (2) and loss (3). Similiar to the ideas of Fahrmeir & Tutz (1994) a dynamic ordered paired comparison system is assumed. The components of β_t represent the abilities of each team changing over time. Due to identifiability reasons one has to impose a further restriction. Here we set the ability $\beta_{t,18}$ of the last team (in alphabetic order) to zero. Note that the more natural restriction $\sum_i^{18} \beta_{ti} = 0$ can be achieved after estimation simply by subtracting the average ability at time t from $\hat{\beta}_t$. As usual in models for ordered paired comparison data, it is assumed that the difference in ability of two competing teams affects the probability of win, draw and loss. The (possibly time–varying) home court advantage is represented by the unknown threshold parameters γ_t.

For definition of the covariate vector consider one specific game Y_{ti}, when, let say, team 3 is playing at home against team 5. Then the covariate vector x_{ti} consists of zeros except the third and the fifth components, which are set to 1 and -1 respectively. The model equation (2) now specializes to

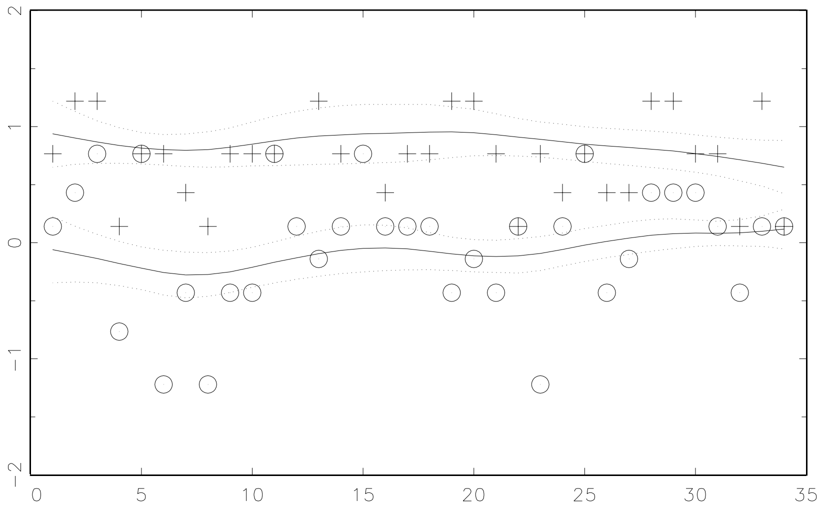
$$P(Y_{ti} \leq j) = \Phi(\gamma_{tj} + \beta_{t3} - \beta_{t5}) \quad j = 1, 2,$$

so the ability of team 3 minus the ability of team 5 at time t affects the (cumulative) probabilities of win, draw and loss.

In our analysis we use conjugate priors for the components of Σ_u with expectation and standard deviation equal to 0.1, for the components of Σ_w we

Figure 5: Estimated Threshold Parameters together with

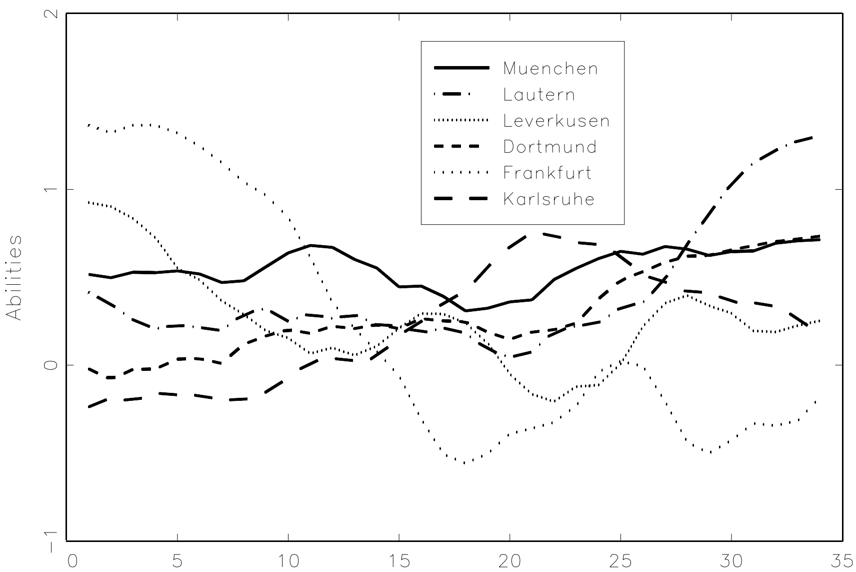
Transformed Cumulative Observed Frequencies



use Jeffrey's noninformative prior. We model the dynamic of the ability and the threshold parameters with a random walk of first and second order respectively. Two questions are of interest: Is there a significant "Beckenbauer effect" when later champion FC Bayern München dismissed coach Erich Ribbeck and hired Franz Beckenbauer at the middle of the season ($t = 21$)? Secondly we focus on the fit and will compare the observed and fitted final table.

