LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386

Biller:

Posterior mode estimation in dynamic generalized linear mixed models. (REVISED, June 2000)

Sonderforschungsbereich 386, Paper 70 (1997)

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

Projektpartner

gsf

MAX-PLANCK-GESELLSCHAFT

TUM

Posterior mode estimation in dynamic generalized linear mixed models

Clemens Biller *

13th June 2000

Summary: Dynamic generalized linear mixed models for longitudinal data combine the generalized linear mixed model and the dynamic generalized linear model dealing with the case that both unit– and time–specific parameters are allowed. We base statistical inference on posterior mode estimation thus avoiding numerical integrations in high dimensions or Monte Carlo simulations which are necessary for posterior mean estimation in a fully Bayesian analysis. This results in a Fisher scoring algorithm with backfitting steps in each scoring iteration, since estimating equations of the unobserved effects mutually contain each other effect. Algorithms for estimation of random effects and dynamic effects can be used in each backfitting step due to the additive definition of the model. Estimation of unknown hyperparameters is done by an EM–type algorithm where posterior modes and curvatures resulting from the Fisher scoring algorithm are substituted for posterior means and covariances. We apply the model to multicategorical business test data.

Keywords: backfitting algorithm; dynamic effects; EM algorithm; Fisher scoring algorithm; posterior mode estimation; random effects.

1 Introduction

Longitudinal data consist of observations (y_{it}, x_{it}) for a population of $i = 1, \ldots, n$ units observed across $t = 1, \ldots, T$ time periods, where y_{it} denotes the response variate and x_{it} a vector of covariates. To deal with nonnormal responses, generalized linear models (Nelder and Wedderburn, 1972, or McCullagh and Nelder, 1989) are a well known regression tool. Here it is assumed that the conditional mean $\mu_{it} = \mathrm{E}(y_{it}|x_{it})$ of the response given the covariates is related to the linear predictor $\eta_{it} = X_{it}'\alpha$ by a response function h, i.e., $\mu_{it} = h(\eta_{it})$, where α denotes the vector of unknown regression parameters independent of time t and unit i. The design matrix X_{it} is formed out of the covariates x_{it}.

To consider unobserved heterogeneity among units, that often can not be measured completely by the observed covariates, the generalized linear mixed model (GLMM or random effects model, see Breslow and Clayton, 1993) extends the generalized linear model by introducing random effects β_i varying independently over units.

*This work was supported by a grant from the German National Science Foundation, Sonderforschungsbereich 386.

Another extension of the generalized linear model is the dynamic generalized linear model (DGLM, see Fahrmeir, 1992) dealing with the assumption, that covariate effects may vary over time, typically following a Markovian transition model.

If we want to model time–varying effects but also take into account unobserved heterogeneity among the units, a combination of the GLMM and the DGLM is attractive. For normal data such models already exist, see e.g. Hsiao (1986) or Rosenberg (1973), whereas in the generalized setting of exponential families little research has been done up to now. Fahrmeir, Kaufmann and Morawitz (1989) used an information matrix filter for a model with parameters varying across time and units. Knorr–Held (1995) considered the "dynamic generalized linear mixed model" (DGLMM) by adding the random effect term of the GLMM to the linear predictor of the DGLM. Estimation of this model was done by Markov Chain Monte Carlo methods. A somewhat related approach, the "generalized additive mixed model", was proposed by Lin and Zhang (1999). Here covariate effects were modelled by additive nonparametric functions, while additive random effects were used to account for heterogeneity. To avoid numerical integrations double penalized quasi–likelihood was considered to make approximate inference.

In this paper we consider an extended version of the DGLMM by additionally adding the fixed covariate effect α to the linear predictor of Knorr–Held (1995), thus getting

$$\eta_{it} = X_{it}'\alpha + U_{it}'\beta_i + Z_{it}'\gamma_t, \tag{1}$$

with random effects β_i and time–varying dynamic effects γ_t. A direct approach based on the posterior density of the unknown parameters would involve computationally intractable high–dimensional integrations. Bayesian techniques like the Markov Chain Monte Carlo methods used in Knorr–Held (1995) avoid numerical integrations by taking repeated samples from the posterior distribution. However, the choice of computationally efficient sampling schemes and questions about convergence of the sampling process to equilibrium are drawbacks of these methods. Therefore we use posterior mode estimation, a conceptually simpler approach, that may be used to check the convergence behaviour of the simulation–based Markov Chain Monte Carlo methods. To avoid numerical integrations here we maximize the posterior density of the unknown parameters. This results in Fisher scoring steps, where the estimating equations of each of the unknown parameters mutually contain the other unknown parameters. We solve this problem by using the iterative backfitting or Gauss–Seidel algorithm at each Fisher scoring step. In the derivation of the algorithm, the results of posterior mode estimation of the GLMM and the DGLM can be applied to the DGLMM. An overview of these models is given in Fahrmeir and Tutz (1997).

After introducing the DGLMM in Section 2, the algorithm for estimating the unknown effects is derived in Section 3. The estimating equation of the dynamic effect γ_t in the backfitting algorithm contains the inverse of a block tridiagonal matrix. In the DGLM the generalized extended Kalman filter and smoother presented by Fahrmeir (1992) is used for approximative posterior mode estimation of the dynamic effects, whereby the computation of the mentioned inverse matrix is avoided. But in combination with the backfitting algorithm in the DGLMM this method showed numerical problems leading to

singular covariance matrices. Fahrmeir and Kaufmann (1991) considered a Fisher scoring algorithm with better approximation qualities to the posterior mode based on the so–called LDL' factorization of the block tridiagonal matrix. Section 4 describes a modified version of this algorithm, that shows no numerical problems in connection with backfitting. For the above mentioned algorithms we assume that all hyperparameters of the model, i.e., the parameters of the prior specifications of the varying effects, are known. But in practice these hyperparameters are only known in special cases, so techniques are required to estimate the model parameters and hyperparameters simultaneously. Section 5 illustrates such a technique, a modification of the iterative EM algorithm of Dempster, Laird and Rubin (1977). In Section 6 the DGLMM is applied to multicategorical business test data. A brief summary is given in Section 7.

2 The dynamic generalized linear mixed model

Consider longitudinal data as defined in Section 1. Response y_{it} may be univariate in the case of count data or binomial data, or a vector $y_{it} = (y_{it1}, \ldots, y_{itl})'$ in the case of categorical data with $k = l + 1$ (unordered or ordered) categories. Here we consider the more general case of the l–dimensional response vector.

For notational convenience, observations at time t are collected in panel waves $y_t = (y'_{1t}, \ldots, y'_{nt})'$, $x_t = (x'_{1t}, \ldots, x'_{nt})'$, and histories of responses and covariates up to time t are denoted by $y^*_t = (y'_1, \ldots, y'_t)'$, $x^*_t = (x'_1, \ldots, x'_t)'$.

The DGLMM is defined by an observation model and a parameter model. The observation model assumes that the distribution of y_{it} belongs to a simple exponential family with density $p(y_{it}|y^*_{t-1}, x^*_t, \alpha, \beta_i, \gamma_t)$ and mean $\mu_{it} = \mathrm{E}(y_{it}|y^*_{t-1}, x^*_t, \alpha, \beta_i, \gamma_t) = h(\eta_{it})$, defined in dependence of the covariates with response function h and linear predictor (1). The design matrices X_{it}, U_{it} and Z_{it} are built from the covariates x_{it} and have dimensions $(s \times l)$, $(p \times l)$ and $(q \times l)$, respectively. Z_{it} may contain past observations (y^*_{t-1}, x^*_{t-1}). As pointed out in Section 1, we include the following unknown parameters: the s–dimensional fixed effect α, the p–dimensional random effect β_i, varying independently over the units i, and the q–dimensional dynamic parameter γ_t varying over time t.

