Biller:

Discrete Duration Models combining Dynamic and Random Effects. (REVISED, February 2000)

Sonderforschungsbereich 386, Paper 88 (1997)

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

Projektpartner

Discrete Duration Models combining Dynamic and Random Effects

Clemens Biller

February 7, 2000

Abstract

Survival data may include two different sources of variation, namely variation over time and variation over units. If both of these variations are present, neglecting one of them can cause serious bias in the estimations. Here we present an approach for discrete duration data that includes both time–varying and unit–specific effects to model these two variations simultaneously. The approach is a combination of a dynamic survival model with dynamic time–varying baseline and covariate effects and a frailty model measuring unobserved heterogeneity with random effects varying independently over units. Estimation is based on posterior modes, i.e., we maximize the joint posterior distribution of the unknown parameters to avoid numerical integration and simulation techniques, that are necessary in a full Bayesian analysis. Estimation of unknown hyperparameters is achieved by an EM–type algorithm. Finally, the proposed method is applied to data of the Veteran's Administration Lung Cancer Trial.

Keywords

Dynamic effects, EM–type algorithm, Fisher scoring, frailty, generalized linear model, posterior mode estimation.

1 Introduction

In event history analysis, we often know only interval–censored events, i.e., we only know that the occurrence of an event lies in a time interval between two consecutive follow ups. Time therefore is divided into $\overline{T}+1$ intervals $[b_0, b_1)$, $[b_1, b_2)$, $\ldots$, $[b_{\overline{T}-1}, b_{\overline{T}})$, $[b_{\overline{T}}, b_\infty)$ to discretize the original but unknown continuous duration time. Survival time T then is measured as a discrete variable, where $T = t$ denotes failure or end of duration within the interval $[b_{t-1}, b_t)$, for $t \in \{1, \ldots, \overline{T}\}$. We assume $b_0 = 0$ and $b_{\overline{T}}$ denotes the final follow up. In addition to duration time T, a vector c_t of possibly time–dependent covariates is observed. The discrete duration process is described by the hazard function

$$\lambda(t|c_t) = P(T = t|T \geq t, c_t), \tag{1}$$

$t = 1, \ldots, \overline{T}$, the conditional probability of failure in interval $[b_{t-1}, b_t)$, given that interval is reached. The probability of surviving interval $[b_{t-1}, b_t)$ is given by the survival function

$$S(t|c_t^*) = P(T > t|c_t^*) = \prod_{s=1}^{t}(1 - \lambda(s|c_s)),$$

where $c_t^* = (c_1, \ldots, c_t)$ denotes history of covariates up to time t. To model the discrete hazard function (1) with dependence of the covariates, the binary logit model

$$\lambda(t|c_t) = P(T = t|T \geq t, c_t) = \frac{\exp(\gamma_t + c_t'\beta)}{1 + \exp(\gamma_t + c_t'\beta)} \tag{2}$$

is a standard choice, where the parameters γ_t and β are baseline and covariate effects. Alternatively, the grouped proportional hazards model, a discrete version of the Cox model, defines the discrete hazard function as

$$\lambda(t|x_t) = 1 - \exp(-\exp(\gamma_t + x_t'\beta)). \tag{3}$$

A summary of discrete time duration models is given by Fahrmeir and Tutz (1997), with generalizations to discrete models for multiple modes of failure. For more detailed information see Hamerle and Tutz (1989).

Model (2) treats the effects $\gamma_1, \ldots, \gamma_{\overline{T}}$ and β as fixed. If the number of time intervals is large, the number of unknown parameters becomes dangerously high. As a consequence, unrestricted modelling and fitting will often lead to nonexistence of maximum likelihood estimates and divergence of the used iterative estimating procedures due to the many parameters involved, especially if the data become sparse for larger $\overline{T}$. To avoid such problems, we could use a more parsimonious parameterization, for instance by using piecewise polynomials for baseline hazard functions. However, by imposing such parametric forms one may overlook unexpected patterns like peaks or seasonal

effects. Fahrmeir (1994) and Fahrmeir and Wagenpfeil (1996) propose dynamic models with a more parsimonious parameterization within the framework of generalized linear models. They allow baseline and covariate effects to vary over time, thereby following a Markovian transition model. Estimation is done by generalized Kalman filters and smoothers. This approach allows a rather flexible modelling, but it does not include unobserved heterogeneity or frailty, which often appears in event history analysis due to different responding of the observed individuals to some treatment, e.g. a chemotherapy. Neglecting heterogeneity can lead to bias in the estimates of covariate effects. But often it is not possible to collect enough covariates to account for heterogeneity. A solution is the introduction of individual–specific random effects, varying independently over the subjects. The approach including both parametric covariate effects and random effects to model frailty is standard in event history analysis, see e.g. Scheike and Jensen (1997). But this approach considers covariate effects as time–fixed, an assumption that is not always justified, since we often observe individuals over a longer period of time, where covariate effects may vary with time. See, for example, the application in Section 4.

Hence, the two mentioned approaches, the dynamic model and the random effects model, allow either modelling of dynamic effects varying over time, or the modelling of random effects, varying randomly over units. But combining these two models appears to be reasonable. Petersen, Andersen and Gill (1996) interpret the hazard function as "a result of two different sources of variation, one within subjects reflecting the risk changing over time of a given subject, and the other, the selection of individuals prone to failure, reflecting the variation among subjects. If both of these sources of variation are present and we do not include them in our analysis, both the interpretability of the hazard rate as the evolution of individual risk over time, as well as the estimates of say, treatment effects, are at best obscured and at worst seriously biased." Several authors describe the effect of neglected heterogeneity on estimation in duration models with fixed effects, see e.g. Vaupel and Yashin (1985). In the case of longitudinal data, Knorr–Held (1995) and Biller (1997) propose the dynamic generalized linear mixed model (DGLMM), a combination of the dynamic generalized linear model and the generalized linear model with random effects, that simultaneously model dynamic time–varying effects and unit–specific random effects. In the DGLMM, a direct Bayesian approach based on the posterior density of the unknown parameters involves computationally intractable high–dimensional integrations. In Knorr–Held (1995), Bayesian analysis therefore is based on Markov Chain Monte Carlo methods. An alternative, that is applied in Breslow and Clayton (1993) to generalized linear mixed models or in Fahrmeir and Wagenpfeil (1996) to dynamic generalized linear models, is maximizing the posterior density, i.e., posterior mode estimation. Based on the results of Biller

(1997) for longitudinal data, in this paper we propose the DGLMM for discrete duration data, where estimation of the unknown parameters is founded on the posterior mode principle, resulting in a Fisher scoring algorithm with backfitting steps in each scoring iteration. Estimation of unknown hyperparameters, i.e., the parameters describing the prior knowledge about the time–varying and unit–specific random effects, 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 to avoid the use of numerical integration techniques.

