Stark, Wagenpfeil, Nekarda:

Estimating Time-Varying Effects of Prognostic Factors for Stomach Cancer Patients within a Dynamic Grouped Cox Model

Sonderforschungsbereich 386, Paper 29 (1996)

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

Projektpartner

Estimating Time-Varying Effects of Prognostic Factors for Stomach Cancer Patients within a Dynamic Grouped Cox Model

MONIKA STARK

Institut für medizinische Statistik und Epidemiologie, Technische Universität München, 81675 München, Ismaningerstr.22, Germany, +89/4140-4347, monika.stark@imse.med.tu-muenchen.de

STEFAN WAGENPFEIL

Institut für Statistik, Ludwig-Maximilians-Universität München, 80539 München, Ludwigstr.33, Germany, +89/2180-2232, stefan@stat.uni-muenchen.de

AND

HJALMAR NEKARDA

Chirurgische Klinik, Technische Universität München, 81675 München, Ismaningerstr. 22, Germany, +89/4140-4086

Estimating Time-Varying Effects of Prognostic Factors for Stomach Cancer Patients within a Dynamic Grouped Cox Model

SUMMARY

We describe the identification of prognostic factors in the framework of a completely resected stomach cancer survival-study. For the analysis the dynamic grouped Cox-Model was used allowing for time-varying covariate effects. Therefore the hazard rate might be non-proportional. As estimation concept we applied the posterior mode, computed by iteratively weighted Kalman filtering and smoothing steps. The medical study and questions are described, the statistical method is illustrated, the results are given and interpreted and the method is discussed.

KEYWORDS: Fisher-scoring, Iteratively weighted Kalman filtering and smoothing, Non-proportional hazards, Penalized likelihood, Posterior mode estimation, State space model, Survival analysis.

1. INTRODUCTION

In common regression models for survival analysis the prognostic factors or effects of the covariates are assumed to be constant over observation time t, e.g. using Cox-regression with the proportional-hazards assumption. Sometimes in the analysis of an extended follow-up the impact of a prognostic factor seems to vary. One reason for such a development can be a change of the effect over observation time, and the proportional hazards assumption does no longer meet. In this sense the effects of two proteolytic factors for stomach cancer patients had to be examined.

In Section 2 we introduce the data and describe the study. As the data were measured in months, we applied a discrete-time regression model for our analysis. The dynamic grouped Cox model considered in Section 3 is one of the models covering both features: time-varying effects and a discrete time-axis. It can be derived as a grouped version of the time-continuous Cox model, confer for example Kalbfleisch and Prentice[1]. To ensure the existence of the maximum penalized likelihood estimate (MPL), a transition model for the process of the covariate effects is added, acting as roughness penalty within a penalized likelihood criterion. A basic introduction and detailled description of dynamic, discrete-time survival models is given in Fahrmeir and Tutz[2].

Fahrmeir and Wagenpfeil[3] show that, from a Bayesian point of view, the posterior mode estimate can be obtained by maximizing a penalized likelihood criterion. The relevant Fisher-scoring steps are efficiently carried out by recursively applying the well-known Kalman filter and smoother method to a so-called „working score function“. The resulting algorithm is called the iteratively weighted Kalman filter and smoother, outlined in Section 4.

As a result we get the estimated time-varying effects. Furthermore, we can regard their development over time and their absolute values. This leads to the results described in Section 5. Section 6 concludes.

2. STUDY

At the Klinikum Rechts der Isar of the Technische Universität München between the years 1987 and 1994, 179 stomach cancer patients were enrolled after resection of the tumor. 84 of these patients died yet. The survival time of each person is measured in months. The median is 48 months and the maximum survival time is 92 months as can be seen from Figure 1 . Various factors were collected and their prognostic impact was investigated by Nekarda *et al* [4].

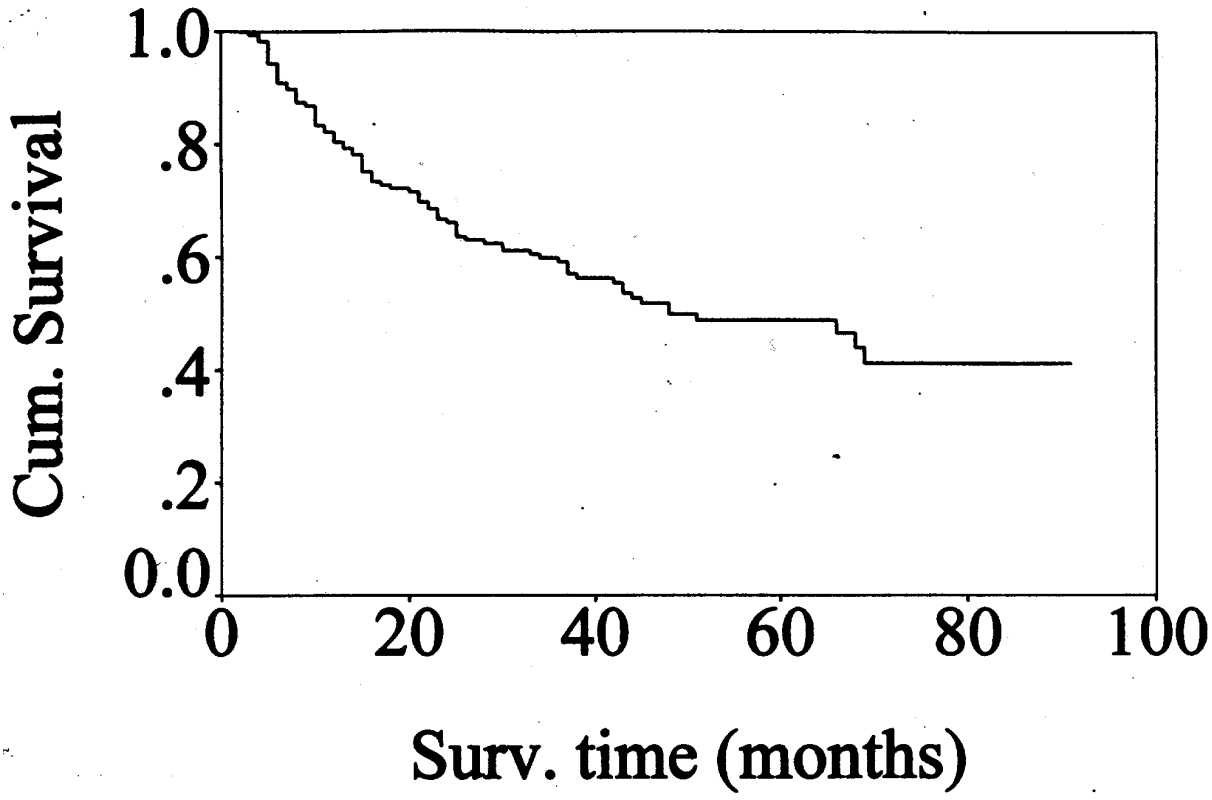

Figure 1. Survial function for stomach cancer patients after surgery (n=179).