The parameter model specifies prior information for the unknown parameters. As usual in the GLMM (see, e.g., Stiratelli, Laird and Ware, 1984, or Breslow and Clayton, 1993), the unit–specific parameters β_i are supposed to be independent and identical normally distributed with mean zero and covariance matrix H, i.e., $\beta_i \sim N(0, H)$ for $i = 1, \ldots, n$. Since the estimating procedure in Section 3 is based on the joint posterior density of the unknown parameters, a flat prior density with covariance matrix $\Gamma \to \infty$ is assigned to the parameter α. The composed parameter vector $b = (\alpha', \beta')'$ with $\beta = (\beta'_1, \ldots, \beta'_n)'$ therefore has the limiting prior density $p(b; H, \Gamma) \propto p(\beta; H)$ (as $\Gamma^{-1} \to 0$).

As in Fahrmeir (1992), with $\gamma_0 = a_0 + v_0$ the sequence of dynamic effects γ_t is defined by the linear Markovian transition equations

$$\gamma_t = T_t \gamma_{t-1} + v_t, \quad t = 1, 2, \ldots . \tag{2}$$

The initial value a_0 and transition matrices $T_1, T_2, \ldots$ are assumed to be known, the error terms v_t are supposed to be normally distributed with mean zero and covariance matrices Q_t. Furthermore, the v_t are independent of past error terms $v_{t-1}, \ldots, v_0$, of past response values y^*_{t-1} and of covariates x^*_t. Histories of dynamic effects up to time t are again denoted by $\gamma^*_t = (\gamma'_0, \gamma'_1, \ldots, \gamma'_t)'$, $t = 0, 1, \ldots, T$, and we define $\gamma = \gamma^*_T$. Independence is assumed between time–varying parameters γ and the composed parameter b.

For a complete model specification, the following additional independence assumptions are required: conditional on γ_t, b and y^*_{t-1}, both the current observation y_t is independent of γ^*_{t-1}, and the individual responses y_{it} within y_t are independent; conditional on y^*_{t-1} and x^*_{t-1}, the covariates x_t are independent of γ^*_{t-1} and b.

As a combination of the GLMM and the DGLM, model (1) comprises two further submodels, the first one consisting only of α and γ_t, the other of β_i and γ_t. The estimating procedures of these submodels result in a straightforward way from the procedure for model (1) in Sections 3 and 5 and therefore are omitted.

3 Posterior mode estimation

In this section we propose an algorithm for estimating the unknown parameters $\varphi = (\alpha', \beta', \gamma')'$ of the DGLMM. In the derivation of the algorithm we refer to results of posterior mode estimation in the GLMM and the DGLM, see Chapters 7 and 8 in Fahrmeir and Tutz (1997). All hyperparameters of the model, i.e., the covariance matrix H of β_i, the initial values a_0, Q_0 and the covariance matrices Q_t of transition equations (2), are assumed to be known. This assumption is dropped in Section 5, where we introduce an algorithm for the simultaneous estimation of parameters and hyperparameters.

Estimation of the parameters φ is based on the posterior density $p(\varphi|y^*_T, x^*_T)$ given the observations (y^*_T, x^*_T). A fully Bayesian analysis would require numerical integration in high dimensions. To avoid this, we estimate parameters by maximizing this density. Repeated application of Bayes' law and using the independence assumptions of Section 2 yields

$$p(\varphi|y^*_T, x^*_T) \propto \prod_{t=1}^{T} \prod_{i=1}^{n} p(y_{it}|y^*_{t-1}, x^*_t, \gamma_t, b) \prod_{t=1}^{T} p(\gamma_t|\gamma_{t-1}) p(\gamma_0) p(\beta).$$

Taking logarithms and neglecting constant terms, we obtain the penalized log–likelihood

$$PL(\varphi) = l(\varphi) + a(\varphi) \tag{3}$$

where $l(\varphi) = \sum_{i=1}^{n} \sum_{t=1}^{T} l_{it}(\varphi) + l_0(\varphi)$ is the sum of the individual log–likelihoods $l_{it}(\varphi) = \ln p(y_{it}|y^*_{t-1}, x^*_t, \gamma_t, b)$ and the log–prior $l_0(\varphi) = -\frac{1}{2}(\gamma_0 - a_0)' Q_0^{-1} (\gamma_0 - a_0)$ of γ_0. The penalty term $a(\varphi) = -\frac{1}{2}\gamma' P \gamma - \frac{1}{2}\beta' L \beta$ includes prior information on the unknown varying effects. The block diagonal matrix L is given by $L = \mathrm{diag}(H^{-1}, \ldots, H^{-1})$, while the penalty matrix

$$P = \begin{pmatrix} P_{00} & P_{01} & & 0 \\ P'_{01} & P_{11} & \ddots & \\ & \ddots & \ddots & P_{T-1,T} \\ 0 & & P'_{T-1,T} & P_{TT} \end{pmatrix} \tag{4}$$

is block tridiagonal with blocks $P_{00} = T_1' Q_1^{-1} T_1$, $P_{t-1,t} = -T_t' Q_t^{-1}$ for $t = 1, \ldots, T$, $P_{tt} = Q_t^{-1} + T_{t+1}' Q_{t+1}^{-1} T_{t+1}$ for $t = 1, \ldots, T-1$ and $P_{TT} = Q_T^{-1}$ (see Fahrmeir and Tutz, 1997, Section 8.1.2).

To derive estimates of φ, we need the score function $s(\varphi) = (s(\alpha)', s(\beta)', s(\gamma)')' = \partial PL(\varphi)/\partial\varphi$ and the expected information matrix $F(\varphi) = \mathrm{E}(-\partial^2 PL(\varphi)/\partial\varphi\partial\varphi')$.

As for the random effects model in Fahrmeir and Tutz (1997), Section 7.3.3, the first two components of the score function (with $s(\beta) = \partial PL(\varphi)/\partial\beta = (s(\beta_1)', \ldots, s(\beta_n)')'$) result in

$$\begin{aligned} s(\alpha) &= \partial PL(\varphi)/\partial\alpha = \sum_{i=1}^{n}\sum_{t=1}^{T} X_{it} D_{it} \Sigma_{it}^{-1}(y_{it} - \mu_{it}) \\ s(\beta_i) &= \partial PL(\varphi)/\partial\beta_i = \sum_{t=1}^{T} U_{it} D_{it} \Sigma_{it}^{-1}(y_{it} - \mu_{it}) - H^{-1}\beta_i, \quad i = 1, \ldots, n, \end{aligned}$$

with $D_{it} = \partial h(\eta_{it})/\partial\eta_{it}$, $\Sigma_{it} = \mathrm{cov}(y_{it}|\varphi)$ and $\mu_{it} = h(\eta_{it})$. Defining matrices $X_i' = (X_{i1}, \ldots, X_{iT})$, $X' = (X_1', \ldots, X_n')$, $U_i' = (U_{i1}, \ldots, U_{iT})$, $U = \mathrm{diag}(U_1, \ldots, U_n)$, $y_i = (y_{i1}', \ldots, y_{iT}')'$, $y = (y_1', \ldots, y_n')'$, $\mu_i = (\mu_{i1}', \ldots, \mu_{iT}')'$, $\mu = (\mu_1', \ldots, \mu_n')'$, $D_i = \mathrm{diag}(D_{i1}, \ldots, D_{iT})$, $D = \mathrm{diag}(D_1, \ldots, D_n)$, $\Sigma_i = \mathrm{diag}(\Sigma_{i1}, \ldots, \Sigma_{iT})$, $\Sigma = \mathrm{diag}(\Sigma_1, \ldots, \Sigma_n)$ we obtain the functions

$$\begin{aligned} s(\alpha) &= X' D \Sigma^{-1}(y - \mu) \\ s(\beta) &= U' D \Sigma^{-1}(y - \mu) - L\beta. \end{aligned}$$