Section 2 introduces the data situation and the DGLMM for discrete duration data, and in Section 3 we present the algorithms for the estimation of the unknown parameters, first the Fisher scoring algorithm for the model parameters, i.e., the dynamic time–varying effects and the unit–specific random effects, and then the EM–type algorithm for estimating the hyperparameters. We give here only a brief summary of the results, for details we refer to Appendices A and B. An application of the DGLMM to discrete duration data is given in Section 4, and concluding remarks in Section 5.

2 Data situation and model definition

Observing a sample of $i = 1, \ldots, n$ units, for each of these units the true discrete survival or duration time T_i exists, but is not observable for all units. For some units we only know the censoring time C_i, the time of the last possible observation of survival of unit i. Censoring is assumed to occur at the end of the interval $[b_{t-1}, b_t)$. In the concept of random censoring the survival time T_i and censoring time C_i are independent random variables. The data are now represented by (t_i, δ_i, c_i), $i = 1, \ldots, n$, where $t_i = \min(T_i, C_i)$ denotes the observed survival time, and $\delta_i = I(T_i < C_i)$, an indicator variable for censoring. Additionally we observe possibly time–varying covariates $c_i = \{c_{i1}, \ldots, c_{i,t_i}\}$. In order to define discrete duration models in the framework of generalized linear models, we define event indicators by $y_{it} = 1$, if an event occurs in $[b_{t-1}, b_t)$ for unit i, and $y_{it} = 0$, if no event occurs in $[b_{t-1}, b_t)$ for $t = 1, \ldots, t_i$, $i = 1, \ldots, n$. With the risk set $R_t = \{i : t \leq t_i\}$, i.e., the set of units still at risk in interval $[b_{t-1}, b_t)$, we collect observations at time t in vectors $y_t = (y_{it}, i \in R_t)$, $c_t = (c_{it}, i \in R_t)$, and denote histories of event indicators and covariates up to time t by $y_t^* = (y_1, \ldots, y_t)$, $c_t^* = (c_1, \ldots, c_t)$. Given y_{t-1}^* and c_t^*, the event indicator y_{it} follows a binomial distribution

$$y_{it} | y_{t-1}^*, c_t^* \sim B(1, \mu_{it}),$$

with expectation μ_{it}. Hence, the discrete hazard function (1) of unit i can be represented as

$$\lambda(t|c_{it}) = P(y_{it} = 1|c_t^*, y_{i1} = \ldots = y_{i,t-1} = 0) = \mu_{it}.$$

To analyse the effect of the covariates on survival of the units, we link the hazard function to a linear predictor η_{it} using a response function $\lambda(t|c_{it}) = h(\eta_{it})$, where h may be the logistic distribution function (2) or the extreme–minimal–value distribution function (3).

The covariates are incorporated into the predictor by

$$\eta_{it} = X_{it}'\alpha + U_{it}'\beta_i + Z_{it}'\gamma_t. \tag{4}$$

Design vectors X_{it}, U_{it} and Z_{it} are built from the covariates c_i. The effects of the covariates are represented by the parameters α, β_i and γ_t. Here α denotes fixed effects, independent of unit i and time t, β_i measures the unit–specific effect of unit i, and $\gamma_t = (\gamma_{0t}, \tilde{\gamma}_t')'$ includes the time–varying baseline–parameter γ_{0t} and the time–varying covariate effects $\tilde{\gamma}_t$. Correspondingly the design vector $Z_{it}' = (1, \tilde{Z}_{it}')$ is built from basic covariates. If β_i and γ_t are modelled as fixed effects, we have to introduce for each unit i a unit–specific dummy variable and for each time t a time–specific dummy variable in the linear predictor. This parameterization leads to a great number of parameters to be estimated, which often results in nonexistence of maximum likelihood estimates, i.e., at least one component of the parameters tends to infinity, and as consequence, the estimating procedure diverges. For a more parsimonious parameterization we define the unit–specific and time–varying effects β_i and γ_t as random variables.

As usual in generalized linear mixed models (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 identically normal with mean zero and covariance matrix H, i.e.,

$$\beta_i \overset{iid}{\sim} N(0, H), \quad i = 1, \ldots, n.$$

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 (as $\Gamma^{-1} \to 0$)

$$p(b; H, \Gamma) \propto p(\beta; H) = \prod_{i=1}^{n} p(\beta_i; H).$$

For a parsimonious parameterization of the time–varying effects, Fahrmeir and Wagenpfeil (1996) propose to define the sequence of γ_t by the linear Markovian transition

equations

$$\begin{aligned} \gamma_t &= T_t\gamma_{t-1} + v_t, \quad t = 1, 2, \ldots \\ \gamma_0 &= a_0 + v_0. \end{aligned} \tag{5}$$

With known transition matrices $T_1, T_2, \ldots$ and independent error terms $v_t \sim N(0, Q)$, for $t \geq 1$, and $v_0 \sim N(0, Q_0)$, this approach allows the effects γ_t to vary flexibly over time t, but also penalizes for unsmooth paths of γ_t. A simple example of transition equation (5), that is used in the application in Section 4, is the first–order random walk defined by $T_t = I$, the identity matrix. From the transformation $\gamma_t - \gamma_{t-1} = v_t$ we see, that covariance $Q = 0$ leads to time–fixed effects, while $Q > 0$ results in a stochastic trend model for the γ_t with varying differences between successive effects. Histories of dynamic effects up to time t are denoted by $\gamma_t^* = (\gamma_0', \gamma_1', \ldots, \gamma_t')'$, $t = 0, 1, \ldots, \overline{T}$, and we define $\gamma = \gamma_{\overline{T}}^*$, the history up to time $\overline{T}$. 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.

As pointed out in Section 1, model (4) is a combination of the dynamic and the random effects model. Hence, these two models are submodels of (4). But there are two other submodels, the first one consisting only of α and γ_t, the other of β_i and γ_t. The derivation of the estimating procedures of these submodels from the procedure for model (4) in Section 3 is straightforward and therefore omitted.

3 Estimation of unknown parameters

In this section we present an algorithm for the simultaneous estimation of the unknown model parameters $\varphi = (\alpha', \beta', \gamma')'$ and the unknown hyperparameters $\theta = (H, a_0, Q_0, Q)$, i.e., the parameters specifying the prior distribution of the varying effects β_i and γ_t. In a first step we present an algorithm for computing the model parameters φ, assuming that hyperparameters θ are known. Afterwards, we drop this assumption and derive an algorithm for estimating θ.