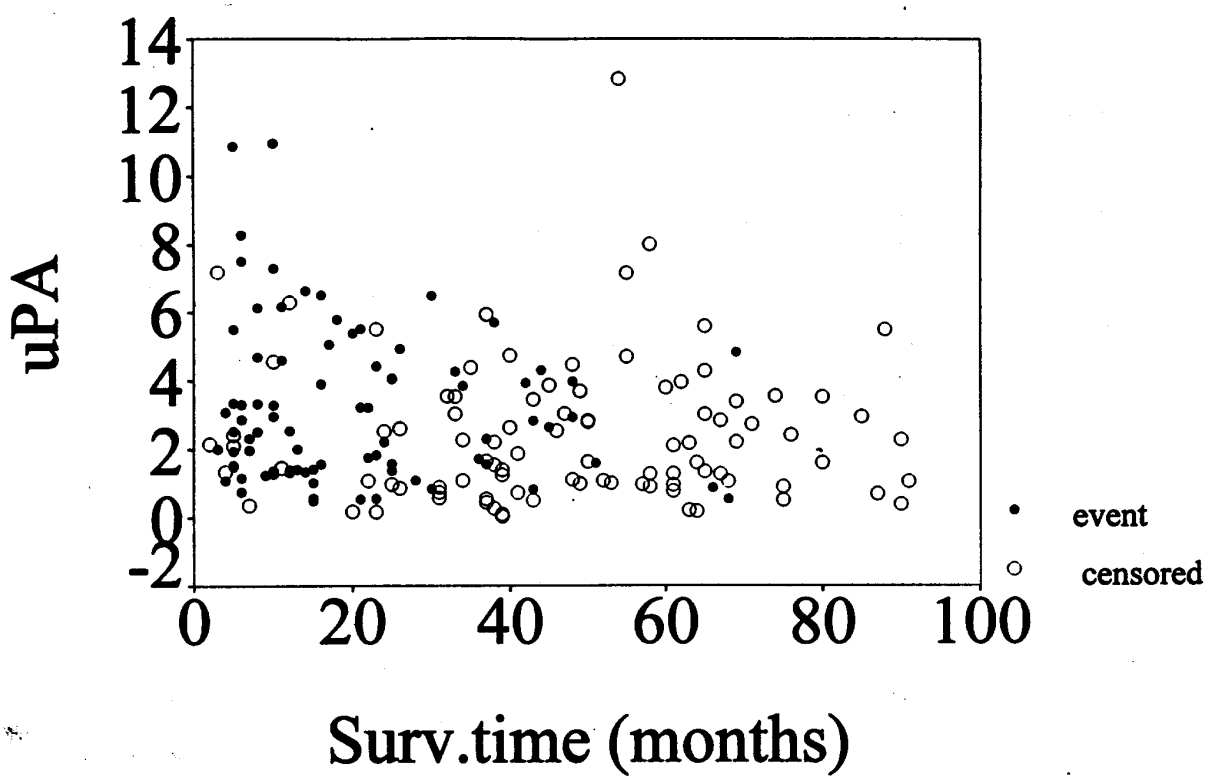

Figure 2. Scattergram of observed uPA-values, depending on survival time of patients.

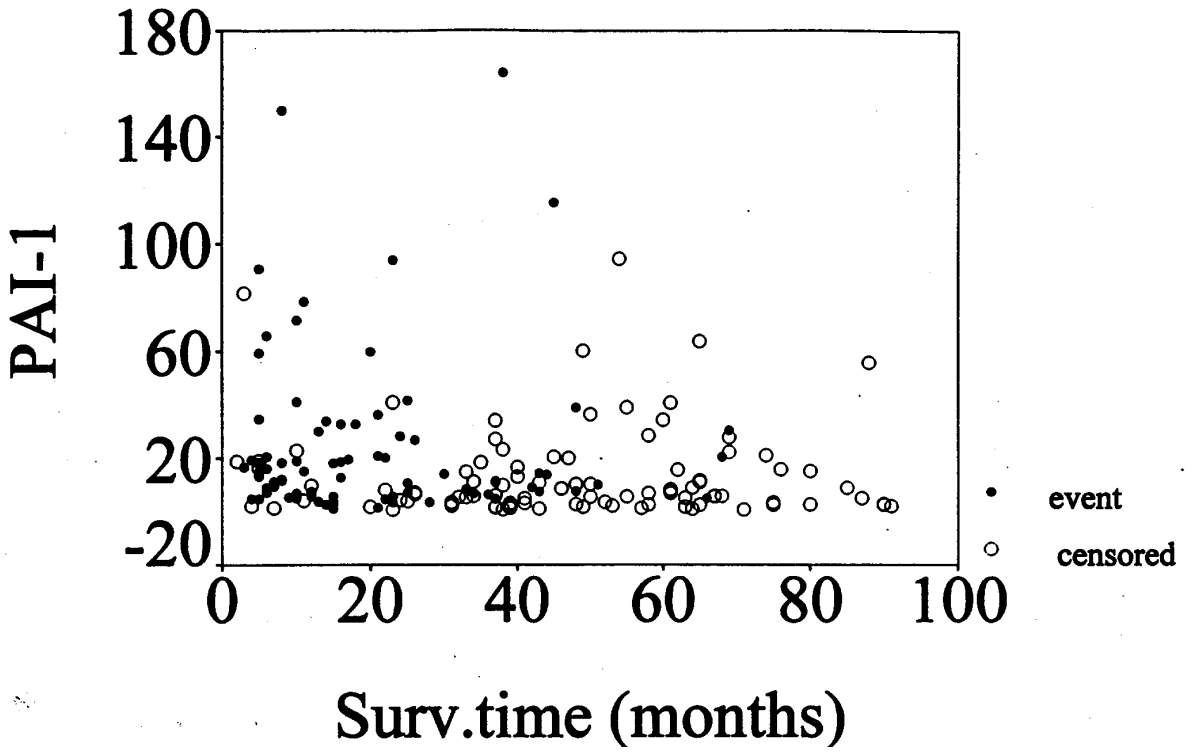

Figure 3. Scattergram of observed PAI-1-values, depending on survival time of patients.

During earlier observation time periods, two of the histomorphological factors, the urokinase-like plasminogen activator ´uPA´ and its inhibitor ´PAI-1´, showed effects different from those measured for patients with longer survival times. The plots of the uPA- and PAI-1 - values against survival time in the Figures 2 and 3 empirically indicate that both factors have a prognostic impact. For demonstration, the observation time was divided into two intervals: the first two years (24 months) and the time afterwards. UPA and PAI-1 were binary coded by a cut-off value, according to prognosis (uPA: 1.08, PAI-1: 6.23). So we set the covariate x_U = 1 if uPA $\geq$ 1.08 and x_U = 0 otherwise. In the same way we define the covariate x_P = 1 if PAI-1 $\geq$ 6.23 and x_P = 0 otherwise. From the first to the second time interval, the relative risk, estimated by two simple Cox-regression models, decreased for high uPA-values (x_U=1) from 2.5 [1.2; 5.3] to 1.9 [0.7; 5.0] and increased for high PAI-1 - values (x_P=1) from 1.9 [1.1; 3.4] to 5.9 [1.8; 19.7].