The elements of $s(\gamma) = \partial PL(\varphi)/\partial\gamma = \partial l(\varphi)/\partial\gamma + \partial a(\varphi)/\partial\gamma$ are given by

$$\begin{aligned} \partial l(\varphi)/\partial\gamma_0 &= Q_0^{-1}(a_0 - \gamma_0) \\ \partial l(\varphi)/\partial\gamma_t &= \sum_{i=1}^{n} Z_{it} D_{it} \Sigma_{it}^{-1}(y_{it} - \mu_{it}), \quad t = 1, \ldots, T \\ \partial a(\varphi)/\partial\gamma &= -P\,\gamma. \end{aligned}$$

If we introduce matrices $Z_i' = \mathrm{diag}(Z_{i1}, \ldots, Z_{iT})$, $Z' = (Z_1', \ldots, Z_n')$, $\mathbf{Z} = \mathrm{diag}(I, Z)$, $\mathbf{D} = \mathrm{diag}(I, D)$, $\mathbf{\Sigma} = \mathrm{diag}(Q_0, \Sigma)$, $\mathbf{y} = (a_0', y')'$, $\boldsymbol{\mu} = (\gamma_0', \mu')'$, the third component of the score function has the form

$$s(\gamma) = \begin{pmatrix} Q_0^{-1}(a_0 - \gamma_0) \\ Z' D \Sigma^{-1}(y - \mu) \end{pmatrix} - P\,\gamma = \mathbf{Z}'\mathbf{D}\mathbf{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}) - P\,\gamma.$$

For a unified representation we additionally define matrices $\mathbf{X}' = (O', X')$, $\mathbf{U}' = (O', U')$ and $K = \mathrm{diag}(O, L, P)$ (with O in each case a matrix of zeros of appropriate dimension) yielding the score function

$$s(\varphi) = \begin{pmatrix} s(b) \\ s(\gamma) \end{pmatrix} = (\mathbf{X}, \mathbf{U}, \mathbf{Z})'\mathbf{D}\mathbf{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}) - K\varphi, \tag{5}$$

with $s(b) = (s(\alpha)', s(\beta)')'$.

The expected information matrix is

$$F(\varphi) = \begin{pmatrix} F(b) & F(b,\gamma) \\ F(\gamma,b) & F(\gamma) \end{pmatrix} = (\mathbf{X},\mathbf{U},\mathbf{Z})'\,\mathbf{W}\,(\mathbf{X},\mathbf{U},\mathbf{Z}) + K \tag{6}$$

with weight matrix $\mathbf{W} = \mathbf{D}\mathbf{\Sigma}^{-1}\mathbf{D}'$, components

$$F(b) = \begin{pmatrix} F_{\alpha\alpha} & F_{\alpha\beta} \\ F_{\beta\alpha} & F_{\beta\beta} \end{pmatrix} \qquad F(b,\gamma) = \begin{pmatrix} F_{\alpha\gamma} \\ F_{\beta\gamma} \end{pmatrix} = F(\gamma,b)' \qquad F(\gamma) = F_{\gamma\gamma}$$

and $F_{\alpha\alpha} = \mathbf{X}'\mathbf{W}\mathbf{X}$, $F_{\alpha\beta} = \mathbf{X}'\mathbf{W}\mathbf{U} = F'_{\beta\alpha}$, $F_{\alpha\gamma} = \mathbf{X}'\mathbf{W}\mathbf{Z} = F'_{\gamma\alpha}$, $F_{\beta\beta} = \mathbf{U}'\mathbf{W}\mathbf{U} + L$, $F_{\beta\gamma} = \mathbf{U}'\mathbf{W}\mathbf{Z} = F'_{\gamma\beta}$, $F_{\gamma\gamma} = \mathbf{Z}'\mathbf{W}\mathbf{Z} + P$.

The estimate $\hat{\varphi}$ of φ is computed as the solution of the likelihood equation $s(\hat{\varphi}) = 0$, which is in general nonlinear and has to be solved iteratively. We use Fisher scoring

$$\hat{\varphi}^{(k+1)} = \hat{\varphi}^{(k)} + F^{-1}(\hat{\varphi}^{(k)})s(\hat{\varphi}^{(k)}), \quad k = 0,1,2,\ldots,$$

starting with an initial estimate $\hat{\varphi}^{(0)}$. Direct computation of the Fisher scoring steps will lead to problems if dimensions are to high. Using the partitioning of the score function (5) and the information matrix (6), therefore these steps are transformed to

$$\begin{pmatrix} F(\hat{b}^{(k)}) & F(\hat{b}^{(k)},\hat{\gamma}^{(k)}) \\ F(\hat{\gamma}^{(k)},\hat{b}^{(k)}) & F(\hat{\gamma}^{(k)}) \end{pmatrix} \begin{pmatrix} \hat{b}^{(k+1)} - \hat{b}^{(k)} \\ \hat{\gamma}^{(k+1)} - \hat{\gamma}^{(k)} \end{pmatrix} = \begin{pmatrix} s(\hat{b}^{(k)}) \\ s(\hat{\gamma}^{(k)}) \end{pmatrix},$$

which results in the two equations

$$F(\hat{b}^{(k)})(\hat{b}^{(k+1)} - \hat{b}^{(k)}) + F(\hat{b}^{(k)},\hat{\gamma}^{(k)})(\hat{\gamma}^{(k+1)} - \hat{\gamma}^{(k)}) = s(\hat{b}^{(k)}) \tag{7}$$

$$F(\hat{\gamma}^{(k)},\hat{b}^{(k)})(\hat{b}^{(k+1)} - \hat{b}^{(k)}) + F(\hat{\gamma}^{(k)})(\hat{\gamma}^{(k+1)} - \hat{\gamma}^{(k)}) = s(\hat{\gamma}^{(k)}). \tag{8}$$

Equation (7) can be transformed to

$$\begin{pmatrix} F^{(k)}_{\alpha\alpha} & F^{(k)}_{\alpha\beta} \\ F^{(k)}_{\beta\alpha} & F^{(k)}_{\beta\beta} \end{pmatrix} \begin{pmatrix} (\hat{\alpha}^{(k+1)} - \hat{\alpha}^{(k)}) \\ (\hat{\beta}^{(k+1)} - \hat{\beta}^{(k)}) \end{pmatrix} = \begin{pmatrix} s^*(\alpha) \\ s^*(\beta) \end{pmatrix},$$

defining $s^*(\alpha) = s(\hat{\alpha}^{(k)}) + F^{(k)}_{\alpha\gamma}(\hat{\gamma}^{(k)} - \hat{\gamma}^{(k+1)})$ and $s^*(\beta) = s(\hat{\beta}^{(k)}) + F^{(k)}_{\beta\gamma}(\hat{\gamma}^{(k)} - \hat{\gamma}^{(k+1)})$. Some more transformations yield the following sequential equations to obtain first

$$\hat{\alpha}^{(k+1)} = \hat{\alpha}^{(k)} + \left(F^{(k)}_{\alpha\alpha} - F^{(k)}_{\alpha\beta}(F^{(k)}_{\beta\beta})^{-1}F^{(k)}_{\beta\alpha}\right)^{-1}\left(s^*(\alpha) - F^{(k)}_{\alpha\beta}(F^{(k)}_{\beta\beta})^{-1}s^*(\beta)\right) \tag{9}$$

and then

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + (F^{(k)}_{\beta\beta})^{-1}\left(s^*(\beta) - F^{(k)}_{\beta\alpha}(\hat{\alpha}^{(k+1)} - \hat{\alpha}^{(k)})\right). \tag{10}$$

From (8) we get the estimating equation for γ,

$$\hat{\gamma}^{(k+1)} = (\mathbf{Z}'\mathbf{W}^{(k)}\mathbf{Z} + P)^{-1}\mathbf{Z}'\mathbf{W}^{(k)}(\mathbf{y}^{(k)}_w - \mathbf{X}\hat{\alpha}^{(k+1)} - \mathbf{U}\hat{\beta}^{(k+1)}) \tag{11}$$

with the working observation $\mathbf{y}^{(k)}_w = \mathbf{X}\hat{\alpha}^{(k)} + \mathbf{U}\hat{\beta}^{(k)} + \mathbf{Z}\hat{\gamma}^{(k)} + (\mathbf{D}^{(k)\prime})^{-1}(\mathbf{y} - \boldsymbol{\mu}^{(k)})$. In all equations above $M^{(k)}$ means evaluation of the respective matrix M at $\varphi = \hat{\varphi}^{(k)}$.

