Survival Analysis Applied to an Actuarial Problem

Axel Gandy, Uwe Jensen and Constanze Lütkebohmert
University of Hohenheim

Abstract

Survival analysis often deals with biological and medical data. In this article we apply methods of survival analysis to an actuarial dataset to identify significant covariates for canceling a contract. In our approach we use the well known risk models suggested by Cox and Aalen which we recapitulate briefly. Furthermore, we compare the methods using a goodness of fit test.

Keywords: Survival Analysis, Actuarial Mathematics, Cox Model, Aalen Model, Goodness-of-Fit Test

Axel Gandy, e-mail: agandy@uni-hohenheim.de,
Uwe Jensen, e-mail: jensen@uni-hohenheim.de,
Constanze Lütkebohmert, e-mail: cluetkeb@uni-hohenheim.de

Institute of Applied Mathematics and Statistics
University of Hohenheim
70593 Stuttgart
Germany

1 Introduction

Usually, survival analysis is applied in medicine or biology. But there are also other fields of applications, e.g. system or software reliability (e.g. Gandy & Jensen [8]) or actuarial mathematics (e.g. Czado & Rudolph [6]). In our paper we will use models of survival analysis in an actuarial context. The dataset we are considering stems from a German insurance company and contains information about private accident insurance contracts. Generally, in survival analysis the time to death of an individual or a system is examined. We investigate the time from the conclusion until the cancellation of a contract. Our dataset does not only consist of information about the time to cancellation. There are several other attributes given about the insurance holder and the person insured: age, number of persons insured, amount of the annual premium, insurance sums covering death or disablement, etc. Our main goal is to investigate in which way the attributes influence the cancellation of contracts. The dataset consists of more than 100 000 contracts each with about 70 attributes. For our analysis we extracted a smaller dataset of about 30 000 contracts.

The models we use to examine the data are well known in survival analysis: the Cox model and the Aalen model. Cox [5] and Aalen[1] provide methods for exploring the association of covariates with failure rates. Both models are discussed in detail in Andersen et al. [3] and Fleming & Harrington [7]. We want to introduce them briefly. Of course, there exist several other regression models, see e.g. Scheike & Zhang [12] but we confine ourselves to these two.

We will consider n different contracts in both models and henceforth $\mathbf{N}(t) = (N_i(t),\ i = 1, ..., n)$ is an n-variate counting process. In our application, the time, indicated by t, measures the duration of a contract and $N_i(t) = 1$ if contract i has been canceled up to time t and $N_i(t) = 0$ otherwise. Furthermore, we transform our observed attributes into numerical covariates. For each contract i we have p covariates.

In Aalen's model we denote the value of the covariates at time t with $Y_{i1}(t), ..., Y_{ip}(t)$, which are all 0 if the individual is not at risk. We assume that the intensity $\boldsymbol{\lambda}(t) = (\lambda_1(t), ..., \lambda_n(t))$ of $\mathbf{N}(t)$ can be written as

$$\lambda_i(t) = \sum_{j=1}^{p} Y_{ij}(t)\alpha_j(t)$$

where $\alpha_j(t)$ are unknown deterministic baseline intensities which need to be estimated. In the Cox model we denote the covariates with $Z_{i1}(t), ..., Z_{ip}(t)$. It is possible that these covariates differ from those of the Aalen model. Now the intensity $\boldsymbol{\lambda}(t) = (\lambda_1(t), ..., \lambda_n(t))$ is given by

$$\lambda_i(t) = \lambda_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\} R_i(t),$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is an unknown parameter vector, $\lambda_0(t)$ is the baseline hazard function and $R_i(t)$ is a process taking only values 1 or 0 which indicates whether an individual is at risk or not.

The paper is organized as follows: Section 2 provides an introduction to the Cox model. The estimation of its parameter $\boldsymbol{\beta}$ and the integrated baseline hazard function as well as the use of martingale residuals for detecting the functional form of a covariate are described. The Aalen model and the estimation of the integrated intensity $\int_0^t \boldsymbol{\alpha}(s) \mathrm{d}s$ are explained in Section 3. A goodness of fit test of the Aalen model against the Cox model concludes this section. The paper closes with the description of the dataset and the results we found in our analysis.