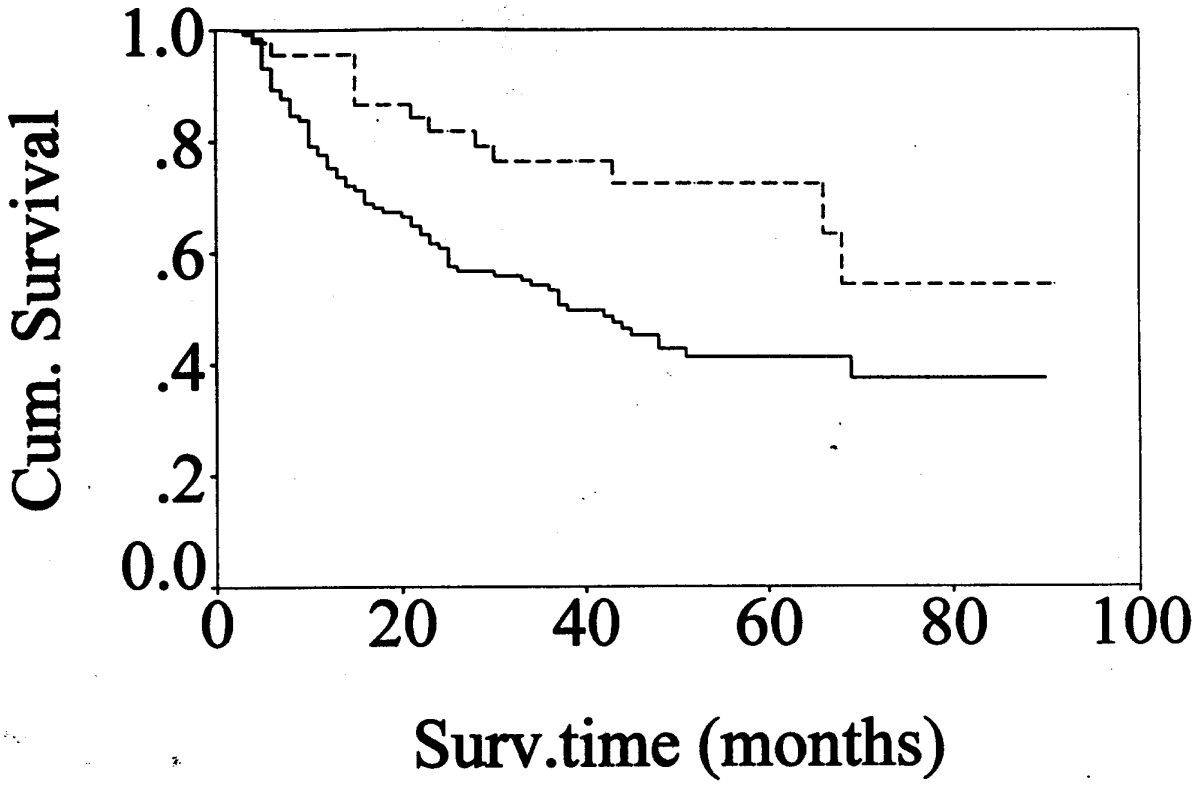

Figure 4. Survival functions for patients with low uPA (---) and with high uPA(—)

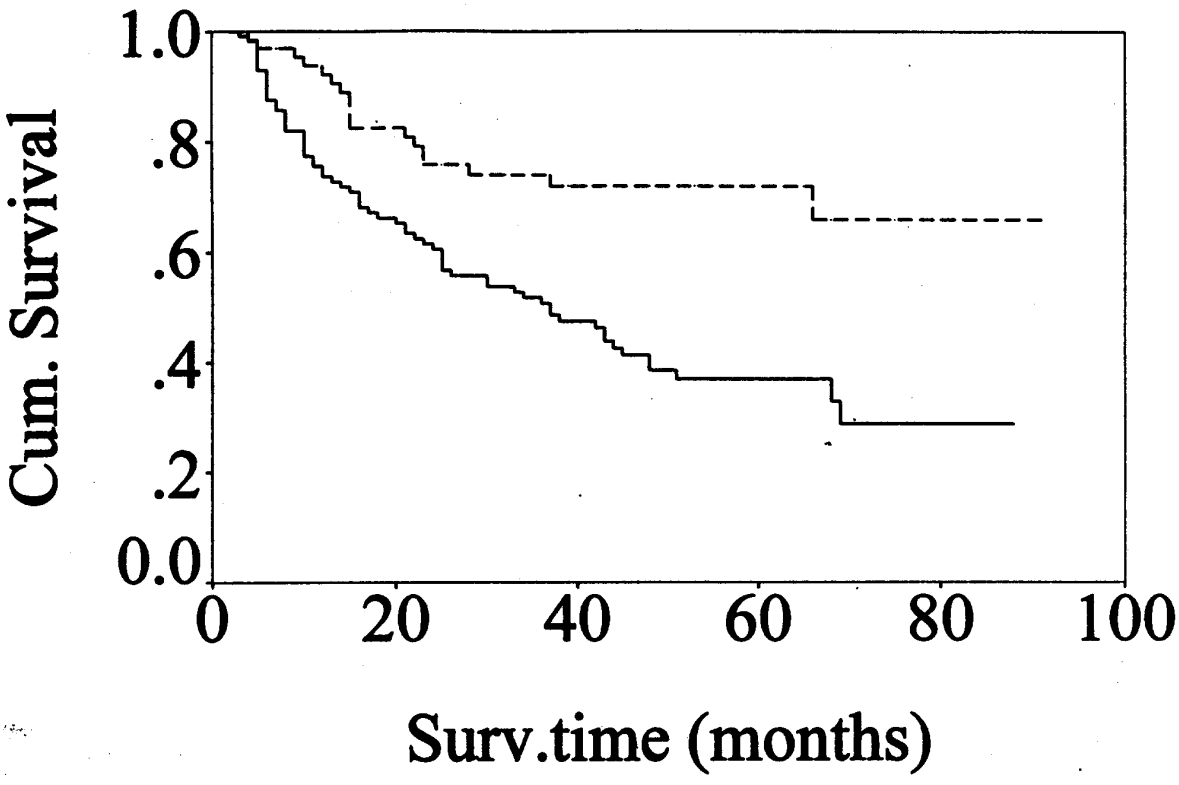

Figure 5. Survival functions for patients with low PAI-1 (---) and with high PAI-1(—)

In Figure 4, the cumulative survival functions of patients with high uPA values versus patients with low uPA were plotted, and the same is displayed for PAI-1 in Figure 5. For the first two years uPA shows an effect as the deviation between the two cumulative survivor curves increases. Later on, the effect seems to vanish. For PAI-1 the deviation between the two survivor curves continuously increases, confer Figure 5. Therefore, PAI-1 seems to have an effect that is constant or increasing over time.

The request was to investigate whether there is a change of impact of the prognostic factors uPA and PAI-1 over time or not, to describe time-varying effects by a suitable model and to compute the estimated effects in the framework of the proposed model.

3. MODEL

In the framework of survival data analysis a commonly used model specification for the continuous-time hazard function

$$\lambda(t) = \lim_{\Delta \to 0} \frac{P(t < T \leq t + \Delta | T > t)}{\Delta}$$

was introduced by Cox[5],

$$\lambda(t) = \lambda_0(t) \exp\{z'\beta\}, \tag{1}$$

where T is a continuous random variable denoting survival time and t its realization, $z = (z_1, \ldots, z_m)' \in \mathbb{R}^m$ the vector of covariates, $\beta \in \mathbb{R}^m$ a vector of unknown regression parameters modelling the effect of the covariates and $\lambda_0(t)$ the baseline hazard function with $z = (0,\ldots,0)' \in \mathbb{R}^m$.