Figure 5 shows the estimated threshold parameters together with observed cumulative frequencies $\hat{\pi}_{tj}$ transformed into the same scale by $\Phi^{-1}(\hat{\pi}_{tj})$. Note that for $t = 12$ and $t = 15$ no game was lost and therefore transformed observed cumulative frequency $\Phi^{-1}(\hat{\pi}_{t2})$ is equal to ∞ and is not plotted. In Figure 6 the abilities of the six best teams are shown. This picture is in close agreement with the leading position of Eintracht Frankfurt for the first ten playing days as well as the "final spurt" of Kaiserslautern at the end of the season. Finally Figure 7 shows the estimated ability within corresponding

Figure 6: Estimated Abilities of the Six Best Teams



naive confidence band of Bayern München. There is no significant but a noticeable increase after coach substitution ($t = 21$). Figure 8 shows observed and fitted points of the final table. One team gets two points for a win and one point for a draw, so the fitted points are calculated by 2∗estimated probability of win + 1∗estimated probability of draw. There is a close agreement between observed and fitted points so the model seems to be quite appropriate.

Acknowledgements

I like to thank Beate Elfinger and Elisabeth Pritscher for several discussions and suggestions. Special thanks go to Ludwig Fahrmeir for supervising this research project.

Figure 7: Ability of Bayern Muenchen

Estimate and Naive Confidence Band

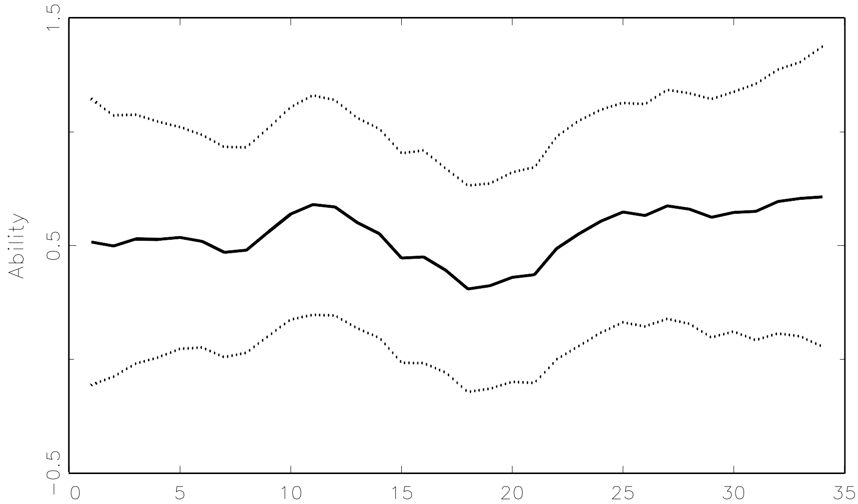
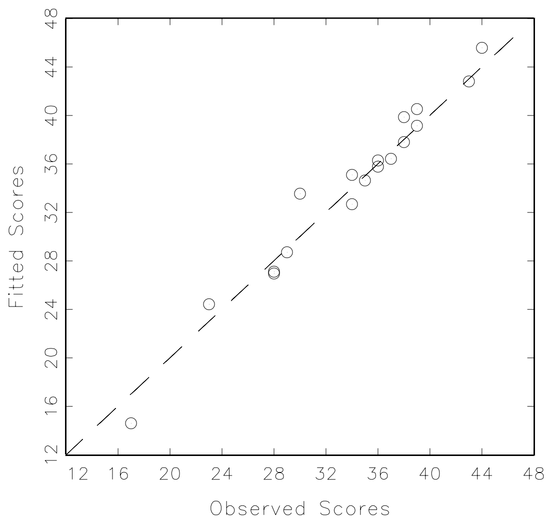


Figure 8:

Bundesliga 93/94: Observed versus Fitted



Appendix: The Conditional Distribution of β_t in an AR Model

Assume that parameter vector β_t follows an AR model

$$\sum_{i=0}^{z} F_i \beta_{t-i} = u_t \quad t = z+1, \ldots, T$$

with $F_0 = I$ and independent starting values $\beta_1, \ldots, \beta_z$, normally distributed with mean μ_t and covariance matrix Σ_t, $t = 1, \ldots, z$. Let the error term u_t be independent normally distributed with mean zero and covariance matrix Σ_u and of course independent of the starting values.