2 The Cox Model

2.1 The Model

The Cox model is a multiplicative intensity model first introduced by Cox [5] and generalized for counting processes by Andersen & Gill [2]. Consider a multivariate counting process $\mathbf{N}(t) = (N_i(t), i = 1, ..., n)$, $t \in \mathcal{T} = [0, \tau]$, $0 < \tau < \infty$, where each component $N_i(t)$ indicates the number of observed events up to time t for the ith subject. In our case there is at most one event per subject: a contract can be canceled at most once. The intensity $\lambda_i(t)$ of $N_i(t)$ is related to a vector of covariates $\mathbf{Z}_i(t)$ in the following way

$$\lambda_i(t) = \lambda_0(t) R_i(t) \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\}, \ i = 1, ..., n,$$

where $R_i(t)$ is a predictable process taking only values 0 or 1 indicating whether the ith subject is at risk at time t. The vector of regression coefficients is denoted by $\boldsymbol{\beta}$. The deterministic baseline hazard function $\lambda_0(t)$ is non-negative and is left completely unspecified.

2.2 Estimation in the Cox Model

The vector of regression parameters $\boldsymbol{\beta}$ and the integrated baseline hazard

$$\Lambda_0(t) = \int_0^t \lambda_0(u)\mathrm{d}u$$

are usually estimated by partial likelihood methods (see Andersen et al. [3]). The partial likelihood is given by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{t=0}^{\tau} \left\{ \frac{R_i(t) \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(t)\}}{\sum_{j=1}^{n} R_j(t) \exp\{\boldsymbol{\beta}^T \mathbf{Z}_j(t)\}} \right\}^{\Delta N_i(t)} \tag{1}$$

where $\Delta N_i(t) = N_i(t) - N_i(t-)$. The maximizer $\hat{\boldsymbol{\beta}}$ of L is called partial maximum likelihood estimator. The cumulative hazard Λ_0 can be estimated by the Breslow estimator

$$\hat{\Lambda}_0(t) = \sum_{i=1}^{n} \int_0^t \frac{\mathrm{d}N_i(s)}{\sum_{j=1}^{n} R_j(s) \exp\{\hat{\boldsymbol{\beta}}^T \mathbf{Z}_j(s)\}}.$$

2.3 Martingale Residuals and Functional Form

The Cox model heavily relies on the functional form of the covariates $\mathbf{Z}$. In applications it is not clear whether one of the covariates, say X, should better be included in a different functional form like X^2 or $\log X$. Therneau et al. [13] suggested to use martingale residuals to determine the functional form of covariates graphically. Arguing differently we derive similar results.

We consider only one individual and drop the index i. Let X and $\mathbf{Z}$ be stochastically independent random covariates constant over time. We assume that the counting process N admits the following intensity

$$\lambda(t) = h(X) \exp\{\boldsymbol{\beta}^T \mathbf{Z}\} R(t) \lambda_0(t) = h(X)\lambda^*(t),$$

where h is an unknown positive function. Hence,

$$M(t) = N(t) - \int_0^t h(X)\lambda^*(s)\mathrm{d}s = N(t) - h(X)\Lambda^*(t)$$

is a mean zero martingale. Forming conditional expectation with respect to X we get

$$\mathbb{E}[M(t)|X] = \mathbb{E}[N(t)|X] - h(X)\mathbb{E}[\Lambda^*(t)|X].$$

Since this is again a mean zero martingale we set, heuristically, $\mathbb{E}[M(\tau)|X]$ equal to zero and get

$$h(X) \approx \frac{\mathbb{E}[N(\tau)|X]}{\mathbb{E}[\Lambda^*(\tau)|X]} = \left(1 - \frac{\mathbb{E}[N(\tau) - \Lambda^*(\tau)|X]}{\mathbb{E}[N(\tau)|X]}\right)^{-1}.$$

In particular, we are interested in $f(X) := \log h(X)$. Using a first order Taylor expansion we get

$$f(X) \approx -\log\left(1 - \frac{\mathbb{E}[N(\tau) - \Lambda^*(\tau)|X]}{\mathbb{E}[N(\tau)|X]}\right) \approx \frac{\mathbb{E}[N(\tau) - \Lambda^*(\tau)|X]}{\mathbb{E}[N(\tau)|X]}.$$

Treating $c = \mathbb{E}[N(\tau)|X]$ as constant it remains to estimate $\mathbb{E}[N(\tau) - \Lambda^*(\tau)|X]$ for which we use the martingale residuals

$$\hat{M}(\tau) = N(\tau) - \int_0^\tau R(s)\exp\{\hat{\boldsymbol{\beta}}^T \mathbf{Z}\}\mathrm{d}\hat{\Lambda}_0(s)$$

resulting from the Cox model ignoring X. To do so we smooth a scatterplot of $\hat{M}_i(\tau)$ against X_i via robust locally weighted regression (see Cleveland [4]). To sum up plotting the martingale residuals against X should give an idea of the functional form of X. A linear scatterplot indicates that no further transformation of X is necessary. We have carried out several simulation studies, which supported the validity of this heuristical method.