For discrete-time data with $T \in \{1,2,\ldots\}$ Kalbfleisch and Prentice[1] derive the grouped Cox model. In this case, the discrete hazard rate

$$\lambda(t) = P(T = t | T \geq t) \tag{2}$$

is modelled by

$$\lambda(t) = 1 - \exp\{-\exp(\beta_0(t) + z'\beta)\}, \tag{3}$$

where $\beta_0(t)$ is the baseline parameter depending on time t. Within a dynamic model not only the baseline parameter but also the covariate effects are allowed to vary over time. We now

formulate a state space model and notationally we replace β with β_t and gather $\beta_0(t)$ and β_t in the state vector $\alpha_t = (\alpha_0(t), ..., \alpha_m(t))' \in \mathbb{R}^{m+1}$, confer Fahrmeir[6]. Furthermore one has to specify the way in which the unknown and unobservable state vectors α_t may vary over time. Therefor we assume a parameter transition model of the form

$$\alpha_t = F\alpha_{t-1} + \xi_t, \tag{4}$$

where F is a nonrandom transition matrix, $\{\xi_t \sim N(0, Q_t)\}$ a white noise sequence which is independent of the initial state $\alpha_0 \sim N(a_0, Q_0)$. The transition model (4) covers many well-known structural time series models like the random walks of first or higher order (e. g. Fahrmeir and Tutz[2]). In our analysis we used the random walk of first order, i. e $F = I \in \mathbb{R}^{m+1,m+1}$, where I is the unit matrix.

For each patient i = 1,...,n and time interval t = 1,...,t_{max} , the survival data are given by $(y_{it}, \delta_i, x_{U,i}, x_{P,i})$, where the values y_{it} are the failure indicators of patient i defined as

$$y_{it} = \begin{cases} 1 , & \text{patient i died in time interval t} \\ 0 , & \text{otherwise} \end{cases} , \tag{5}$$

δ_i the censoring indicator with

$$\delta_i = \begin{cases} 1 , & \text{failure in time interval t} \\ 0 , & \text{censoring in time interval t} \end{cases} \tag{6}$$

and $x_{U,i}, x_{P,i}$ the binary coded covariate values of uPA and PAI-1, respectively. Furthermore we define the risk set $R_t := \{ i \mid y_{i,1} = y_{i,2} = \ldots = y_{it} = 0 \}$ containing all patients under risk in time interval t.

Our dynamic or state space model now consists of two parts: An observation model defining the conditional likelihood of the observations given unknown baseline parameters and possibly time-varying effects, and a transition model (4) governing the stochastic transition of these parameters.

Since $\{y_{it}\}$ can be regarded as a sequence of binomial responses (death or survival), we assume as observation model

$$y_{it} \mid z_i, \delta_i, \alpha_t \sim B(1, \lambda_i(t)), \tag{7}$$

where $\lambda_i(t) = 1 - \exp\{-\exp(\eta_{it})\}$ according to (3) is the conditional probability of patient i with covariate values z_i dying in time interval t, and $\eta_{it} = (1, z_i)'\alpha_t$ is the linear predictor. Depending on whether we want to examine uPA or PAI-1, we set $z_i = x_{U,i}$ or $z_i = x_{P,i}$. Hence $\eta_{it} = \beta_0(t) + x_{U,i}\,\beta_t$ or $\eta_{it} = \beta_0(t) + x_{P,i}\,\beta_t$.

Our aim is to estimate the states or vectors of regression parameters α_t in the framework of the state space model (4) and (7).

4. ESTIMATION

In the following we gather failure indicators y_{it} of time interval t and censoring indicators δ_i in the column vectors

$$y_t = (y_{it}, i \in R_t) \text{ and } \delta^* = (\delta_1, \delta_2, ..., \delta_n).$$

Furthermore we denote histories of states and failure indicators by

$$\alpha^* = (\alpha_0, \alpha_1, ..., \alpha_{t_{max}}) \text{ and } y_t^* = (y_1, ..., y_t), \tag{8}$$

respectively. The posterior density of the unobservable states α^* given the data is $p(\alpha^* | y_{t_{max}}^*, \delta^*)$. One way of estimating the vector α^* is to compute the posterior mean. A comprehensive survey on that kind of field can be found in Fahrmeir and Tutz[2]. The method we follow is to determine the posterior mode estimate, that is the maximizer of the posterior density.

Fahrmeir[6] and Fahrmeir and Wagenpfeil[7] show that under appropriate assumptions the following proportionality holds for the posterior density:

$$p(\alpha^* | y_n^*, \delta^*) \propto \prod_{t=1}^{t_{max}} \prod_{i \in R_t} p\left(y_{it} | \alpha_t, y_{t-1}^*, \delta^*\right) \prod_{t=1}^{n} p(\alpha_t | \alpha_{t-1}) \cdot p(\alpha_0). \tag{9}$$

Incorporating the binomial density from the observation model (4) and the Gaussian density from the transition model (7), assuming that Q_0 and Q_t , t = 1,..., t_{max}, are regular and taking logarithms, we obtain the penalized loglikelihood function

$$PL(\alpha^*) = l(\alpha^*) - \frac{1}{2}\sum_{t=1}^{t_{max}} (\alpha_t - F\alpha_{t-1})' Q_t^{-1} (\alpha_t - F\alpha_{t-1}) - \frac{1}{2}(\alpha_0 - a_0)' Q_0^{-1} (\alpha_0 - a_0) \tag{10}$$

with $l(\alpha) := \sum_{t=1}^{t_{max}} \sum_{i \in R_t} l_{it}(\alpha_t)$ and the individual log-likelihood contribution

$$l_{it}(\alpha_t) := (1 - y_{it}) \log\{1 - \lambda_i(t)\} + y_{it} \log \lambda_i(t).$$

The first term in (10) measures the deviance between the data and the fit. The second and third terms penalize deviations from the transition mechanism for the parameters specified in (4), thus acting as roughness penalties and enforcing smoothness of the parameter sequence $\{\alpha_t\}$. The roughness is weighted by Q_t.

The maximization of $PL(\alpha^*)$ in (10) with respect to α^* and thus the computation of the posterior mode estimates can be achieved by various methods from nonlinear optimization theory. Fahrmeir and Wagenpfeil[3] suggest an efficient Fisher scoring method. They show that each single Fisher scoring step can be carried out by applying the well-known Kalman filter and smoother algorithm (confer Anderson and Moore[8]) to a so-called working score function $\tilde{s}_t(\alpha_t^j) := s_t(\alpha_t^j) - S_t(\alpha_t^j)\{a_{t|t-1} - \alpha_t^j\}$, where

$$s_t(\alpha_t^j) = \partial\left\{\sum_{i \in R_t} l_{it}(\alpha_t^j)\right\} \Big/ \partial\alpha_t^j = \sum_{i \in R_t} \begin{pmatrix} 1 \\ z_i \end{pmatrix} D_{it}(\alpha_t^j) \Sigma_{it}^{-1}(\alpha_t^j)\{y_{it} - \lambda_i(t)\}$$

is the score function of the likelihood and

$$S_t(\alpha_t^j) = \sum_{i \in R_t} \begin{pmatrix} 1 \\ z_i \end{pmatrix} D_{it}(\alpha_t^j) \Sigma_{it}^{-1}(\alpha_t^j) D_{it}'(\alpha_t^j) (1 \quad z_i)$$

the expected Fisher information matrix of the likelihood defined by the observation model (7). Hereby,

$$D_{it}(\alpha_t^j) = \left.\frac{\partial \lambda_i(t)}{\partial \eta_{it}}\right|_{\alpha_t^j},$$