Model parameters φ

Estimation is based on the posterior mode principle, i.e., we maximize the logarithm of the joint posterior density $p(\varphi|y_{\overline{T}}^*, c_{\overline{T}}^*)$ of the parameters φ (given the data $(y_{\overline{T}}^*, c_{\overline{T}}^*)$). Taking logarithms, this leads to the penalized log-likelihood

$$PL(\varphi) = l(\varphi) + a(\varphi), \tag{6}$$

with

$$l(\varphi) = \sum_{i=1}^{n} \sum_{t=1}^{t_i} \log p(y_{it}|y^*_{t-1}, c^*_t, \gamma_t, b) - \frac{1}{2}(\gamma_0 - a_0)' Q_0^{-1} (\gamma_0 - a_0),$$

the sum of individual log–likelihoods and the log–prior of γ_0. In the penalty term

$$a(\varphi) = -\frac{1}{2}\gamma' P \gamma - \frac{1}{2}\beta' L \beta \tag{7}$$

the first part results from the prior specifications of the time–varying effects γ_t with the block tridiagonal penalty matrix P. For the definition of P see Fahrmeir and Wagenpfeil (1996). The second part of (7) results from the prior specifications of the unit–specific effects β_i with the block diagonal matrix $L = \text{diag}(H^{-1}, \ldots, H^{-1})$.

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

Since the likelihood equation $s(\hat{\varphi}) = 0$ is nonlinear in $\hat{\varphi}$, we use the iterative Fisher scoring algorithm to compute the estimate $\hat{\varphi}$ of φ,

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

starting with an initial value $\hat{\varphi}^{(0)}$. Dimensions in this equation are too high for direct inversion of $F(\hat{\varphi}^{(k)})$. Therefore the Fisher scoring step is transformed into equations for each parameter, using the partitioning of the score function $s(\varphi)$ and the information matrix $F(\varphi)$. We get the scoring equations

$$\hat{\alpha}^{(k+1)} = \hat{\alpha}^{(k)} + \left(F^{(k)}_{\alpha\alpha} - F^{(k)}_{\alpha\beta}(F^{(k)}_{\beta\beta})^{-1} F^{(k)}_{\beta\alpha}\right)^{-1} \left(s^*(\alpha) - F^{(k)}_{\alpha\beta}(F^{(k)}_{\beta\beta})^{-1} s^*(\beta)\right) \tag{8}$$

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

$$\hat{\gamma}^{(k+1)} = (\mathbf{Z}'\mathbf{W}^{(k)}\mathbf{Z} + P)^{-1}\mathbf{Z}'\mathbf{W}^{(k)}(\mathbf{y}^{(k)}_w - \mathbf{X}\hat{\alpha}^{(k+1)} - \mathbf{U}\hat{\beta}^{(k+1)}) \tag{10}$$

with working matrices

$$\begin{aligned} s^*(\alpha) &= s(\hat{\alpha}^{(k)}) + F^{(k)}_{\alpha\gamma}(\hat{\gamma}^{(k)} - \hat{\gamma}^{(k+1)}) \\ s^*(\beta) &= s(\hat{\beta}^{(k)}) + F^{(k)}_{\beta\gamma}(\hat{\gamma}^{(k)} - \hat{\gamma}^{(k+1)}) \\ \mathbf{y}^{(k)}_w &= \mathbf{X}\hat{\alpha}^{(k)} + \mathbf{U}\hat{\beta}^{(k)} + \mathbf{Z}\hat{\gamma}^{(k)} + (\mathbf{D}^{(k)'})^{-1}(\mathbf{y} - \boldsymbol{\mu}^{(k)}) \,. \end{aligned}$$

A direct computation of (8) to (10) is not possible, since each of these equations contains the remaining unknown effects. Therefore we apply the iterative backfitting algorithm (see Hastie and Tibshirani, 1990) to solve each Fisher scoring step by repeated computation of (8) to (10) until convergence of each of these three parameters.

To avoid direct inversion of the block tridiagonal matrix $F_{\gamma\gamma} = \mathbf{Z}'\mathbf{W}\mathbf{Z} + P$ in equation (10), we use an algorithm that utilizes in an efficient way the structure of the matrix.

Following Biller (1997) we use an algorithm based on the LDL' factorization of $F_{\gamma\gamma}$, but adapted to the case of a diffuse prior distribution for the starting value $\gamma_0 \sim N(a_0, Q_0)$ of equation (5). Wagenpfeil (1996, p. 94) points out, that a diffuse starting prior for state space models, defined by $Q_0^{-1} = 0$, avoids problems in the estimation of a_0 and Q_0 in the EM–type algorithm, namely strong underestimation and abnormally narrow confidence bands around time $t = 0$. For details of the algorithm see Appendix B.

In contrast to the inversion of $F_{\gamma\gamma}$ there are no problems in computing the necessary inversions in (8) and (9) since $F_{\beta\beta} = \mathbf{U}'\mathbf{W}\mathbf{U} + L$ is block diagonal.

Hyperparameters $\boldsymbol{\theta}$

For estimating the hyperparameters θ, we present a modified version of the EM algorithm, which is a maximum likelihood method in incomplete data situations (see Dempster, Laird and Rubin, 1977). Here, $(y_{\overline{T}}^*, c_{\overline{T}}^*)$ are the observed, but incomplete data, the model parameters φ are the unobserved data. The joint density of the complete data $(y_{\overline{T}}^*, c_{\overline{T}}^*, \varphi)$, whose logarithm is proportional to the penalized log–likelihood $PL(\varphi)$ (see (6)), depends on the unknown hyperparameters $\theta = (H, a_0, Q_0, Q)$. For estimation of θ, the EM algorithm avoids numerical integration of the joint density of the complete data with respect to φ. With a starting value $\theta^{(0)}$ in each iteration $r = 0, 1, \ldots$ of the EM algorithm expectation– and maximization–steps are carried out until convergence. The expectation–step computes the conditional expectation $M(\theta|\theta^{(r)}) = \mathrm{E}(PL(\varphi|\theta)|\theta^{(r)})$ given the current value $\theta^{(r)}$ of the hyperparameters. The maximization–step maximizes $M(\theta|\theta^{(r)})$ with respect to θ, to obtain estimates $\theta^{(r+1)}$ of the hyperparameters. The resulting estimating equations contain posterior expectations and covariances of β_i and γ_t, that require numerical computation of high dimensional integrals. To avoid numerical integration techniques we substitute posterior modes and curvatures resulting from the Fisher scoring algorithm for posterior means and covariances of β_i and γ_t, as in Stiratelli, Laird and Ware (1984). The curvatures, i.e., the covariance matrices of the posterior mode estimates, are the corresponding submatrices on the main diagonal of the inverse of the information matrix $F^{-1}(\varphi)$. But even with methods for inverting partitioned matrices it is not possible to obtain the necessary submatrices $V(\beta_i)$ 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,

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

Hence we use $V(\beta_i) = \widehat{\text{cov}}(\beta_i|\theta^{(r)})$ and $V(\gamma_t) = \widehat{\text{cov}}(\gamma_t|\theta^{(r)})$ instead of the posterior covariances of β_i and γ_t. $V(\gamma_t)$ (and also matrices B_t, that are needed below) results from the algorithm to invert the block tridiagonal matrix $F_{\gamma\gamma}$ mentioned above.