3 The Aalen Model

3.1 The Model

An alternative to the Cox model is the additive risk model of Aalen [1]. As before let $\mathcal{T} = [0, \tau]$, $0 < \tau < \infty$ be a fixed time interval and consider an n-variate counting process $\mathbf{N}(t) = (N_i(t), i = 1, ..., n)$ together with a matrix of covariates $Y_{ij}(t), j = 1, ..., p$, $p \leq n$ observed for each component $N_i(t)$. The covariate $Y_{ij}(t)$ is set equal to 0 if the individual i is not at risk. The link between the covariates and the counting process is given by the assumption that the intensity process $\lambda_i(t)$ of $N_i(t)$ can be written as

$$\lambda_i(t) = \sum_{j=1}^{p} Y_{ij}(t)\alpha_j(t), \quad t \in \mathcal{T}$$

where $\alpha_j(t)$ are deterministic baseline intensities that are left unspecified except for some regularity conditions. In our application the functions $\alpha_j(t)$ represent the unknown, time-

dependent influences of the covariates on the cancellation of contracts.
In the next section we describe estimators for $\int_0^t \boldsymbol{\alpha}(s)\mathrm{d}s$ and $\boldsymbol{\alpha}(t)$.

3.2 Nelson-Aalen Estimator

An estimator for the integrated baseline intensity $\mathbf{B}(t) = \int_0^t \boldsymbol{\alpha}(s)\mathrm{d}s$ is given by a generalized Nelson-Aalen estimator

$$\hat{\mathbf{B}}(t) := \int_0^t \mathbf{Y}^-(s)\mathrm{d}\mathbf{N}(s)$$

where $\mathbf{Y}^-(t)$ is a generalized inverse of $\mathbf{Y}(t)$, i.e. a $p \times n$ matrix satisfying $\mathbf{Y}^-(t)\mathbf{Y}(t) = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. For simplicity, we will assume that $\mathbf{Y}(t)$ has full rank. Motivated by a least squares argument, Aalen [1] suggested to use $\mathbf{Y}^-(t) = (\mathbf{Y}^T(t)\mathbf{Y}(t))^{-1}\mathbf{Y}^T(t)$.
Usually, we are not interested in estimating $\mathbf{B}(t)$ but in estimating $\boldsymbol{\alpha}(t)$ itself. For this a kernel estimator can be used. A kernel is a measurable bounded function $K : \mathbb{R} \to \mathbb{R}_+$ which vanishes outside [-1,1] and satisfies $\int_{-1}^1 K(t)\mathrm{d}t = 1$. Let $b > 0$ and K a kernel then an estimator for $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}}(t) = \frac{1}{b}\int_{\mathcal{T}} K\left(\frac{t-s}{b}\right)\mathrm{d}\hat{\mathbf{B}}(s), \quad t \in [b, \tau - b].$$

3.3 Model Checking

Goodness of fit tests are used to examine whether a model is adequate. The following methods for testing goodness of fit in the Aalen model are based on Gandy & Jensen [9]. Assume that $\boldsymbol{c}(t) = (c_1(t), ..., c_n(t))^T$ is a vector of predictable stochastic processes such that $\boldsymbol{c}(t)$ is perpendicular to the columns of the matrix of covariates $\mathbf{Y}(t)$ in the Aalen model, i.e. $\mathbf{Y}^T(t)\mathbf{c}(t) = 0$ for all $t \geq 0$. Then under some regularity conditions

$$\hat{T}(t) := \frac{1}{\sqrt{n}}\int_0^t \boldsymbol{c}^T(s)\mathrm{d}\mathbf{N}(s)$$

is a local martingale. The process $\mathbf{c}(t)$ can be defined by a projection of some vector $\mathbf{d}(t)$ onto the orthogonal complement of the column space of $\mathbf{Y}(t)$. With the corresponding projection matrix $\mathbf{P}(t)$ we get $\mathbf{c}(t) = \mathbf{P}(t)\mathbf{d}(t)$. If $\mathbf{Y}(t)$ has full rank we can set $\mathbf{P}(t) = \mathbf{I} - \mathbf{Y}(t)(\mathbf{Y}^T(t)\mathbf{Y}(t))^{-1}\mathbf{Y}(t)$.
The tests which are constructed in Gandy & Jensen [9] are used to check the hypothesis