$\lambda_i(t)$ from (3), is the first derivative of the hazard rate modelling with respect to the linear predictor η_{it}, evaluated at α_t^j,

$$\Sigma_{it}(\alpha_t^j) = \lambda_i(t)\{1 - \lambda_i(t)\}$$

is the variance of the binomial distribution in (7), and $\alpha^j = ((\alpha_0^j)', (\alpha_1^j)', \ldots, (\alpha_{t_{max}}^j)')'$ the starting vector of a Fisher scoring step. Let $a_{t|t-1}$, $a_{t|t}$ and $a_{t|t_{max}}$, t = 1,...,t_{max}, be the predicted, corrected and smoothed values of α_t^{j+1}, respectively, where $\alpha^{j+1} = ((\alpha_0^{j+1})', (\alpha_1^{j+1})', \ldots, (\alpha_{t_{max}}^{j+1})')'$

denotes the solution of the Fisher scoring step from α^j to α^{j+1}. Then the Fisher scoring step can be carried out by the following working Kalman filter and smoother recursions:

Working Kalman filter and smoother:

Initialization: $a_{0|0} = a_0$, $V_{0|0} = Q_0$.

For t = 1,..,t_{max}:

$$\text{prediction step:} \quad a_{t|t-1} = F_t a_{t-1|t-1},$$
$$V_{t|t-1} = F_t V_{t-1|t-1} F_t' + Q_t.$$
$$\text{correction step:} \quad V_{t|t} = \left\{ V_{t|t-1}^{-1} + S_t\left(\alpha_t^j\right) \right\}^{-1},$$
$$a_{t|t} = a_{t|t-1} + V_{t|t} \tilde{s}_t\left(\alpha_t^j\right).$$

For t = t_{max},...,1:

$$\text{smoother step:} \quad B_t = V_{t-1|t-1} F_t' V_{t|t-1}^{-1}$$
$$a_{t-1|92} = a_{t-1|t-1} + B_t(a_{t|92} - a_{t|t-1}),$$
$$V_{t-1|92} = V_{t-1|t-1} + B_t(V_{t|92} - V_{t|t-1})B_t' .$$

Note that the prediction vector $a_{t|t-1}$, necessary for the computation of $\tilde{s}_t\left(\alpha_t^j\right)$, is determined in the previous prediction step. The matrices $V_{t|t_{max}}$, t = 1,..., t_{max}, are the main-diagonal blocks of the inverse Fisher information matrix of (10), confer Fahrmeir and Kaufmann[9] . They were used in our analysis to construct pointwise confidence bands via the δ-method as in Fahrmeir and Wagenpfeil[3] .

The generalized extended Kalman filter and smoother described in Fahrmeir[10] is a special case of the working Kalman filter and smoother. A complete Fisher scoring algorithm is given by the iteratively weighted Kalman filter and smoother defined as follows:

Iteratively weighted Kalman Filter and smoother:

Initialization: Choose a starting vector α^0 and set the iteration index j = 0 .

Step 1: Starting with α^j, compute α^{j+1} by application of the working Kalman filter and smoother.

Step 2: If a convergence criterion is fulfilled: STOP,

else set j = j+1 and go to Step 1 .

If the convergence criterion is fulfilled in iteration k, then α^k is taken as the posterior mode estimate. The iteratively weighted Kalman filter and smoother allows for joint estimation of time-varying effects and discrete hazard functions via model (3).

So far, the hyperparameters a_0, Q_0 and Q_t in (4) were supposed to be known. In a general setting, and especially within our real data example, however, the a_0, Q_0 and Q_t will be unknown hyperparameters. The estimation can be carried out by applying the EM-principle described for this situation in Fahrmeir and Wagenpfeil[7]. We assumed $Q = Q_1 = Q_2 = \ldots = Q_{t_{max}}$ and a diffuse prior distribution of the random vector α_0 for the examination and analysis of our data. Diffuse versions of the Kalman filter algorithm are given in de Jong[11] and Wagenpfeil[12]. Wagenpfeil[13] sketches and compares further ways of hyperparameter estimation in the framework of exponential family state space models.

5. RESULTS

The analysis was done for the factors uPA and PAI-1 separately, i. e. two simple models of the form (4), (7) were used with m = 1, n=179 and t_{max} =92. In addition, uPA and PAI-1 were entered as binary factors, optimally coded belonging to prognosis, as given in Section 2. The results obtained with the iteratively weighted Kalman filter and smoother are shown in Figures 6 and 8. The time-constant coefficient delivered by the common Cox-regression is outlined (thick dashed line). It corresponds well to the time-varying estimation within the framework of model (4), (7). The baseline effects $\beta_0(t)$ are also displayed (Figures 7 and 9). The EM-type-estimated hyperparameters $\hat{Q}$ were diag(0.0007748, 0.05840038) for the uPA-model with linear predictor $\eta_{it} = \beta_0(t) + x_{U,i}\,\beta_t$ and diag(0.0010091, 0.0009996) for the PAI-1-model with $\eta_{it} = \beta_0(t) + x_{P,i}\,\beta_t$.

For uPA there is a positive effect to the hazard rate or a negative effect to survival, respectively, for the first two years. Afterwards, there is no significant effect obvious. The effect seems to be time-varying because the pointwise confidence bands are separated totally at least after one and after five years. This corresponds to the estimated survival curves, displayed in Figure 10.

For PAI-1 there is a significant negative effect to survival over the whole observation period. The effect does not seem to be time-varying because the confidence bands cover the dashed line indicating the value of the PAI-1 effect from the Cox-regression.

In the analysis containing uPA, the coefficient of the baseline risk seems to be constant, whereas in the analysis containing PAI-1, the baseline risk slightly decreases during observation time.

Using the common multiple Cox-regression over the whole observation time using uPA and PAI-1 as binary coded covariates, uPA showed no significant effect. A multiple analysis using model (4), (7) with uPA and PAI-1 as binary coded covariates showed results with a structure similar to the results in Fig. 6 and 8.