Each Fisher scoring step consists of computation of equations (9), (10) and (11). But this form still allows no direct solution of the problem, since the estimating equations of $\hat{\alpha}^{(k+1)}$ and $\hat{\beta}^{(k+1)}$ include $\hat{\gamma}^{(k+1)}$ (in $s^*(\alpha)$ and $s^*(\beta)$) and vice versa. But the equations can be solved by applying the backfitting or Gauss–Seidel algorithm, used by Hastie and Tibshirani (1990) to solve generalized additive models, in each Fisher scoring step. Starting with an initial value for $\hat{\gamma}^{(k+1)}$ first $\hat{\alpha}^{(k+1)}$ and $\hat{\beta}^{(k+1)}$ are estimated following equations (9) and (10). $\hat{\alpha}^{(k+1)}$ and $\hat{\beta}^{(k+1)}$ are now initial values to compute $\hat{\gamma}^{(k+1)}$ following equation (11). This is again a starting value to compute $\hat{\alpha}^{(k+1)}$ and $\hat{\beta}^{(k+1)}$. We repeat this procedure until some termination criterion is reached.

This procedure yields the following Fisher scoring algorithm with backfitting in each scoring step to solve the unknown parameters α, β and γ of the DGLMM:

Starting values : $\hat{\alpha}^{(0)} = 0, \quad \hat{\beta}^{(0)} = 0, \quad \hat{\gamma}^{(0)} = 0$

Fisher scoring steps for $k = 0, 1, 2, \ldots$:

Solve the Fisher scoring step by the inner backfitting loop: Set $\hat{\gamma}^{(k+1)} = \hat{\gamma}^{(k)}$ and repeat computation of $\hat{\alpha}^{(k+1)}$, $\hat{\beta}^{(k+1)}$ and $\hat{\gamma}^{(k+1)}$ using equations (9), (10) and (11) until convergence of each of the three parameters.

Stop if some termination criterion is reached.

The dimensions in the estimating equations (9), (10) and (11) are still high. A further reduction is necessary to avoid problems. Due to the block diagonal structure of $F_{\beta\beta}$ equation (9) can be represented with summations over submatrices of $F^{(k)}_{\alpha\beta}$, $F^{(k)}_{\beta\beta}$ and $F^{(k)}_{\beta\alpha}$, similarly to Fahrmeir and Tutz (1997), Section 7.3.3. Equivalently, equation (10) can be replaced by separate equations for each $\hat{\beta}_i^{(k+1)}$ for $i = 1, \ldots, n$. Due to the block tridiagonal structure of $F_{\gamma\gamma} = \mathbf{Z}'\mathbf{W}\mathbf{Z} + P$ a similar partitioning for equation (11) of γ is not possible. To avoid direct inversion of $F_{\gamma\gamma}$ we use an algorithm, which utilizes the so–called LDL' factorization of the positive definite block tridiagonal matrix, as pointed out in Section 1. Fahrmeir and Kaufmann (1991) propose this algorithm, which is described in the next section, in similar form as alternative to the Kalman filter and smoother in the DGLM. In combination with the backfitting algorithm it is possible to implement this algorithm in a faster and numerically more stable way than the Kalman filter and smoother.

4 Inversion of the block tridiagonal matrix

Assuming only regular covariance matrices Q_t of the error terms v_t in the transition equations (2), the forward–backward recursive algorithm to solve the inverse of $F_{\gamma\gamma}$ presented in Fahrmeir and Kaufmann (1991) can be simplified as given below. This simplification has the effect, that the algorithm is faster than the original one. Since the whole estimating procedure given in this paper with EM algorithm (see Section 5), Fisher scoring and backfitting is very time consuming, this simplification is really useful.

A simple transformation of equation (11) yields

$$F_{\gamma\gamma}\, \hat{\gamma}^{(k+1)} = s^*_\gamma, \tag{12}$$

defining $s_\gamma^* = \mathbf{Z}'\mathbf{W}^{(k)}(\mathbf{y}_w^{(k)} - \mathbf{X}\hat{\alpha}^{(k+1)} - \mathbf{U}\hat{\beta}^{(k+1)}) = (s_0', s_1', \ldots, s_T')'$. The block tridiagonal matrix $F_{\gamma\gamma}$ has the same form as matrix P in definition (4), but with submatrices P_{rs} replaced by submatrices F_{rs} given by $F_{00} = Q_0^{-1} + P_{00}$, $F_{tt} = \sum_{i=1}^n Z_{it}W_{it}Z_{it}' + P_{tt}$ and $F_{t-1,t} = P_{t-1,t}$ for $t = 1, \ldots, T$. $F_{\gamma\gamma}$ can uniquely be factorized into the form $F_{\gamma\gamma} = LDL'$ with $D = \mathrm{diag}(D_0, D_1, \ldots, D_T)$, where $D_0, \ldots, D_T$ are positive definite matrices, and

$$L = \begin{pmatrix} I & & & 0 \\ -B_1' & I & & \\ & \ddots & \ddots & \\ 0 & & -B_T' & I \end{pmatrix}.$$

Multiplying out the factorization LDL' and comparing the result with the components of $F_{\gamma\gamma}$ yields $D_0 = F_{00}$, $B_t = -D_{t-1}^{-1}F_{t-1,t}$ and $D_t = F_{tt} + B_t'F_{t-1,t}$ for $t = 1, \ldots, T$. Equation (12) now can be solved by forward–backward recursion. First the system $L\varepsilon_T^* = s_\gamma^*$ is solved by forward recursion for the auxiliary vector $\varepsilon_T^* = (\varepsilon_0', \ldots, \varepsilon_T')'$ and then $DL'\hat{\gamma}^{(k+1)} = \varepsilon_T^*$ by backward recursion for $\hat{\gamma}^{(k+1)} = (\hat{\gamma}_0', \hat{\gamma}_1', \ldots, \hat{\gamma}_T')'$. This leads to the following algorithm solving one backfitting step for γ:

Initialization : $\quad \varepsilon_0 = s_0, \quad D_0 = F_{00}$

Forward recursion, for $t = 1, \ldots, T$:
$$\begin{aligned} B_t &= -D_{t-1}^{-1}F_{t-1,t} \\ D_t &= F_{tt} + B_t'F_{t-1,t} \\ \varepsilon_t &= s_t + B_t'\varepsilon_{t-1} \end{aligned}$$

Filter correction : $\quad \hat{\gamma}_T = D_T^{-1}\varepsilon_T$

Smoother corrections, for $t = T, \ldots, 1$: $\quad \hat{\gamma}_{t-1} = D_{t-1}^{-1}\varepsilon_{t-1} + B_t\hat{\gamma}_t$

In addition Fahrmeir and Kaufmann (1991) give a formula to get approximate covariance matrices $V_{t|T}$ of γ_t. Starting with $V_{T|T} = D_T^{-1}$ they obtain

$$V_{t-1|T} = D_{t-1}^{-1} + B_tV_{t|T}B_t', \quad t = T, \ldots, 1, \tag{13}$$

by backward recursion. They also show, that

$$B_t = V_{t-1|t-1}T_t'V_{t|t-1}^{-1}, \quad t = 1, \ldots, T, \tag{14}$$

a result from the Kalman filter and smoother, holds for the above defined B_t. Both (13) and (14) are necessary for the estimation of hyperparameters in the next section.