Then the conditional distribution of $\beta_t | \beta_{s \neq t}, \Sigma_u$ is a multivariate normal distribution $N(A_t a_t, A_t)$, where the terms A_t and a_t are defined by

$$A_t^{-1} = \begin{cases} \Sigma_t^{-1} + \sum\limits_{j=1+z-t}^{\min(z,T-t)} F_j' \Sigma_u^{-1} F_j & t = 1, \ldots, z \\ \sum\limits_{j=0}^{\min(z,T-t)} F_j' \Sigma_u^{-1} F_j & t = z+1, \ldots, T \end{cases}$$

and

$$a_t = \begin{cases} -\Sigma_t^{-1} \mu_t - \sum\limits_{j=1+z-t}^{\min(z,T-t)} F_j' \Sigma_u^{-1} \left(\sum\limits_{i=0, i \neq j}^{z} F_i \beta_{t+j-i} \right) & t = 1, \ldots, z \\ - \sum\limits_{j=0}^{\min(z,T-t)} F_j' \Sigma_u^{-1} \left(\sum\limits_{i=0, i \neq j}^{z} F_i \beta_{t+j-i} \right) & t = z+1, \ldots, T \end{cases}.$$

Proof:

By Bayes's theorem, the density $p(\beta_t | \beta_{s \neq t}, \Sigma_u)$ is proportional to

$$\begin{cases} p(\beta_t) \prod\limits_{j=1+z-t}^{\min(z,T-t)} p(\beta_{t+j} | \beta_{t+j-1}, \ldots, \beta_{t+j-z}, \Sigma_u) & t = 1, \ldots, z \\ \prod\limits_{j=0}^{\min(z,T-t)} p(\beta_{t+j} | \beta_{t+j-1}, \ldots, \beta_{t+j-z}, \Sigma_u) & t = z+1, \ldots, T. \end{cases}$$

For simplicity let us just consider the case $z < t \leq T - z$; the formulas for the endpoint cases can be derived in a similar manner. Then we have

$$p(\beta_t|\beta_{s\neq t}, \Sigma_u) \propto \prod_{j=0}^{z} p(\beta_{t+j}|\beta_{t+j-1}, \ldots, \beta_{t+j-z}, \Sigma_u),$$

where $p(\beta_{t+j}|\beta_{t+j-1}, \ldots, \beta_{t+j-z}, \Sigma_u)$ is the density function of a normal distribution with mean $-\sum_{i=1}^{z} F_i\beta_{t+j-i}$ and covariance matrix Σ_u. Therefore $-2\log p(\beta_t|\beta_{s\neq t}, \Sigma_u)$ is, up to a normalizing constant, a sum of $z+1$ terms:

$$\begin{aligned}
& \sum_{j=0}^{z}\left[\left(\beta_{t+j} + \sum_{i=1}^{z} F_i\beta_{t+j-i}\right)' \Sigma_u^{-1} \left(\beta_{t+j} + \sum_{i=1}^{z} F_i\beta_{t+j-i}\right)\right] \\
= & \sum_{j=0}^{z}\left[\left(\sum_{i=0}^{z} F_i\beta_{t+j-i}\right)' \Sigma_u^{-1} \left(\sum_{i=0}^{z} F_i\beta_{t+j-i}\right)\right] \\
= & \sum_{j=0}^{z}\left[\left(\beta_t + F_j^{-1}\left(\sum_{i=0,i\neq j}^{z} F_i\beta_{t+j-i}\right)\right)' F_j'\Sigma_u^{-1}F_j \left(\beta_t + F_j^{-1}\left(\sum_{i=0,i\neq j}^{z} F_i\beta_{t+j-i}\right)\right)\right].
\end{aligned}$$

Combining these quadratic forms (see for example Box & Tiao, p. 418) leads to the desired result.