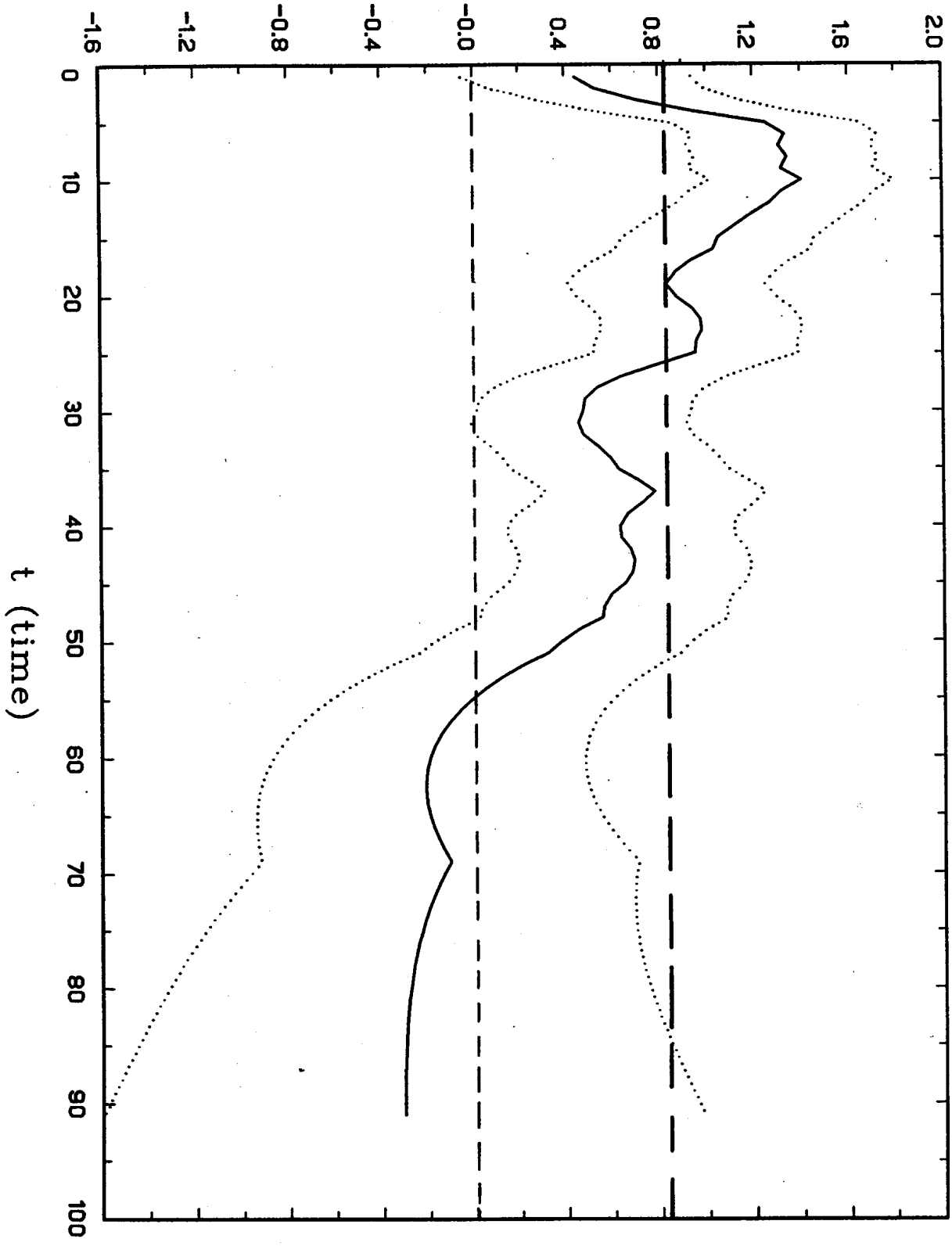

Figure 6. Estimated time-varying coefficient β_t and confidence bands for uPA.

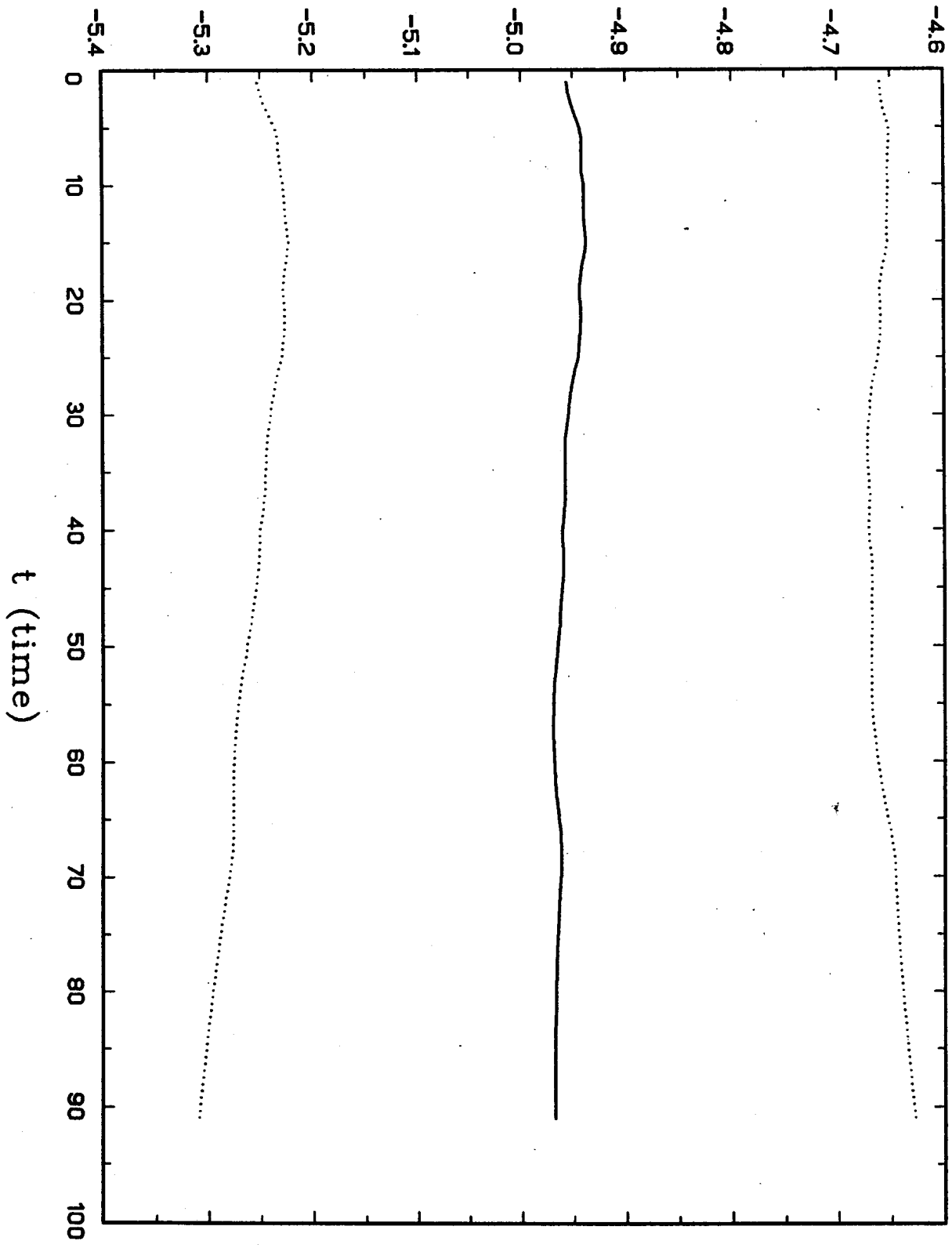

Figure 7. Estimated baseline effect $\beta_0(t)$ of the uPA-model.