$H_0 :\quad \boldsymbol{\lambda}(t) = \mathbf{Y}(t)\boldsymbol{\alpha}(t)$ for some deterministic, bounded, measurable function $\boldsymbol{\alpha} : \mathcal{T} \to \mathbb{R}^p$, which means that Aalen's model is appropriate. Three different alternative hypotheses are discussed. In the first hypothesis the intensity of the alternative model is completely known. The second alternative is that there is an additional covariate. The third alternative hypothesis, the Cox model is true, is given by

$$\begin{aligned} H_c \quad : \quad & \lambda_i(t) = \lambda_0(t) R_i(t) \exp\{\boldsymbol{\beta}^{0^T} \mathbf{Z}_i(t)\},\ i = 1, ..., n \\ & \text{for some } \boldsymbol{\beta}^0 \text{ and some deterministic, bounded, measurable } \lambda_0 : \mathcal{T} \to [0, \infty] \end{aligned}$$

where $\mathbf{Z}(t)$ describes the covariates in the Cox model, which do not have to be the same as in the Aalen model (see 2.1).

Under the hypothesis H_0 and some additional assumptions $\hat{T}(t)$ converges to a mean zero Gaussian martingale whose variance can be estimated consistently by

$$[\hat{T}](t) = \frac{1}{n} \int_0^t \mathbf{d}^T(s) \mathbf{P}(s) \ \mathrm{diag}(\mathrm{d}\mathbf{N}(s)) \mathbf{P}(s) \mathbf{d}(s),$$

where $\mathrm{diag}(\mathrm{d}\mathbf{N}(s))$ is a diagonal matrix with entries $\mathrm{d}N_i(s), i \in \{1, ..., n\}$. Under regularity conditions $n^{-\frac{1}{2}} \hat{T}(t)$ converges uniformly in probability to a process $H(t)$ under the alternative hypotheses. In the first and third alternative $H(t) \geq 0$ and therefore we can use one-sided tests. The simplest test statistic we can construct is given by

$$T := \frac{1}{\sqrt{[\hat{T}](\tau)}} \hat{T}(\tau) \xrightarrow{d} \mathcal{N}(0, 1),$$

which converges as $n \to \infty$ in distribution to a standard normal random variable. To test against the hypothesis H_c, in Gandy & Jensen [9] it is suggested to choose $\mathbf{d}(t)$ as follows. Let $0 < t_0 < \tau$ and define for some $t_0 > 0$

$$d_i(s) := \begin{cases} 0 & , s \leq t_0 \\ R_i(s) \exp\{\mathbf{Z}_i(s) \hat{\boldsymbol{\beta}}(s-)\} & , t_0 < s \leq \tau, \end{cases}$$

where $\hat{\boldsymbol{\beta}}(s-) = \lim_{\substack{u \to s \\ u < s}} \hat{\boldsymbol{\beta}}(u)$. The estimator $\hat{\boldsymbol{\beta}}(u)$ is the partial maximum likelihood estimator for Cox's model which uses only information up to time u, i.e. $\hat{\boldsymbol{\beta}}(u)$ is the maximizer at (1) where τ is replaced by u.

4 Dataset

The dataset we use for our analysis emanates from a database provided by a German insurance company. It contains more than 100 000 private accident insurance contracts and for each contract several different attributes of the insurance holder and the insured person are known. Special features of these contracts are that more than one person can be insured in a contract and that the insurance holder does not have to be insured in it. There also exist aggregated covariates like the average insurance sum per insured person in the contract. Furthermore, attributes like age of the insurance holder, the annual premium, the duration of the contract, different sums insured, etc. are given. In total each contract offers about 70 attributes. For our analysis we have transformed some attributes into numerical covariates and deleted some due to too small frequency, e.g. the covariate which is 1 if the premium is paid in advance appeared only once. The cancellation of a contract could only be observed during the period of May 1st, 2002 to April 30th, 2003. About 91 percent of all contracts were censored meaning that they were not canceled during this period. Since this dataset is quite big, we reduced our analysis to a smaller dataset. There we only contemplate contracts belonging to insurance holders working in similar professions which were 31298 contracts. In our analysis we focused on 43 covariates since some of the covariates were redundant. In the smaller dataset in the first, second, third and ninth year no contracts have been canceled, which one must have in mind using the two survival models. Furthermore there are only few contracts which have a duration longer than 30 years. The longest duration of a contract is given by 44 years.