This procedure yields estimates for H and Q, which also result in the generalized mixed and the generalized dynamic model (see Fahrmeir and Tutz, 1997, Sections 7.3 and 8.3):

$$\begin{aligned}
\hat{H} &= \frac{1}{n}\sum_{i=1}^{n}\left(V(\beta_i) + \hat{\beta}_i\hat{\beta}_i'\right) \\
\hat{Q} &= \frac{1}{\overline{T}-2}\sum_{t=2}^{\overline{T}}\Big((\hat{\gamma}_t - T_t\hat{\gamma}_{t-1})(\hat{\gamma}_t - T_t\hat{\gamma}_{t-1})' \\
&\quad + V(\gamma_t) - V(\gamma_t)'B_t'T_t' - T_tB_tV(\gamma_t) + T_tV(\gamma_{t-1})T_t'\Big)
\end{aligned}$$

Since we use a diffuse prior for the starting value γ_0, estimates for a_0 and Q_0 are not needed.

4 Veteran's Administration Lung Cancer Trial

This section illustrates joint modelling of unit–specific random effects and time–varying dynamic effects. The aim is to show what happens with the estimated heterogeneity, if we include or omit covariates with time–constant or time–varying effects, i.e., to show, which parts of the model make a contribution to the explanation of heterogeneity.

The data of the Veteran's Administration Lung Cancer Trial, presented in Kalbfleisch and Prentice (1980), includes the survival times of 137 patients with lung cancer. To compare a standard and a test chemotherapy, patients were randomized to one of the two therapy groups. In addition to the right censored survival times and the indicator for therapy group other covariates were observed. Kalbfleisch and Prentice (1980) used a Weibull and a proportional hazards model to estimate the fixed covariate effects. In their analyses only the covariates Karnofsky rating (patients' performance status measured on a scale ranging from 10 to 90, with high values indicating good performance) and tumor cell type had significant effects. Mau (1986) considered dynamic covariate effects with the regression model of Aalen (1989). Here Karnofsky rating was the only significant covariate (significance tests with the Cox and the two–step Anderson and Senthilselvan model, see Mau, 1986, for references). In both analyses the covariate therapy group was insignificant.

Since the survival times are sparse at the end of the observation period, we group them to monthly survival times, i.e., we can use the model for discrete times considered in Section 2. Following the results of Kalbfleisch and Prentice (1980) and Mau (1986) we consider only the covariates therapy group and Karnofsky rating.

In all models presented below, the time–varying effects γ_t are modelled with a first–order random walk, i.e., transition model (5) with identity transition matrices $T_t = I$. The unit–specific random effects β_i are assumed to be independent with normal distribution $N(0, H)$. The following graphs of time–varying effects γ_t in each case contain pointwise confidence bands, i.e., we do not estimate simultaneous confidence bands, but the confidence intervals of γ_t are estimated separately for each time point t by using the diagonal of the matrix $V(\gamma_t)$ presented at the end of Section 3.

We start with a simple model only including the time–varying baseline parameter:

$$\text{model 1:} \quad \eta_{it} = \gamma_{0t}.$$

Estimates are obtained by Fisher scoring combined with the EM–type algorithm. For Q we get the estimate $\hat{Q} = 0.0263$. Figure 1 showes the estimate of the dynamic baseline parameter. Here the baseline, which has to capture all possible variations over time and units, is negative and, except for the last months, strongly decreasing with time.

Figure 1 about here

Including a random effect to measure possibly existing heterogeneity yields

$$\text{model 2:} \quad \eta_{it} = \gamma_{0t} + \beta_i.$$

Now we get $\hat{Q} = 0.0017$ for the random walk variance, resulting in a nearly time–constant and negative baseline, see Figure 2. For the unknown variance H of β_i we get the estimate $\hat{H} = 0.4268$, which is a strong indication for existing heterogeneity in the data. Figure 3 shows the estimates $\hat{\beta}_i$ of the unit–specific random effects. On the x–axis the patients are ordered respective to the individual failure times t_i, i.e., at the beginning (left) is the patient with the smallest t_i, at the end (right) the patient with the greatest t_i. Here we recognize a strong time–dependence of the estimated unit–specific effects. Patients with small observed failure times have all the same positive effects, the longer the patients live the smaller (and negative) the effects $\hat{\beta}_i$ are. Patients with the greatest observed failure times have the smallest $\hat{\beta}_i$. Also patients with censored failure times ($\circ$) have negative individual effects. This result is in agreement with Aalen (1988), who points out, that individuals have different frailties (measured in our example with the unit–specific effects $\hat{\beta}_i$), and that those who are most frail (i.e., individuals with the largest effects $\hat{\beta}_i$) will die earlier than the others. The results of model 2 show, that all the variation in the data, measured in model 1 by the baseline parameter, is now measured by the unit–specific random effects $\hat{\beta}_i$, since the baseline parameter in model 2 is nearly time constant.

Figure 2 about here

Figure 3 about here

Now we include the covariate therapy group to compare the test chemotherapy ($group = 1$) with the standard chemotherapy ($group = -1$, reference category). Here we use effect coding of the dichotomous covariate, since dummy coding in dynamic models shows an unsatisfactory asymmetry in the declaration of the reference category (see Section 5.1 of Knorr–Held, 1997). Assuming the effect of *group* as fixed yields

$$\text{model 3:} \quad \eta_{it} = \gamma_{0t} + group \cdot \alpha + \beta_i.$$

The effect of *group* is estimated as $\hat{\alpha} = 0.0416$. If we have a look at the standard deviation of 0.1212 we can consider the effect as not significant as in the analyses of Kalbfleisch and Prentice (1980) and Mau (1986). All other parameters have nearly the same estimates as in model 2. The variance of the random effects β_i is with the value $\hat{H} = 0.4378$ even a little bit higher as in model 2. Here we see, that the fixed–effect approach of the insignificant covariate therapy group makes no contribution to the explanation of heterogeneity.