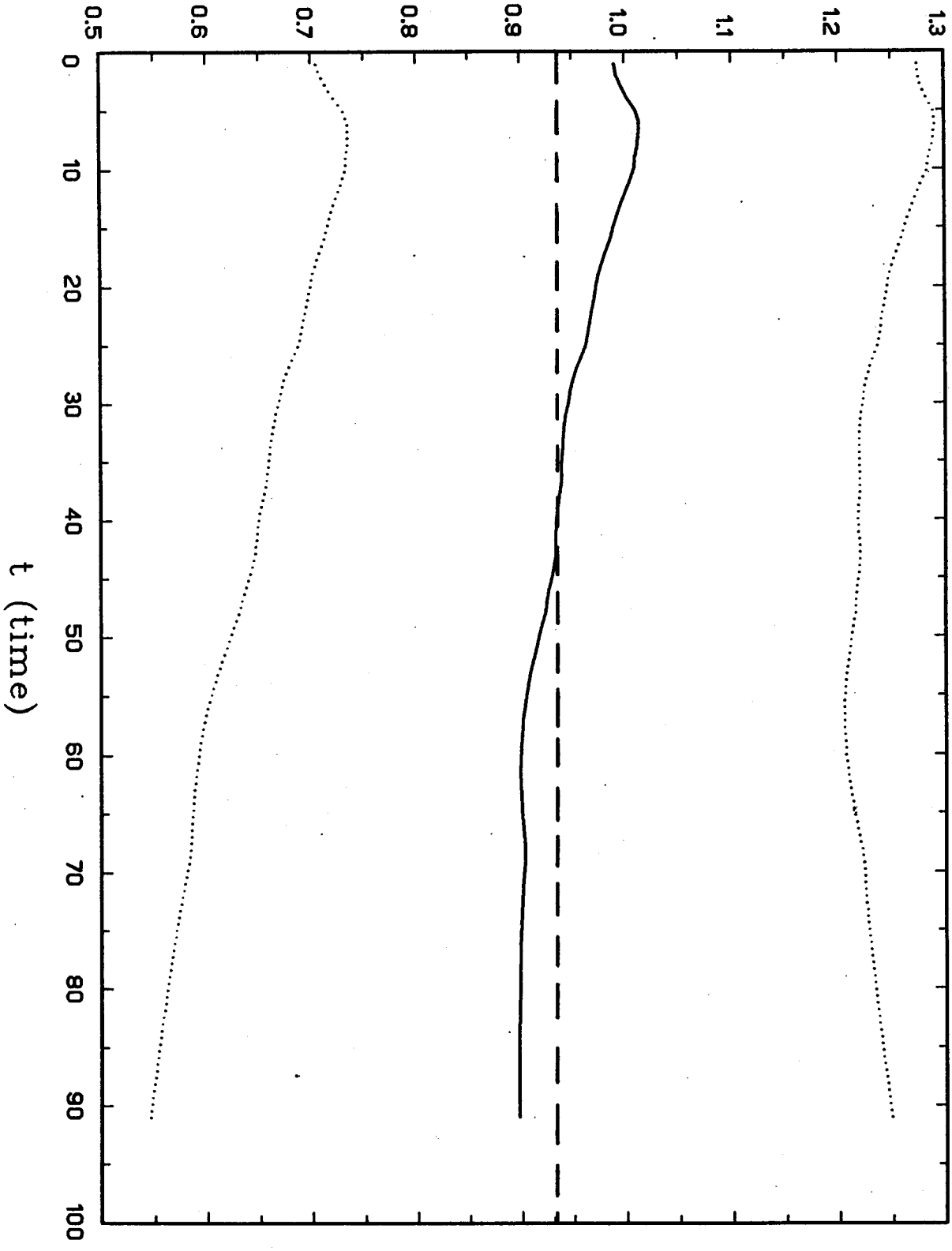

Figure 8. Estimated time-varying coefficient β_t and confidence bands for PAI-1.

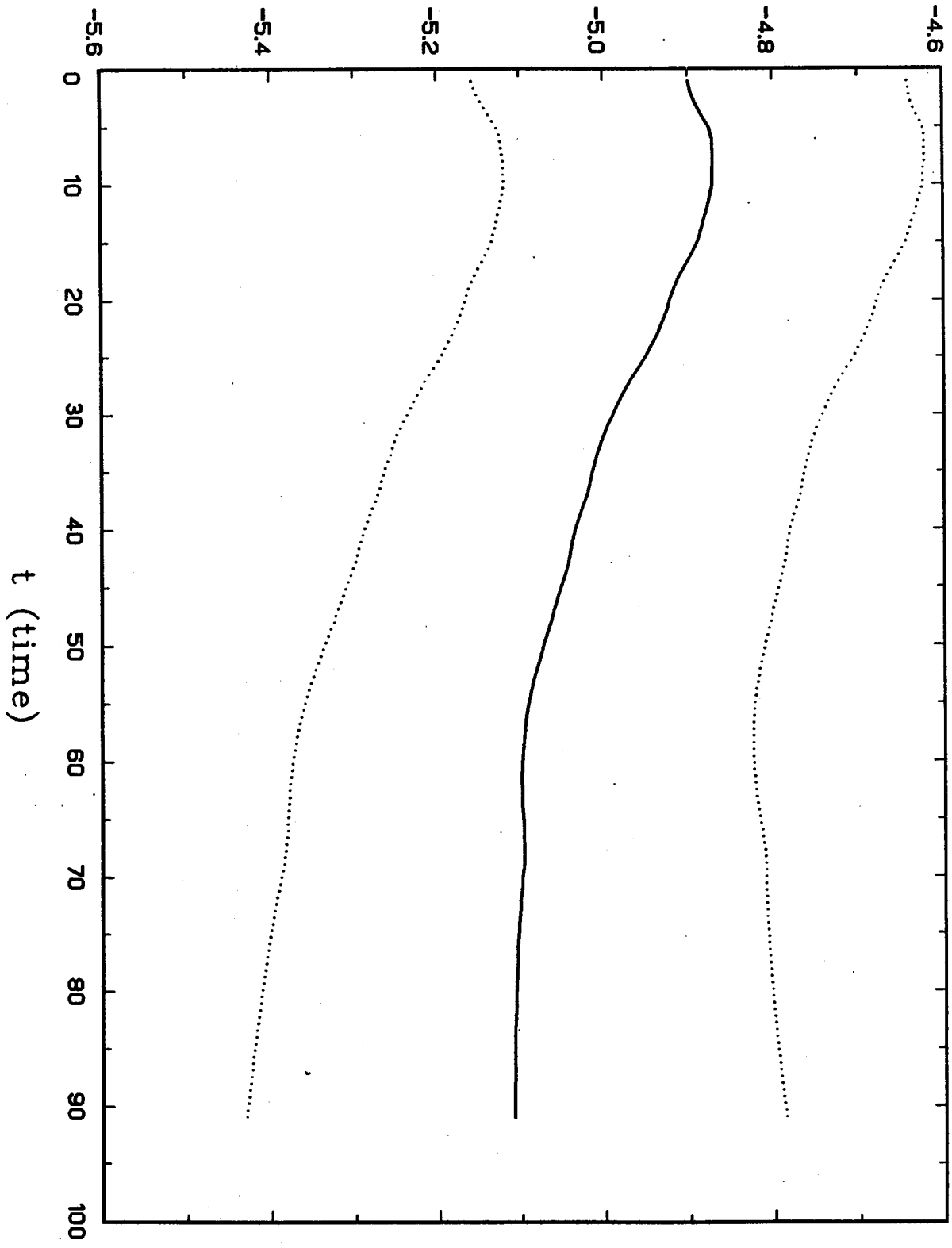

Figure 9. Estimated baseline effect $\beta_0(t)$ of the PAI-1-model.