5 Results

The conclusions we want to present are drawn from the smaller dataset containing 43 covariates. First we analyzed our dataset by using two different variable selection methods to exclude the least significant covariates. Here we confine ourselves to the forward selection method since the backward selection method has produced similar results. Conducting the forward selection method in the Cox model we first estimate parameters for covariates forced into the model (see Krall et al. [11]). Then we compute adjusted χ^2-statistics for each covariate and examine the largest of these statistics. If it is significant at a 5 percent level the corresponding covariate is added to the model and stays in the model in all the following steps. In the Aalen model we include as a first covariate in the forward selection the baseline covariate, i.e. the covariate which is 1 for all contracts under risk. Then we test the hypothesis H_0 as described in Subsection 3.3 against the hypothesis that there exists an additional covariate, i.e. we test against all other variables and include the covariate having the smallest p-value into the model. We stop our selection when the p-values of all covariates in the test are greater than the level of 5 percent. The analysis of our dataset yields nearly the same significant covariates by using the forward selection method in both models. In the Cox model as well in the Aalen model the forward selection methods suggest to include 17 covariates in the models, see Table 1 for the covariates in the Cox model.

Covariates	Description	Parameter Estimate	Standard Error
beitrag_zw_2	paying the premium every 6 months	0.23833	0.05627
beitrag_zw_12	paying the premium every month	0.19082	0.04340
iart_1	paying the annual premium by direct debit	-0.43466	0.05085
taetig_art_2	employee	-0.10646	0.04210
taetig_art_3	executive employee	-0.26966	0.11709
vp1_stat_ta_12	standardized single insurance	-0.24778	0.05623
vp1_stat_gg_6	risk group B of first ins. person	-0.18729	0.05398
vp1_vsu_inv_unf	insurance sum for disability of first ins. person	-5.193E-6	1.024E-6
vp1_risk_jbeitr	risk premium of first ins. person	0.00284	0.00036
vp1_stat_geschl_2	first insured person is female	0.08593	0.04808
alter_vn	age of the insurance holder	-0.01330	0.00173
vp_vn	insurance holder equals first person insured	-0.22325	0.05460
vsu_rent_mean	average accident benefits per person insured	-0.000365	0.000102
vsu_tg1_mean	average daily benefits per person insured	-0.03044	0.00657
vsu_ktg_mean	average hospital daily benefits per person insured	-0.00551	0.00156
dyn	dynamic in the contract	0.21613	0.04258
anz_er_2	number of adults insured	0.26230	0.05245