In the next step we include the covariate Karnofsky rating with the three categories rating 10-30, rating 40-60 and rating 70-90 following Kalbfleisch and Prentice (1980), p. 60. Since in dynamic models the same problems occur for effect coding with more than two categories as for dummy coding mentioned above (see Knorr–Held, 1997) here we use dummy coding with the two dummy variables *karnof1* and *karnof2* built as

karnof1	*karnof2*	
1	0	rating 10–30, completely hospitalized
0	1	rating 40–60, partial confinement
0	0	rating 70–90, able to care for self (reference category)

and get

$$\begin{aligned} \text{model 4:} \quad \eta_{it} &= \gamma_{0t} + group \cdot \alpha_1 + karnof1 \cdot \alpha_2 + karnof2 \cdot \alpha_3 + \beta_i \\ &= \gamma_{0t} + x_{it}'\alpha + \beta_i, \end{aligned}$$

with $x_{it}' = (group, karnof1, karnof2)$, $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$. The baseline effect γ_{0t} is again nearly time–constant and negative with random walk variance $\hat{Q} = 0.0010$. For the fixed effects α we get the estimates (and estimated standard deviations):

group :	0.0499	(0.1174)
karnof1 :	2.6956	(0.4146)
karnof2 :	0.8034	(0.1779)

As in model 3 the effect of therapy group is slightly positive and not significant. The dummies *karnof1* and *karnof2* have strongly positive effects with the effect of *karnof1* clearly above the effect of *karnof2*. That means, the patients with bad performance status (Karnofsky rating 10–30) have a much higher risk of death when compared with the reference group, i.e., the patients with high Karnofsky rating 70–90. Also patients with Karnofsky rating 40–60 (*karnof2*), have a higher risk when compared with the reference group, but a smaller risk as the patients with Karnofsky rating 10–30 (*karnof1*). In considering the standard deviations of the effects of *karnof1* and *karnof2* we could characterize the two dummies as significant. For the variance H of the unit–specific random effects we get the estimate $\hat{H} = 0.2074$, i.e., the amount of heterogeneity is only half of the value of models 2 and 3. Figure 4 shows the estimates $\hat{\beta}_i$. Here again we see the same tendency as in models 2 and 3, that patients with small observed failure times have positive effects and patients with great observed failure times or censored failure times have negative effects. But the estimates spread more around zero and the amount of the estimates is smaller, ranging only from about -0.6 to $+0.2$. From the results of model 4 we can recognize the following: with the inclusion of the significant covariate Karnofsky rating the amount of the heterogeneity in the data decreases, i.e., the significant covariate Karnofsky rating explains a great part of the heterogeneity, in contrast to the insignificant covariate therapy group, see the comments to model 2. But the estimate $\hat{H} = 0.2074$ indicates, that there is still a meaningful amount of unobserved heterogeneity in the data.

Figure 4 about here

Now we consider the covariate effects as time–varying and first include only the covariate therapy group in

$$\text{model 5:} \quad \eta_{it} \;=\; \gamma_{0t} + \mathit{group} \cdot \tilde{\gamma}_{t1} + \beta_i \;=\; z_{it}'\gamma_t + \beta_i,$$

with $z_{it}' = (1, \mathit{group})$ and $\gamma_t = (\gamma_{0t}, \tilde{\gamma}_{t1})'$. The random walk covariance has the estimate $\hat{Q} = \text{diag}(0.0018, 0.0183)$. For the baseline this small variance again yields a nearly time–constant and negative effect as in the models above, while the estimate of the time–varying effect $\tilde{\gamma}_{t1}$ of the covariate *group*, that is shown in Figure 5, is strongly decreasing till time $t = 24$. At the beginning the effect of the test chemotherapy is positive. From time $t = 4$ the effect is negative to the end of the observation period. This result indicates that from the beginning up to time $t = 4$ the standard therapy is better for survival and then it changes. But as in the models above and in the analyses of Kalbfleisch and Prentice (1980) and Mau (1986) we can consider the effect of the covariate *group* as not significant, since the zero line is included in or is close below the

confidence band. The estimation of the variance H of the unit–specific random effects yields $\hat{H} = 0.3511$, the estimates of β_i look similar as the estimates of model 2 in Figure 3, but with values ranging only from about -0.8 to $+0.3$. Comparing with model 3, where the effect of therapy group was modelled as fixed, we see, that the time–varying dynamic approach of the therapy effect explains some amount of the heterogeneity, but the bigger part still remains.

Figure 5 about here

In the next step we include the covariate Karnofsky rating with time–varying effects yielding

$$\begin{aligned} \text{model 6:} \quad \eta_{it} &= \gamma_{0t} + \mathit{group} \cdot \tilde{\gamma}_{t1} + \mathit{karnof1} \cdot \tilde{\gamma}_{t2} + \mathit{karnof2} \cdot \tilde{\gamma}_{t3} + \beta_i \\ &= z_{it}'\gamma_t + \beta_i, \end{aligned}$$

with $z_{it}' = (1, \mathit{group}, \mathit{karnof1}, \mathit{karnof2})$ and $\gamma_t = (\gamma_{0t}, \tilde{\gamma}_{t1}, \tilde{\gamma}_{t2}, \tilde{\gamma}_{t3})'$. For Q we get the estimate $\hat{Q} = \text{diag}(0.0011, 0.0246,\ 0.0361, 0.1008)$. The baseline parameter γ_{0t} is again negative and nearly time–constant. The estimates of the therapy effect $\tilde{\gamma}_{t1}$ are nearly the same as in model 5, see Figure 5. Figure 6 shows the estimates of the time–varying effects of Karnofsky rating. The dummy *karnof1* has a strongly positive and time–constant effect. That means, the patients with bad performance status (Karnofsky rating 10–30) have a much higher risk of death over the entire course of time when compared with the reference group, i.e., the patients with high Karnofsky rating 70–90. For the dummy *karnof2*, the Karnofsky rating 40–60, we have a significant positive effect on the hazard till time $t = 5$. From $t = 5$ the confidence band includes the zero line, that means, from time $t = 5$ on the effect of *karnof2* is insignificant. For the variance H of the unit–specific random effects we get the estimate $\hat{H} = 0.0245$. The estimates of β_i look similar as the estimates of model 4 in Figure 4, but with much smaller values ranging only from about -0.06 to $+0.02$. In contrast to model 4 with the time–fixed modelling of Karnofsky rating and model 5 with the introduction of the insignificant covariate therapy group with time–varying effects we see here that the time–varying modelling of the effects of the significant covariate Karnofsky rating explains almost all heterogeneity in the data, leaving only a negligible amount of unexplained heterogeneity.