5 Estimation of hyperparameters

In the sections above the hyperparameters – parameters defining the distribution of the unknown parameters φ – are considered to be known. But knowledge of these hyperparameters is given only in few cases. Now we assume that hyperparameters, the covariance matrix H of β_i, the initial values a_0, Q_0 and the covariance matrices $Q_t = Q$ (independent

of $t \geq 1$) of transition model (2), are unknown and have to be estimated from the data. Estimation is based on the maximum likelihood principle and can be realized by procedures like the iterative EM algorithm of Dempster et al. (1977) avoiding numerical integrations.

Here we use a modification of the EM algorithm, the so–called EM–type algorithm, that is given in Fahrmeir and Tutz (1997, Sections 7 and 8) separately for the random effects and the dynamic model. To avoid calculation of posterior means, these are replaced by posterior modes as in Stiratelli, Laird and Ware (1984) or Wong and Mason (1985).

For the estimation of the unknown hyperparameters $\theta = (H, a_0, Q_0, Q)$ we consider the penalized log–likelihood (3) in dependence on θ,

$$PL(\varphi|\theta) = \sum_{i=1}^{n}\sum_{t=1}^{T} l_{it}(\varphi) + l(\beta|H) + l(\gamma_0|a_0, Q_0) + l(\gamma|Q),$$

and neglect constant terms not including θ in $l(\beta|H) = -\frac{n}{2}\log|H| - \frac{1}{2}\sum_{i=1}^{n}\beta_i' H^{-1}\beta_i$, $l(\gamma_0|a_0, Q_0) = -\frac{1}{2}\log|Q_0| - \frac{1}{2}(\gamma_0 - a_0)' Q_0^{-1}(\gamma_0 - a_0)$ and $l(\gamma|Q) = -\frac{T}{2}\log|Q| - \frac{1}{2}\sum_{t=1}^{T}(\gamma_t - T_t\gamma_{t-1})' Q^{-1}(\gamma_t - T_t\gamma_{t-1})$.

The expectation–step of the r–th cycle of the EM algorithm consists in computing the conditional expectation of $PL(\varphi|\theta)$ given the observations and the current iterate $\theta^{(r)}$. Because of the independence of $l_{it}(\varphi)$ of θ we only consider $S(\theta|\theta^{(r)}) = S_1(H|\theta^{(r)}) + S_2(a_0, Q_0|\theta^{(r)}) + S_3(Q|\theta^{(r)})$ with

$$\begin{aligned}
S_1(H|\theta^{(r)}) &= \mathrm{E}(\, l(\beta|H) \mid \theta^{(r)}\,) \\
&= -\frac{n}{2}\log|H| - \frac{1}{2}\sum_{i=1}^{n} \operatorname{trace}\Big(H^{-1}\mathrm{E}(\,\beta_i\beta_i' \mid \theta^{(r)}\,)\Big) \\
S_2(a_0, Q_0|\theta^{(r)}) &= \mathrm{E}(\, l(\gamma_0|a_0, Q_0) \mid \theta^{(r)}\,) \\
&= -\frac{1}{2}\log|Q_0| - \frac{1}{2}\operatorname{trace}\Big(Q_0^{-1}\mathrm{E}(\,(\gamma_0 - a_0)(\gamma_0 - a_0)' \mid \theta^{(r)}\,)\Big) \\
S_3(Q|\theta^{(r)}) &= \mathrm{E}(\, l(\gamma|Q) \mid \theta^{(r)}\,) \\
&= -\frac{T}{2}\log|Q| - \frac{1}{2}\sum_{t=1}^{T} \operatorname{trace}\Big(Q^{-1}\mathrm{E}(\,(\gamma_t - T_t\gamma_{t-1})(\gamma_t - T_t\gamma_{t-1})' \mid \theta^{(r)}\,)\Big).
\end{aligned}$$

The maximization–step of the r–th cycle of the EM algorithm consists in maximizing $S(\theta|\theta^{(r)})$ with respect to θ. This results in $H = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\beta_i\beta_i'|\theta^{(r)})$, $a_0 = \mathrm{E}(\gamma_0|\theta^{(r)})$, $Q_0 = \operatorname{cov}(\gamma_0|\theta^{(r)})$ and $Q = \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}(\,\gamma_t\gamma_t' - \gamma_t\gamma_{t-1}'T_t' - T_t\gamma_{t-1}\gamma_t' + T_t\gamma_{t-1}\gamma_{t-1}'T_t' \mid \theta^{(r)}\,)$. Considering $\mathrm{E}(\beta_i\beta_i'|\theta^{(r)}) = \operatorname{cov}(\beta_i|\theta^{(r)}) + \mathrm{E}(\beta_i|\theta^{(r)})\,\mathrm{E}(\beta_i|\theta^{(r)})'$, $\mathrm{E}(\gamma_t\gamma_t'|\theta^{(r)}) = \operatorname{cov}(\gamma_t|\theta^{(r)}) + \mathrm{E}(\gamma_t|\theta^{(r)})\,\mathrm{E}(\gamma_t|\theta^{(r)})'$ and $\mathrm{E}(\gamma_{t-1}\gamma_t'|\theta^{(r)}) = \operatorname{cov}(\gamma_{t-1}, \gamma_t|\theta^{(r)}) + \mathrm{E}(\gamma_{t-1}|\theta^{(r)})\,\mathrm{E}(\gamma_t|\theta^{(r)})'$, we see, that we need posterior means and covariances of β_i and γ_t to get estimates of the hyperparameters θ. To avoid numerical integrations in high dimensions we replace posterior means by posterior mode estimates $\hat{\beta}_i$ and $\hat{\gamma}_t$ and posterior covariances by the covariance matrices of the posterior mode estimates of β_i and γ_t resulting from the Fisher scoring algorithm of Section 3. This covariance matrices are approximately given by the corresponding submatrices of the inverse of the expected information matrix (6), $F^{-1}(\hat{\varphi})$. But even with methods for inverting partitioned matrices it is not possible to obtain the necessary submatrices, denoted by $V(\beta_i)$ and $V(\gamma_t)$, in closed form, since $F(\varphi)$ contains the

block tridiagonal matrix $F_{\gamma\gamma}$. Therefore, as approximation we consider the inverses of the submatrices $F_{\beta\beta}$ and $F_{\gamma\gamma}$ separately,

$$F_{\beta\beta}^{-1} = \begin{pmatrix} V(\beta_1) & & 0 \\ & \ddots & \\ 0 & & V(\beta_n) \end{pmatrix}, \quad F_{\gamma\gamma}^{-1} = \begin{pmatrix} V(\gamma_0) & & * \\ & \ddots & \\ * & & V(\gamma_T) \end{pmatrix}.$$

This approximation is in spirit to the BRUTO method for penalized likelihood estimation of generalized additive models (see Hastie and Tibshirani, 1990). Here in the definition of the generalized cross validation for selecting the smoothing parameters the joint additive–fit operator of all nonparametric functions is replaced by a sum over the separate linear smoothers of each function.