□

Using noninformative prior distributions for the starting values ($\mu_t = 0$ and $\Sigma_t^{-1} = 0$, $t = 1, \ldots, z$), these formulas simplify to

$$A_t^{-1} = \sum_{j=\max(0,1+z-t)}^{\min(z,T-t)} F_j'\Sigma_u^{-1}F_j \quad t = 1, \ldots, T$$

and

$$a_t = -\sum_{j=\max(0,1+z-t)}^{\min(z,T-t)} F_j'\Sigma_u^{-1}\left(\sum_{i=0,i\neq j}^{z} F_i\beta_{t+j-i}\right) \quad t = 1, \ldots, T.$$

The use of noninformative priors for the starting values is a natural choice, often leading to symmetric results for the beginning and the end of the series

$p(\beta_t|\beta_{s\neq t}, \Sigma_u)$. For example, consider a random walk of second order of a scalar β_t:

$$\beta_t = 2\beta_{t-1} - \beta_{t-2} + u_t, \quad u_t \sim N(0, \sigma_u^2).$$

Then the conditional distribution $p(\beta_t|\beta_{s\neq t}, \sigma_u^2)$ is normal with mean

$$\begin{array}{ll} 2\beta_{t+1} - \beta_{t+2} & (t = 1) \\ (2\beta_{t-1} + 4\beta_{t+1} - \beta_{t+2})/5 & (t = 2) \\ (-\beta_{t-2} + 4\beta_{t-1} + 4\beta_{t+1} - \beta_{t+2})/6 & (t = 3, \ldots, T-2) \\ (-\beta_{t-2} + 4\beta_{t-1} + 2\beta_{t+1})/5 & (t = T-1) \\ -\beta_{t-2} + 2\beta_{t-1} & (t = T) \end{array}$$

and variance

$$\begin{array}{ll} \sigma_u^2 & (t = 1) \\ \sigma_u^2/5 & (t = 2) \\ \sigma_u^2/6 & (t = 3, \ldots, T-2) \\ \sigma_u^2/5 & (t = T-1) \\ \sigma_u^2 & (t = T). \end{array}$$

References

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

Ansley, C.F. & Kohn, R. (1990) Filtering and smoothing in state space models with partially diffuse initial conditions. *Journal of Time Series Analysis*, 11, 277-93.

Besag, J. & Green, P.J. (1993) Spatial Statistics and Bayesian Computing. *Journal of the Royal Statistical Society B* 55 pp. 25-37.

Box, G. & Tiao, G. (1973) *Bayesian Inference in Statistical Analysis*: Addison–Wesley Publishing Company.

Carlin, B.P. & Polsen, N.G. (1993) Monte Carlo Bayesian Methods for Discrete Regression Models and Categorical Time Series. In *Bayesian Statistics 4*, J. Bernado et al., Eds., Oxford University Press, Oxford, pp. 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.

Devroye, L. (1986) *Non–uniform Random Variate Generation.* New York: Springer.

Fahrmeir, L. (1992) Posterior Mode Estimation by Extended Kalman Filtering for Multivariate Dynamic Generalized Linear Models. *Journal of the American Statistical Association*, 87, 501-509.

Fahrmeir, L. (1994) Dynamic Modelling and Penalized Likelihood Estimation for Discrete Time Survival Data. *Biometrika*, 81, 317-330.

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

Fahrmeir, L. & Tutz, G. (1994) Dynamic Stochastic Models for Time–dependent Ordered Paired Comparison Systems. *Journal of the American Statistical Association*, 89, 1438-1449.

Geweke, J. (1991) Efficient Simulation From the Multivariate Normal and Student–T Distributions Subject to Linear Constraints. in *Computing Science and Statistics: Proceedings of the Twenty–Third Symposium* *on the Interface*, Alexandria, VA: American Statistical Association, pp. 571–578.

Raftery, A.E. (1995) Hypothesis testing and model selection with posterior simulation. In *Practical Markov Chain Monte Carlo* (W.R Gilks et al. eds.), London: Chapman and Hall, to appear.

Smith, A. & Roberts, G. (1993) Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society B* 55 pp. 3-23.

Tierney, L. (1994) Markov chains for exploring posterior distributions. *Annals of Statistics*, to appear.