Figure 6 about here

5 Concluding remarks

We propose a model that combines time–varying and unit–specific effects additively without interactions between time and units to consider the two possibly existing sources of variation in duration analysis, namely variation over time and variation over units. The example in Section 4 shows that the algorithm is able to separate the fixed, the unit–specific and the time–varying effects, and that the combination of the two varying effects has interpretable and comprehensible results. Also we recognize, which parts of the model make a contribution to the explanation of heterogeneity. So models 3 and 5 of the example reveal, that the insignificant covariate *group* contributes nothing to the explanation of heterogeneity. Only the time–varying modelling of the effect of *group* decreases the amount of the measured heterogeneity slighty. However, the introduction of the significant covariate Karnofsky rating explains a great amount of the unobserved heterogeneity. While the fixed effect approach reduces the heterogeneity by the half (model 4), the heterogeneity almost vanishes if we model the therapy effect as time–varying (model 6). Summing up, we may say that in the example the insignificant covariates make no contribution to the explanation of heterogeneity, while the significant covariates explain a great proportion, or almost all, of the heterogeneity if we use time–varying effects.

The EM–type algorithm including Fisher scoring with backfitting in each scoring step presented in Section 3 is implemented in *C++*. The software supports analysis in the dynamic generalized linear mixed model for discrete duration data defined in Section 2 including the submodels mentioned at the end of this Section.

Appendix A: Matrices $s(\varphi)$ and $F(\varphi)$

In this Appendix we define the score function $s(\varphi)$ and the expected information matrix $F(\varphi)$. We assume an informative prior distribution for the starting value $\gamma_0 \sim N(a_0, Q_0)$ of transition equation (5). For diffuse starting priors all definitions remain the same, but we have to omit all terms including Q_0 and γ_0 (see comments in Section 3).

As for the random effects model in Fahrmeir and Tutz (1997), Section 7.3.3, the components $s(\alpha)$ and $s(\beta) = (s(\beta_1)', \ldots, s(\beta_n)')'$ of $s(\varphi)$ result in

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

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

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

The component $s(\gamma) = \partial PL(\varphi)/\partial \gamma = \partial l(\varphi)/\partial \gamma + \partial a(\varphi)/\partial \gamma$ of $s(\varphi)$ (with $\partial l(\varphi)/\partial \gamma = (\partial l(\varphi)/\partial \gamma_0, \partial l(\varphi)/\partial \gamma_1, \ldots, \partial l(\varphi)/\partial \gamma_{\overline{T}})'$) is given by the elements

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

Unlike the matrix representation of $s(\gamma)$ in Fahrmeir and Wagenpfeil (1996), here we have to adapt $s(\gamma)$ to the representation of $s(\alpha)$ and $s(\beta)$, where for each i first matrices $Z_i = \mathrm{diag}(Z_{i1}', \ldots, Z_{i\overline{T}}')$ are built, and then the design matrix $Z = (Z_1', \ldots, Z_n')'$. Since here for each unit i we have different numbers t_i of observations, matrices are defined as follows: $\mathbf{Z} = \mathrm{diag}(I, Z)$, $Z = (\mathbf{Z}_1', \ldots, \mathbf{Z}_n')'$, $\mathbf{Z}_i = (Z_i, O_i)$ for $i \notin R_{\overline{T}}$ and $\mathbf{Z}_i = Z_i$ for $i \in R_{\overline{T}}$, with $Z_i = \mathrm{diag}(Z_{i1}', \ldots, Z_{i,t_i}')$ and O_i a matrix of zeros of appropriate dimension. With $\mathbf{D} = \mathrm{diag}(I, D)$, $\mathbf{\Sigma} = \mathrm{diag}(Q_0, \Sigma)$, $\mathbf{y} = (a_0', y')'$, $\boldsymbol{\mu} = (\gamma_0', \mu')'$ we get

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

For a unified representation we define matrices $\mathbf{X}' = (O', X')$, $\mathbf{U}' = (O', U')$, $K = \mathrm{diag}(O, L, P)$ (with O in each case a matrix of zeros of appropriate dimension) and $\mathbf{W} = \mathbf{D}\mathbf{\Sigma}^{-1}\mathbf{D}'$ yielding the score function and the expected information matrix

$$\begin{aligned} s(\varphi) &= (s(\alpha)', s(\beta)', s(\gamma)')' \;=\; (\mathbf{X}, \mathbf{U}, \mathbf{Z})' \mathbf{D} \mathbf{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) - K\varphi \\ F(\varphi) &= \begin{pmatrix} F_{\alpha\alpha} & F_{\alpha\beta} & F_{\alpha\gamma} \\ F_{\beta\alpha} & F_{\beta\beta} & F_{\beta\gamma} \\ F_{\gamma\alpha} & F_{\gamma\beta} & F_{\gamma\gamma} \end{pmatrix} \;=\; (\mathbf{X}, \mathbf{U}, \mathbf{Z})' \mathbf{W} (\mathbf{X}, \mathbf{U}, \mathbf{Z}) + K. \end{aligned}$$

Appendix B: Inversion of the block tridiagonal matrix

To invert the block tridiagonal matrix $F_{\gamma\gamma} = \mathbf{Z}'\mathbf{W}\mathbf{Z} + P$ in equation (10), now we adapt the algorithm in Biller (1997, Section 4) to the case of a diffuse prior distribution for the

starting value $\gamma_0 \sim N(a_0, Q_0)$ of equation (5). A simple transformation of equation (10) yields $F_{\gamma\gamma}\,\hat{\gamma}^{(k+1)} = s^*_\gamma$, with $s^*_\gamma = \mathbf{Z}'\mathbf{W}^{(k)}(\mathbf{y}_w^{(k)} - \mathbf{X}\hat{\alpha}^{(k+1)} - \mathbf{U}\hat{\beta}^{(k+1)}) = (s'_0, s'_1, \ldots, s'_{\overline{T}})'$. Defining the diffuse prior by $Q_0^{-1} = 0$, the initial values of the algorithm result in $\varepsilon_0 = 0$, $D_0 = T_1'Q^{-1}T_1$, and following $B_1 = T_1^{-1}$, $D_1 = \sum_{i\in R_1} Z_{i1}W_{i1}Z'_{i1} + T_2'Q^{-1}T_2$ and $\varepsilon_1 = s_1$. Hence, we get the following algorithm for the diffuse starting priors, without considering the time point $t = 0$:

Initialization: $\varepsilon_1 = s_1, \quad D_1 = \sum_{i\in R_1} Z_{i1}W_{i1}Z'_{i1} + T_2'Q^{-1}T_2$

Forward recursion, for $t = 2, \ldots, \overline{T}$:

$$B_t = -D_{t-1}^{-1}F_{t-1,t}, \quad D_t = F_{tt} + B_t'F_{t-1,t}, \quad \varepsilon_t = s_t + B_t'\varepsilon_{t-1}$$

Filter correction: $\hat{\gamma}_{\overline{T}} = D_{\overline{T}}^{-1}\varepsilon_{\overline{T}}, \quad V_{\overline{T}|\overline{T}} = D_{\overline{T}}^{-1}$

Smoother corrections, for $t = \overline{T}, \ldots, 2$:

$$\hat{\gamma}_{t-1} = D_{t-1}^{-1}\varepsilon_{t-1} + B_t\hat{\gamma}_t, \quad V_{t-1|\overline{T}} = D_{t-1}^{-1} + B_tV_{t|\overline{T}}B_t'.$$

Affiliation of author

Sonderforschungsbereich 386,
Institute of Statistics,
Ludwig Maximilians University Munich,
Germany

Acknowledgement

I would like to thank Prof. Dr. L. Fahrmeir for his stimulating discussions and for supervising my research project. This work was supported by a grant from the German National Science Foundation, Sonderforschungsbereich 386.