Hence, we use $V(\beta_i) = \widehat{\mathrm{cov}}(\beta_i|\theta^{(r)})$ and $V(\gamma_t) = \widehat{\mathrm{cov}}(\gamma_t|\theta^{(r)})$ instead of the posterior covariances of β_i and γ_t, with $V(\gamma_t) = V_{t|T}$ given by (13). Following Schneider (1986), the covariance $\mathrm{cov}(\gamma_{t-1}, \gamma_t|\theta^{(r)})$ can be replaced by $V_{t-1,t|T} = B_t V_{t|T}$ with B_t defined by (14). This results in the equations

$$\hat{H} = \frac{1}{n}\sum_{i=1}^{n}\Big(V(\beta_i) + \hat{\beta}_i\hat{\beta}_i'\Big) \tag{15}$$

$$\hat{a}_0 = \hat{\gamma}_0 \tag{16}$$

$$\hat{Q}_0 = V_{0|T} \tag{17}$$

$$\begin{aligned}\hat{Q} = \frac{1}{T}\sum_{t=1}^{T}\Big(&(\hat{\gamma}_t - T_t\hat{\gamma}_{t-1})(\hat{\gamma}_t - T_t\hat{\gamma}_{t-1})' \\ &+ V_{t|T} - V_{t|T}'B_t'T_t' - T_tB_tV_{t|T} + T_tV_{t-1|T}T_t'\Big).\end{aligned} \tag{18}$$

With these results, the EM–type algorithm for simultaneous estimation of model parameters φ and hyperparameters θ can be defined as follows:

Initialization : Choose starting values $\theta^{(0)} = (\hat{H}^{(0)}, \hat{a}_0^{(0)}, \hat{Q}_0^{(0)}, \hat{Q}^{(0)})$.

For $r = 0, 1, 2, \ldots$:

Smoothing step : Compute posterior mode estimates $\hat{\alpha}$, $\hat{\beta}_i$, $\hat{\gamma}_t$ and posterior curvatures $V(\beta_i)$, $V_{t|T}$ by the Fisher scoring algorithm in Section 3, with hyperparameters θ replaced by their current estimates $\theta^{(r)}$.

EM step : Compute $\theta^{(r+1)}$ using equations (15), (16), (17) and (18).

Stop if some termination criterion is reached.

6 Application: Business test data

In this section we apply the DGLMM to the multicategorical IFO business test data to illustrate the method with a simple econometric model. A similar modelling of this data as given below already was used in Fahrmeir (1992) and Knorr–Held (1995). For a more correct econometric modelling we refer to König, Nerlove and Oudiz (1981). The algorithms of Sections 3 to 5 are implemented in *C++*.

The IFO business test data is based on monthly questions on the tendency of realizations and expectations of variables like production, orders in hands and demand of firms in various industrial branches in Germany. Here we look at data of $n = 55$ firms of the branch "Steine und Erden" for the years 1980 to 1990 ($T = 132$). The aim is to analyze the dependency of the categorical response "production plans" P_t at month t on the explanatory variables "expected business condition" C_t, "orders in hand" O_t and "production plans in the previous month" P_{t-1}. Furthermore, we assume unobserved heterogeneity in the sample and introduce firm–specific random effects β_i to deal with.

Response and covariates P_t, C_t and O_t are given in the three ordered categories "decrease" $(-)$, "no change" $(=)$ and "increase" $(+)$. Each of these trichotomous variables is described by $l = 2$ dummy variables P_t^+, $P_t^=$ (etc.) with reference category "decrease" $(-)$, so $(1,0)$, $(0,1)$ and $(0,0)$ stand for categories $+$, $=$ and $-$, respectively. The multivariate response $y_{it} = (P_t^+, P_t^=)'$ is multinomially distributed, i.e., $y_{it} \sim M_2(1, \mu_{it})$. To deal with ordered categorical response we use the cumulative logistic model (see Fahrmeir and Tutz, 1997, for details) with the response function

$$h(\eta_{it}) = \begin{pmatrix} F(\eta_{it1}) \\ F(\eta_{it2}) - F(\eta_{it1}) \end{pmatrix}$$

and $F(x) = 1/(1 + \exp(-x))$, the logistic distribution function. The components of the linear predictor $\eta_{it} = (\eta_{it1}, \eta_{it2})'$ have the form $\eta_{itr} = \theta_{itr} + \tilde{\eta}_{itr}$ for $r = 1, 2$, where the thresholds θ_{it1}, θ_{it2}, with the restriction $\theta_{it1} < \theta_{it2}$, model the ordering of the response categories.

Model 1

First we assume that the effects of the covariates C_t^+, $C_t^=$, O_t^+, $O_t^=$, P_{t-1}^+ and $P_{t-1}^=$ are all time–dependent, i.e., there is no fixed effect α in the model. The threshold–parameters θ_{itr} are additively decomposed in unit–specific random effects b_{ri} and time–varying effects g_{rt} yielding $\theta_{itr} = b_{ri} + g_{rt}$, $r = 1, 2$. Since we have monthly data the time–varying component g_{rt} of the threshold is again decomposed in a trend component τ_{rt} and a seasonal component s_{rt}, i.e., $g_{rt} = \tau_{rt} + s_{rt}$. The component s_{rt} is modelled with $s = 12$ seasons (months) in a trigonometric form as the sum $s_{rt} = \sum_{j=1}^{[s/2]} s_{jrt}$ of $[s/2]$ cyclical components defined by

$$\begin{aligned} s_{jrt} &= s_{jr,t-1} \cos \lambda_j + \tilde{s}_{jr,t-1} \sin \lambda_j + w_{jrt} \\ \tilde{s}_{jrt} &= -s_{jr,t-1} \sin \lambda_j + \tilde{s}_{jr,t-1} \cos \lambda_j + \tilde{w}_{jrt}, \quad j = 1, \ldots, [s/2], \end{aligned}$$

with seasonal frequencies $\lambda_j = 2\pi j/s$ and mutually independent white noise processes $\{w_{jrt}\}, \{\tilde{w}_{jrt}\}$ with a common variance σ_ω^2 (see Fahrmeir and Tutz, 1997, Section 8.1.1, or Harvey, 1989, pp. 40–43, for details). The components $\tilde{s}_{jrt}$ are only required for recursive definition of s_{jrt}. We use this seasonal model because it yields nonsingular covariance matrices Q_t in transition model (2). These definitions result in parameters $\beta_i = (b_{1i}, b_{2i})'$, $\gamma_t = (\tau_{1t}, s_{11t}, \tilde{s}_{11t}, \ldots, s_{51t}, \tilde{s}_{51t}, s_{61t}, \tau_{2t}, s_{12t}, \tilde{s}_{12t}, \ldots, s_{52t}, \tilde{s}_{52t}, s_{62t}, \tilde{\gamma}_t')'$ and matrices

$$U_{it}' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad Z_{it}' = \begin{pmatrix} 1 & 1 & 0 & \cdots & 1 & 0 & 1 & 0 & & & \cdots & & & 0 & z_{it}' \\ 0 & & & \cdots & & & 0 & 1 & 1 & 0 & \cdots & 1 & 0 & 1 & z_{it}' \end{pmatrix}$$

with covariates $z'_{it} = (C^+_t, C^=_t, O^+_t, O^=_t, P^+_{t-1}, P^=_{t-1})$ and dynamic covariate effects $\tilde{\gamma}'_t = (\gamma_{t1}, \ldots, \gamma_{t6})$. The components of the linear predictor $\eta_{it} = U'_{it}\beta_i + Z'_{it}\gamma_t$ are then defined as $\eta_{itr} = b_{ri} + \tau_{rt} + s_{rt} + z'_{it}\tilde{\gamma}_t$ for $r = 1, 2$. Assuming a random walk of first order for trend components τ_{rt} and dynamic effects $\tilde{\gamma}_t$, the transition matrices T_t of transition model (2) have the form $T_t = \mathrm{diag}(1, \tilde{T}_t, 1, \tilde{T}_t, 1, \ldots, 1)$, with the submatrix

$$\tilde{T}_t = \begin{pmatrix} \cos\lambda_1 & \sin\lambda_1 & & 0 \\ -\sin\lambda_1 & \cos\lambda_1 & & \\ & & \ddots & \\ 0 & & & \cos\lambda_6 \end{pmatrix}.$$

The EM–type algorithm was initialized with $H = \mathrm{diag}(0.1, 0.1)$, $a_0 = (0, \ldots, 0)'$, $Q_0 = Q = \mathrm{diag}(0.1, \ldots, 0.1)$ and converged after 86 EM–steps.