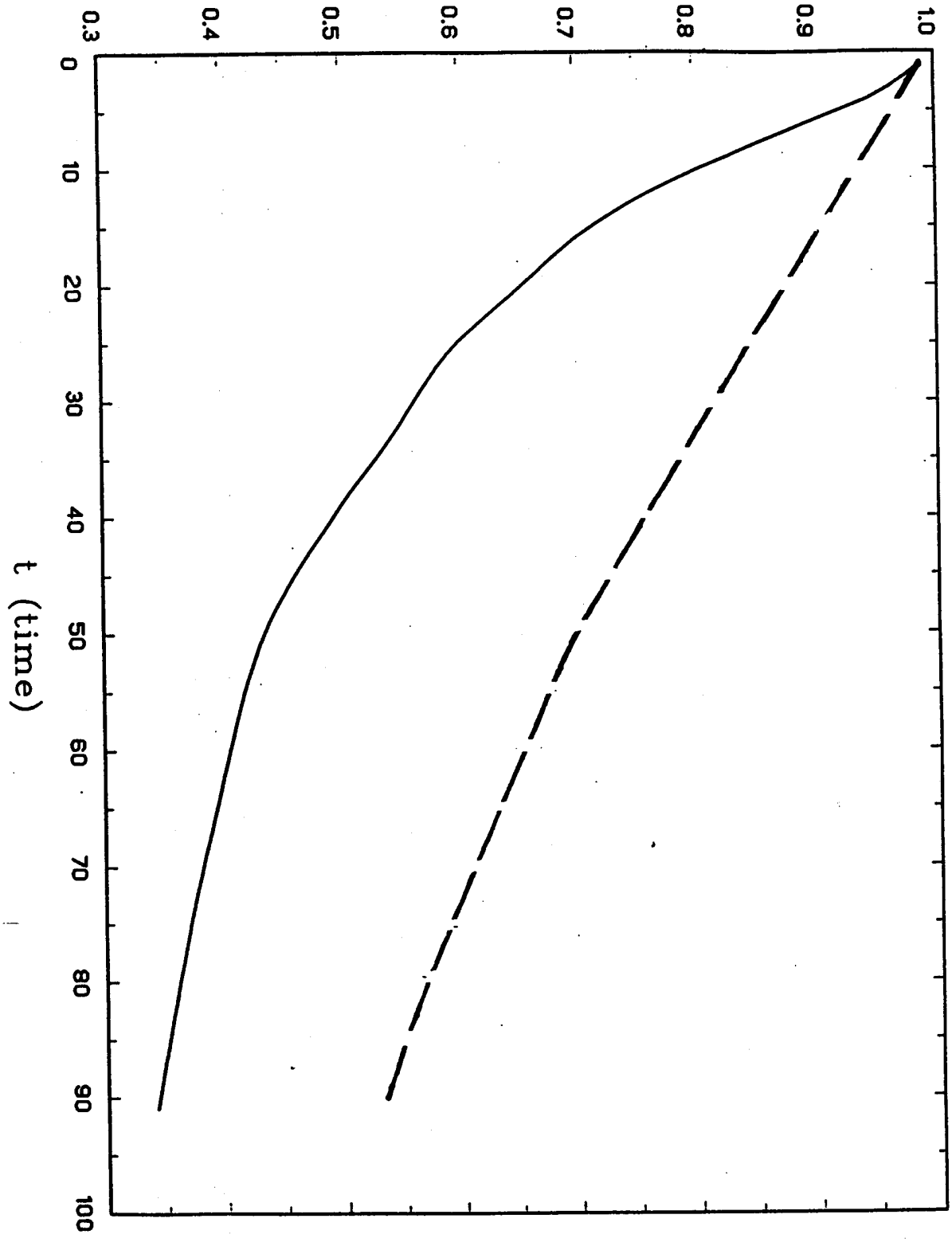

Figure 10. Survival functions for patients with low uPA (---) and with high uPA(—), estimated with the iteratively weighted Kalman filter and smoother

6. DISCUSSION

The progress in using a method, which considers time-varying effects, is made by evaluating simultaneously a) if there are effects, b) if they are time-varying c) to describe time varying-effects over time and d) the possibility to do this in a multiple analysis.

The prognostic impact of factors whose effect varies over time, e. g. changes from positive to negative, may be detected by using the above method. In a more general sense, the above model allows for simultaneous modelling or model revising, respectively.

There is a demand for further research in this method concerning the selection of factors, i. e. the comparison of the impact of factors having time-varying and of factors having time-constant effects. Multiple analyses would then combine the information about one factor: On one hand there is information about the variation of the effect over time and on the other hand there is information about the impact of the effect compared to other factors.

Like for other methods using hyperparameters, the selection of starting values for Q in the EM-type algorithm is a difficult duty and has to be managed responsibly.

Practical conclusions are to take not only PAI-1 but also uPA into account if e. g. the prognosis of patients has to be estimated by a model during earlier observation time periods. During later observation time periods uPA can be neglected.

Although our findings may still be well short of changing clinical practice at the moment, another recommendation would be a close follow-up for patients with high uPA during earlier observation time periods.

Studying the time variation of the risk associated with histomorphological factors may give important insights into their role in tumor cell dissemination and metastasis.

In common, detailed knowledge of the time-dependent risk profile of prognostic factors eventually will enable clinicians to better predict disease recurrance and survival and to individualize follow-up.

ACKNOWLEDGEMENTS

This work was supported financially by the Deutsche Forschungsgemeinschaft. We especially want to thank Prof. Dr. Ludwig Fahrmeir and Prof. Dr. Kurt Ulm for their support.

REFERENCES

1. Kalbfleisch, J.D. and Prentice, R.L. *The Statistical Analysis of Failure Time Data*, John Wiley & Sons,. New York, Chichester, 1980.

2. Fahrmeir, L. and Tutz, G. *Multivariate statistical modelling based on generalized linear models*, Springer Verlag, New York, 1994.

3. Fahrmeir, L. and Wagenpfeil, S. ´Iteratively weighted Kalman filtering and smoothing for exponential family state space models´, Discussion Paper No. 5, SFB, Ludwig-Maximilians-Universität München (1995).

4. Nekarda, H., Schmitt, M., Ulm, K., Wenninger, A., Vogelsang, H., Becker, K., Roder, J. D., Fink, U., Siewert, J. R. ´Prognostic Impact of Urokinase-type Plasminogen Activator uPA and its Inhibitor PAI-1 in Gastric Cancer with complete Resection (R0-Category, UICC)´ , Cancer Research, **54**, 2900-2907 (1994).

5. Cox, D.R. ´Regression models and life-tables (with discussion) ´, J. R. Statist. Soc. B, **34**, 187-220 (1972).

6. Fahrmeir, L. ´Dynamic modelling and penalized likelihood estimation for discrete time survival data´, Biometrica, **81**, 317-330 (1994).

7. Fahrmeir, L. and Wagenpfeil, S. ´Smoothing hazard functions and time-varying effects in discrete duration and competing risks models´, JASA, to appear (1996).

8. Anderson, B.D.O. and Moore, J.B. *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, New Jersey, 1979.

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

10. Fahrmeir, L. ´Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models´, JASA, **87**, 501-509 (1992).

11. De Jong, P. ´The diffuse Kalman Filter´, The Annals of Statistics, **19**, 1073-1083 (1991).

12. Wagenpfeil, S. *Dynamische Modelle zur Ereignisanalyse*, Inaugural-Dissertation, Referent: Prof. Fahrmeir, Koreferent: Prof. Schneeweiß, Ludwig-Maximilians-Universität München, 1995.

13. Wagenpfeil, S. ´Hyperparameter estimation in exponential family state space models´, Discussion Paper No. 6, SFB, Ludwig-Maximilians-Universität München (1995).