References

O. O. Aalen, "Heterogeneity in survival analysis," *Statistics in Medicine* vol. 7 pp. 1121–1137, 1988.

O. O. Aalen, "A linear regression model for the analysis of life times," *Statistics in Medicine* vol. 8 pp. 907–925, 1989.

C. Biller, "Posterior mode estimation in dynamic generalized linear mixed models," *Discussion Paper 70*, Sonderforschungsbereich 386, Ludwig-Maximilians-Universität München, 1997.

N. E. Breslow and D. G. Clayton, "Approximate inference in generalized linear mixed models," *J. A. Statist. Assoc.* vol. 88 pp. 9–25, 1993.

A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc. B* vol. 39 pp. 1–38, 1977.

L. Fahrmeir, "Dynamic modelling and penalized likelihood estimation for discrete time survival data," *Biometrika* vol. 81 pp. 317–330, 1994.

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

L. Fahrmeir and S. Wagenpfeil, "Smoothing hazard functions and time–varying effects in discrete duration and competing risks models," *J. A. Statist. Assoc.* vol. 91 pp. 1584–1594, 1996.

A. Hamerle and G. Tutz, *Diskrete Modelle zur Analyse von Verweildauern und Lebenszeiten*, Campus, New York, 1989.

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

J. Kalbfleisch and R. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley, New York, 1980.

L. Knorr–Held, "Markov chain Monte Carlo simulation in dynamic generalized linear mixed models," *Discussion Paper 8*, Sonderforschungsbereich 386, Ludwig-Maximilians-Universität München, 1995.

L. Knorr–Held, *Hierarchical Modelling of Discrete Longitudinal Data: Applications of Markov Chain Monte Carlo*, Herbert Utz Verlag Wissenschaft, München, 1997.

J. Mau, "On a graphical method for the detection of time–dependent effects of covariates in survival data," *Applied Statistics* vol. 35 pp. 245–255, 1986.

J. H. Petersen, P. K. Andersen and R. D. Gill, "Variance components models for survival data," *Statistica Neerlandica* vol. 50 pp. 193–211, 1996.

T. H. Scheike and T. K. Jensen, "A discrete survival model with random effects: An application to time to pregnancy," *Biometrics* vol. 53 pp. 318–329, 1997.

R. Stiratelli, N. Laird and J. H. Ware, "Random–effects models for serial observations with binary response," *Biometrics* vol. 40 pp. 961–971, 1984.

J. W. Vaupel and A. I. Yashin, "Heterogeneity's ruses: Some surprising effects of selection on population dynamics," *The American Statistician* vol. 39 pp. 176–185, 1985.

S. Wagenpfeil, *Dynamische Modelle zur Ereignisanalyse*, Herbert Utz Verlag Wissenschaft, München, 1996.

Figure 1: Estimates of the baseline parameter γ_{0t} plus/minus one standard deviation (model 1).

Figure 2: Estimates of the baseline parameter γ_{0t} plus/minus one standard deviation (model 2).

Figure 3: Estimates of the unit–specific effects β_i ($\bullet$ for patients with observed failure times, $\circ$ for patients with censored failure times). Patients $i = 1, \ldots, 137$ ordered respective to individual failure times t_i.(model 2)

Figure 4: Estimates of the unit–specific effects β_i ($\bullet$ for patients with observed failure times, $\circ$ for patients with censored failure times). Patients $i = 1, \ldots, 137$ ordered respective to individual failure times t_i. (model 4)

Figure 5: Estimates of the time–varying effect $\tilde{\gamma}_{t1}$ of therapy plus/minus one standard deviation (model 5).

Figure 6: Estimates of the time–varying effects $\tilde{\gamma}_{t2}$ and $\tilde{\gamma}_{t3}$ of Karnofsky rating plus/minus one standard deviation (model 6).