In Figure 1 the unit–specific components of the threshold–parameters are given. In 1 (a) estimates of first and second components are plotted against each other, pointing out a lot of variation within the firms with values between about -1 to $+2$. That indicates how important it is to consider the unobserved heterogeneity. Furthermore the shown effects are highly negative correlated. Firms with negative values of $\hat{b}_{1i}$ and positive $\hat{b}_{2i}$ seem to use often the conservative answer "no change" in the response, while the opposite case with positive $\hat{b}_{1i}$ and negative $\hat{b}_{2i}$ possibly indicates firms avoiding the "no change" category in the response. Figures 1 (b) and (c) show the estimates $\hat{b}_{1i}$ and $\hat{b}_{2i}$ each plotted against the corresponding estimated variance $\mathrm{var}(\hat{b}_{ri})$. It indicates that estimates of firms avoiding the "no change" category (i.e. $\hat{b}_{1i}$ positive and $\hat{b}_{2i}$ negative) are more precise, i.e., have smaller variance. Figure 2 shows the estimated trend components of the first and second threshold parameter within pointwise one standard deviation confidence bands. The first component (lower line) shows a slightly decreasing trend while the second component (upper line) is nearly constant over time. The distinct separation of both trend components provides that the restriction $b_{1i} + \tau_{1t} + s_{1t} < b_{2i} + \tau_{2t} + s_{2t}$ to the thresholds holds. A strong seasonal pattern of the first and second seasonal component with highest points in the first quarter and lowest points in the last quarter of the year can be seen in Figure 3. Estimates of the

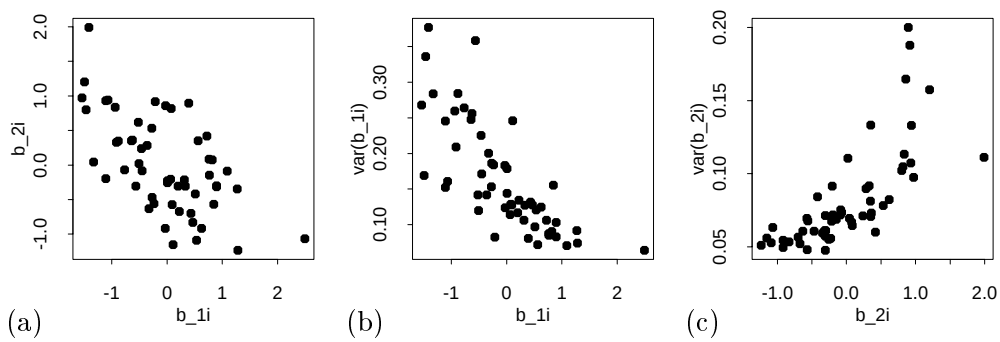

Figure 1: (a) Estimations of unit–specific parameters $\hat{b}_{1i}$ and $\hat{b}_{2i}$ plotted against each other for every unit. (b) Estimate versus estimated variance of parameter b_{1i}. (c) Estimate versus estimated variance of parameter b_{2i}. (Model 1)

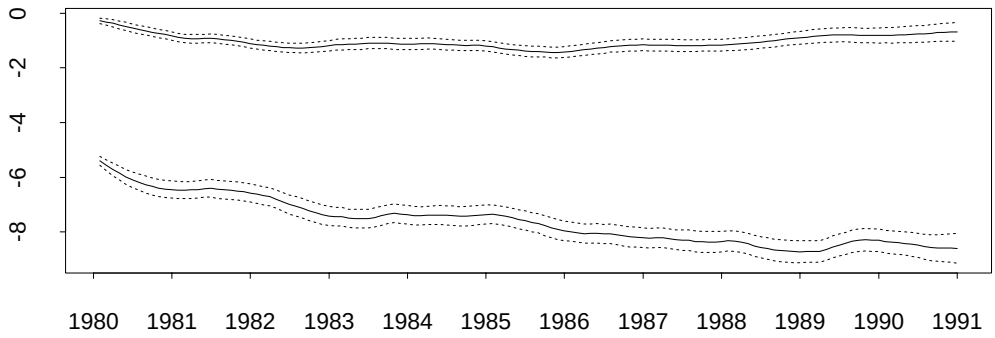

Figure 2: Estimates (solid lines) and estimated pointwise one standard deviation confidence bands (dotted lines) of the trend components of the first and second threshold parameter (model 1).

time–dependent covariate effects are shown in Figure 4. All effects are positive but nearly time–constant, except the effect of the variable "increasing expected business condition" C_t^+. Here a clear temporal variation exists with lowest point at the beginning of 1982. From that time–point the effect increases, which coincides with the establishment of a new government in Germany, what may indicate positive reactions of business to the new government.

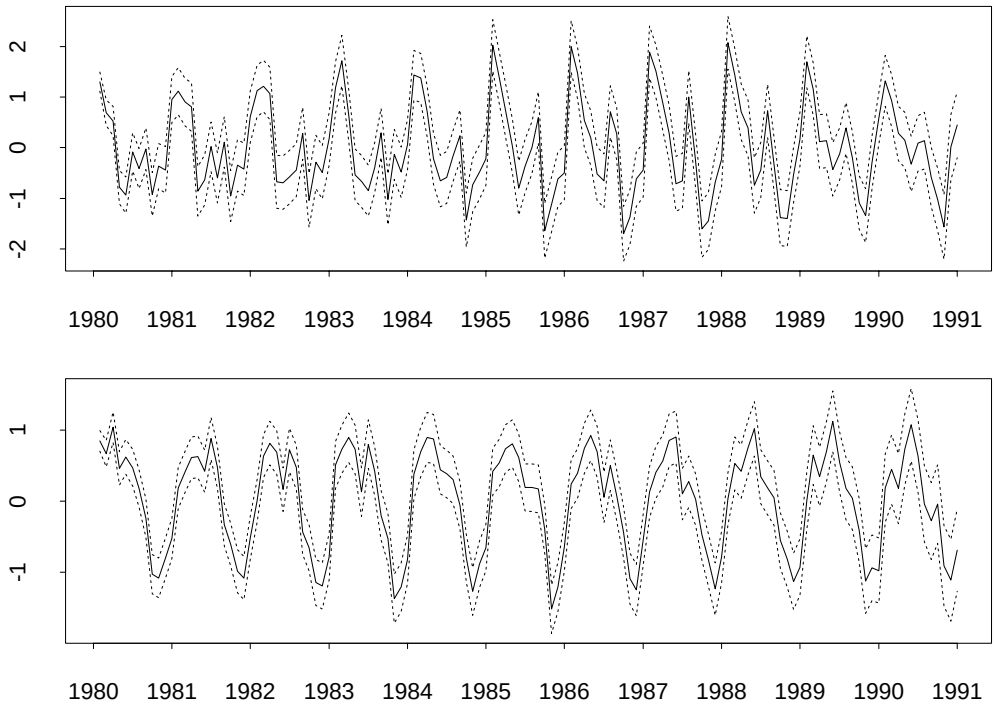

Figure 3: Estimates (solid lines) and estimated pointwise one standard deviation confidence bands (dotted lines) of the seasonal components of first (above) and second threshold parameter (model 1).