Figure 1

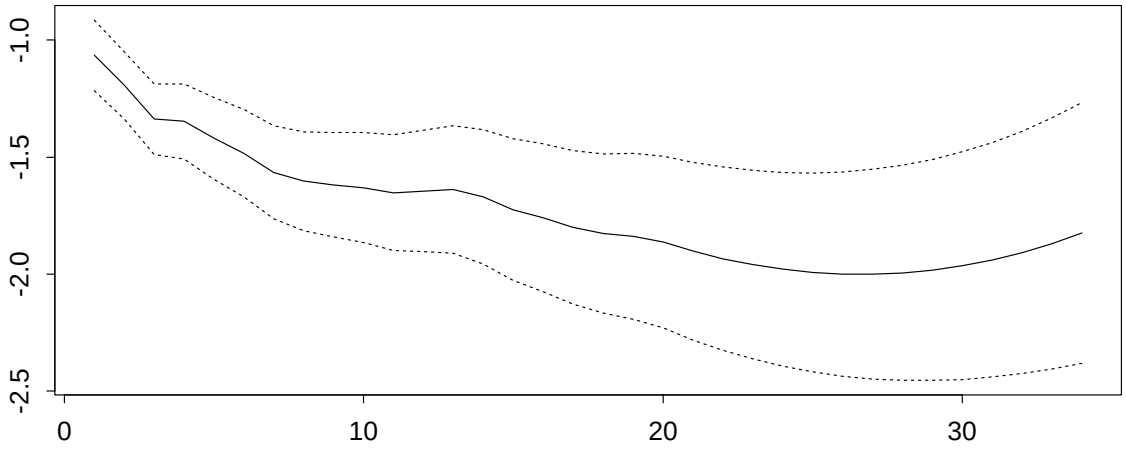

Figure 2

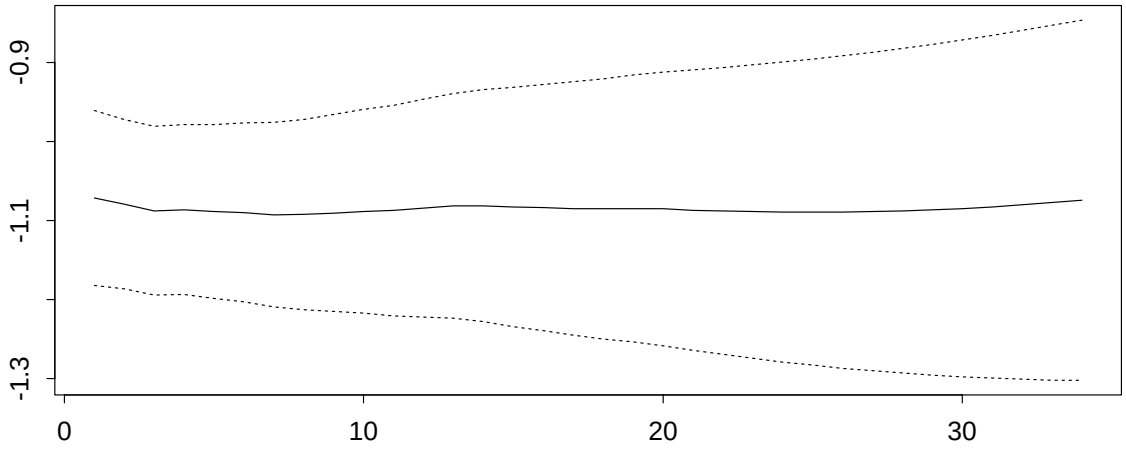

Figure 3

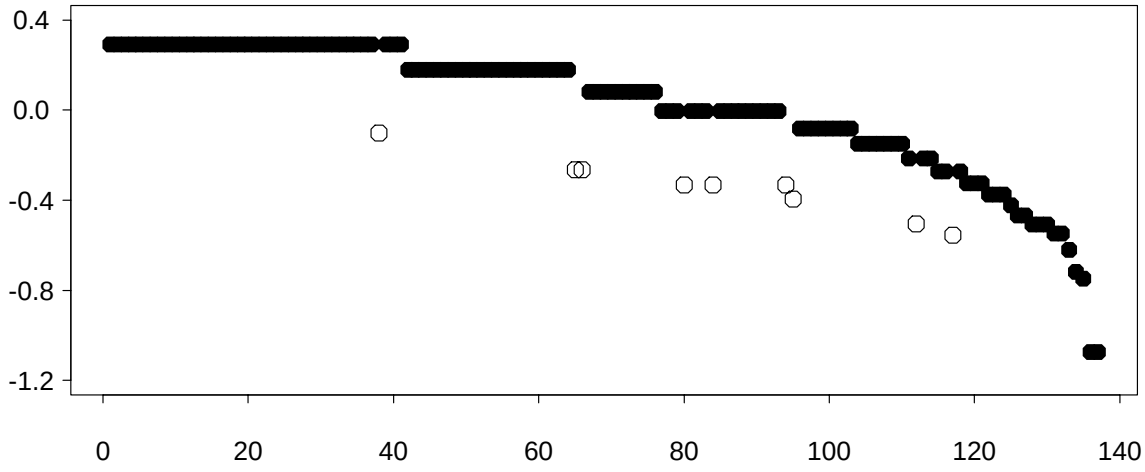

Figure 4

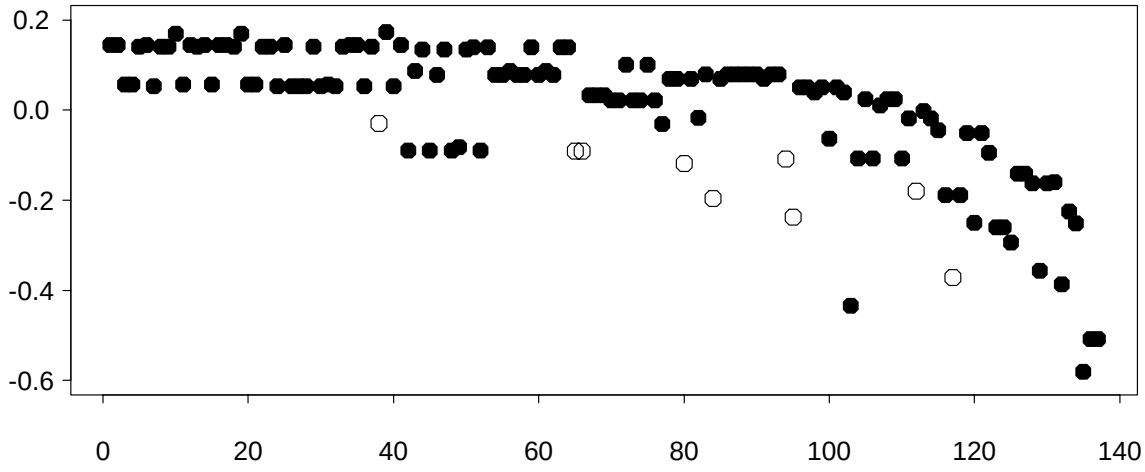

Figure 5

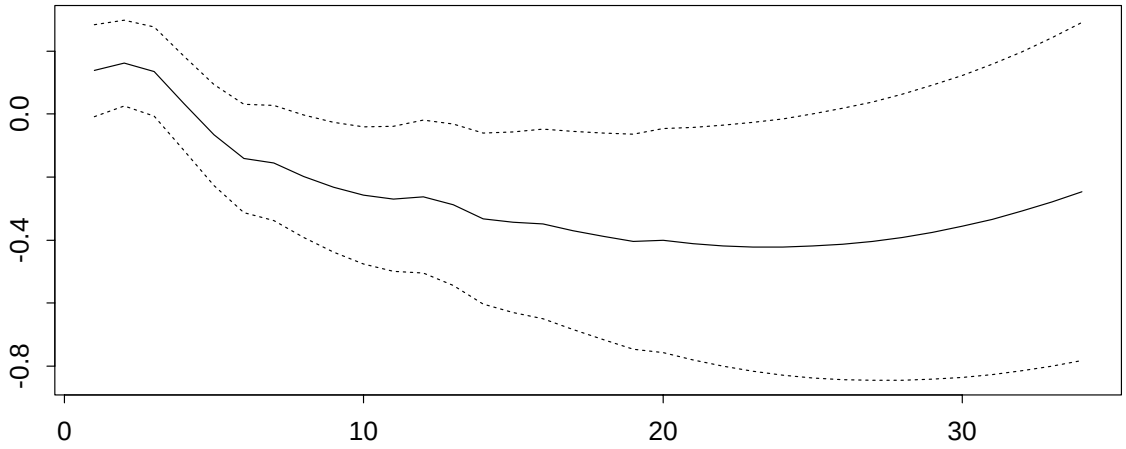

Figure 6