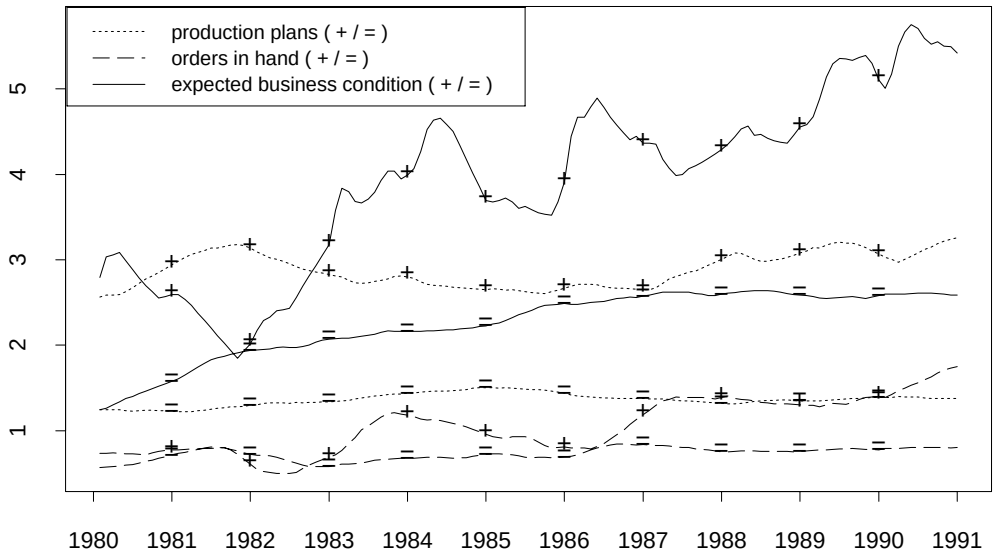

Figure 4: Estimates of time–dependent covariate effects (model 1).

Model 2

Since the effects of the covariates "orders in hand" O_t and "production plans in the previous month" P_{t-1} are nearly time–constant, we now try to model them as fixed effects, what is computationally less expensive than a time–dependent modelling. Now we use the DGLMM with linear predictor $\eta_{it} = X'_{it}\alpha + U'_{it}\beta_i + Z'_{it}\gamma_t$. The fixed effect α contains two threshold components a_1, a_2 and the effects $\tilde{\alpha}' = (\alpha_1, \ldots, \alpha_4)$ of covariates $x'_{it} = (O_t^+, O_t^=, P_{t-1}^+, P_{t-1}^=)$, i.e., α and X'_{it} are given as $\alpha = (a_1, a_2, \tilde{\alpha}')'$ and

$$X'_{it} = \begin{pmatrix} 1 & 0 & x'_{it} \\ 0 & 1 & x'_{it} \end{pmatrix}.$$

Definitions of U_{it}, β_i, Z_{it}, γ_t and T_t remain unchanged, but z_{it} and $\tilde{\gamma}_t$ within Z_{it} and γ_t are reduced to $z'_{it} = (C_t^+, C_t^=)$ and $\tilde{\gamma}'_t = (\gamma_{t1}, \gamma_{t2})$, the time–dependent effects of the covariate "expected business condition". The components of the linear predictor are now defined as $\eta_{itr} = a_r + b_{ri} + \tau_{rt} + s_{rt} + x'_{it}\tilde{\alpha} + z'_{it}\tilde{\gamma}_t$ for $r = 1, 2$. The components a_r can be seen as overall thresholds with the unit– and time–specific deviations b_{ri}, τ_{rt} and s_{rt}.

In model 2 the EM–type algorithm converged after 68 iterations. Estimates of the fixed effects α are given in the Table 1. Comparing the covariate effects with Figure 4 we see that the estimated fixed effects of O_t^+, $O_t^=$, P_{t-1}^+ and $P_{t-1}^=$ are approximately the mean values of

Thresholds		(STD)	Covariate effects		(STD)
a_1	-6.1541	(0.1753)	O_t^+	1.0905	(0.1558)
a_2	-0.0810	(0.1286)	$O_t^=$	0.7583	(0.0850)
			P_{t-1}^+	2.9527	(0.1438)
			$P_{t-1}^=$	1.4274	(0.0816)

Table 1: Estimates of the fixed effects (model 2).

the estimates of the dynamic effects in model 1. The sums of the estimated fixed threshold components and the estimated trend components, $\hat{a}_r + \hat{\tau}_{rt}$, are very near to the estimates of the trend components in model 1, see Figure 2. Also estimates of the remaining effects b_{ri}, s_{rt} and $\tilde{\gamma}_t$ are nearly the same as in model 1 and are omitted here.

7 Concluding remarks

The proposed algorithm is a useful tool for combining the GLMM and the DGLM. Due to the additive definition of the DGLMM we can use the definitions of the GLMM for estimating the fixed and the random effects and the definitions of the DGLM for estimating the dynamic effects of the DGLMM. Instead of the extended Kalman filter and smoother, which is standard in the DGLM to compute the dynamic effects, we use an algorithm that is numerically more stable and faster in combination with the backfitting algorithm. With the EM–type algorithm a simultaneous estimation of model parameters and hyperparameters is possible, where also definitions of the GLMM and the DGLM are used. For the estimation of hyperparameters, where some approximations are required, there exist other possible approaches, which are subject of further research. For estimation in the GLMM Steele (1996) presents Laplace approximations within the EM algorithm, while Wagenpfeil (1995) uses cross–validation in the DGLM. The approach of Lin and Zhang (1999), mentioned in Section 1, applies Laplace approximations in the double penalized quasi–likelihood.

References

Breslow, N. E. and Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models, *J. A. Statist. Assoc.* **88**, 9–25.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. R. Statist. Soc. B* **39**, 1–38.

Fahrmeir, L. (1992). Posterior Mode Estimation by Extended Kalman Filtering for Multivariate Dynamic Generalized Linear Models, *J. A. Statist. Assoc.* **87**, 501–509.

Fahrmeir, L. and Kaufmann, H. (1991). On Kalman Filtering, Posterior Mode Estimation and Fisher Scoring in Dynamic Exponential Family Regression, *Metrika* **38**, 37–60.

Fahrmeir, L. and Tutz, G. (1997). *Multivariate Statistical Modelling Based on Generalized Linear Models*, corrected third printing edn, Springer–Verlag, New York.

Fahrmeir, L., Kaufmann, H. and Morawitz, B. (1989). Varying Parameter Models for Panel Data with Applications to Business Test Data, *Regensburger Beiträge zur Statistik und Ökonometrie 14*, Lehrstuhl für Statistik, Universität Regensburg.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall, London.

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

Knorr–Held, L. (1995). Markov Chain Monte Carlo Simulation in Dynamic Generalized Linear Mixed Models, *Discussion Paper 8*, Sonderforschungsbereich 386, Ludwig–Maximilians–Universität München.

König, H., Nerlove, M. and Oudiz, G. (1981). On the Formation of Price Expectations, An Analysis of Business Test Data by Log–Linear Probability Models, *European Economic Review* **16**, 421–433.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines, *J. R. Statist. Soc. B* **61**, 381–400.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd. edn, Chapman and Hall, New York.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models, *J. R. Statist. Soc. A* **135**, 370–384.

Rosenberg, B. (1973). Random Coefficients Models, *Annals of Economic and Social Measurement* **2**, 399–428.

Schneider, W. (1986). *Der Kalmanfilter als Instrument zur Diagnose und Schätzung variabler Parameterstrukturen in ökonometrischen Modellen*, Physica, Heidelberg.

Steele, B. M. (1996). A Modified EM Algorithm for Estimation in Generalized Mixed Models, *Biometrics* **52**, 1295–1310.

Stiratelli, R., Laird, N. and Ware, J. H. (1984). Random–Effects Models for Serial Observations with Binary Response, *Biometrics* **40**, 961–971.

Wagenpfeil, S. (1995). Hyperparameter Estimation in Exponential Family State Space Models, *Discussion Paper 6*, Sonderforschungsbereich 386, Ludwig–Maximilians–Universität München.

Wong, G. Y. and Mason, W. M. (1985). The Hierarchical Logistic Regression Model for Multilevel Analysis, *J. A. Statist. Assoc.* **80**, 513–524.

Dr. Clemens Biller
Sonderforschungsbereich 386
Institut für Statistik
Ludwig–Maximilians–Universität München

Ludwigstr. 33
D-80539 München