Table 1: Parameter Estimation after Forward Selection in the Cox Model

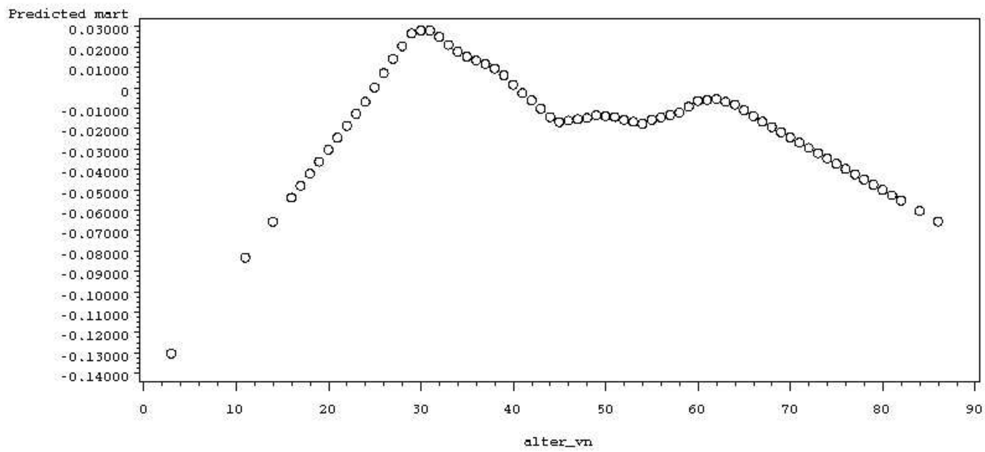

Figure 1: Martingale Residual plot of the covariate age

Variables like the risk premium, the insurance sums, paying with direct debit show effects as one would have expected. For example, a higher risk premium leads to an increasing churn rate. It is apparent that the effect of all different kinds of sums insured is the same. The intensity of a contract being canceled declines as the insurance sum grows. Furthermore, insurance holders paying with direct debit are less likely to cancel their contracts. A closer look at the martingale residuals, following the procedure described in Section 2.3, reveals that we obtain a nearly linear smoothed scatterplot for all investigated covariates that are not 0-1 variables except for the one indicating the age of the insurance holder. Recall that a linear smoothed scatterplot provides evidence that the corresponding covariate has been introduced into the model adequately. The smoothed plots of the martingale residuals against the variable age of the insurance holder (*alter_vn*) and against the variable *vsu_rent_mean* are given in the Figures 1 and 2. As Figure 1 shows the plot of the residuals of the covariate age is nonlinear. Therefore we split the covariate *alter_vn* and allow a piecewise linear influence of the age in the intervals 0-30, 30-44, 44-62, 62 and older. Excluding the variable *alter_vn* and including the 4 new variables gives a better fit. This is established by a likelihood ratio test for which the details are omitted here. The influence of the age of the insurance holder now varies with the age. The intensity of cancellation for an insurance holder of ages 0-30 and 44-62 is increasing whereas this intensity is decreasing for insurance holders of ages 30-44 and 62 and older. After fitting

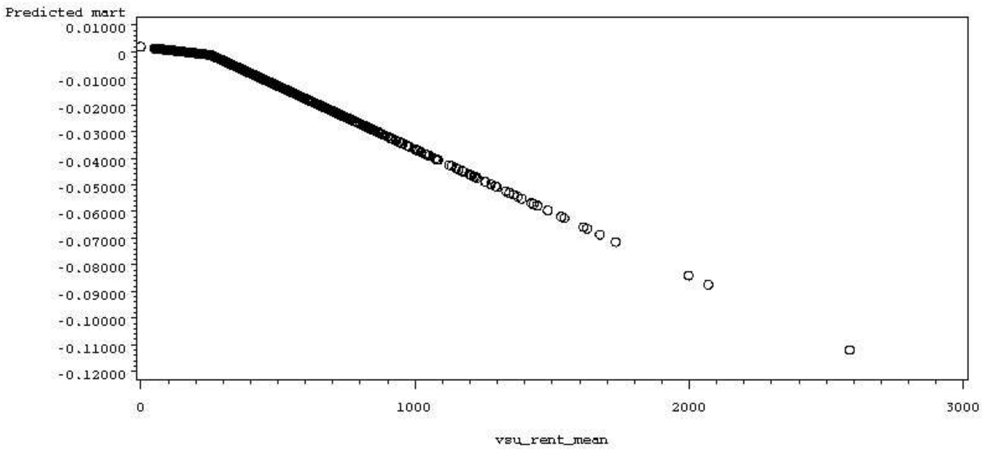

Figure 2: Martingale Residual plot of the covariate *vsu_rent_mean*

the model again with a backward and forward selection method the variable *alter_vn3* (age of an insurance holder between 44-62) does not seem to be significant and is dropped out of the model.

The Aalen model provides nearly the same trends of the variables as those indicated by the Cox model. Whenever the parameter estimate of β for a covariate is positive (negative) in the Cox model, then the estimated integrated intensity $\hat{B}(t)$ of this covariate is increasing (decreasing). This can be seen for example in the Figures 3 and 4. There the estimated integrated intensities of the covariates *iart_1* and *vsu_ktg_mean* are plotted with their pointwise confidence intervals. Testing goodness of fit of the Aalen model against the Cox model using the test explained in Subsection 3.3 with $t_0 = 5$ the hypothesis that Aalen's model is the true model is rejected at a level of 0.00019. One reason for this could be the effect arising from the occurence of several events at the same time.

To sum up the results, we can state that several covariates have been found to be of significant influence. They have been identified by using forward and backward selection methods. The influence of the covariates can be interpreted in a reasonable way. Even in the bigger dataset we are able to observe similar parameter estimates for the covariates in the Cox model. But further investigations are needed to reveal the effect of the same occurrence times of several events in both models.

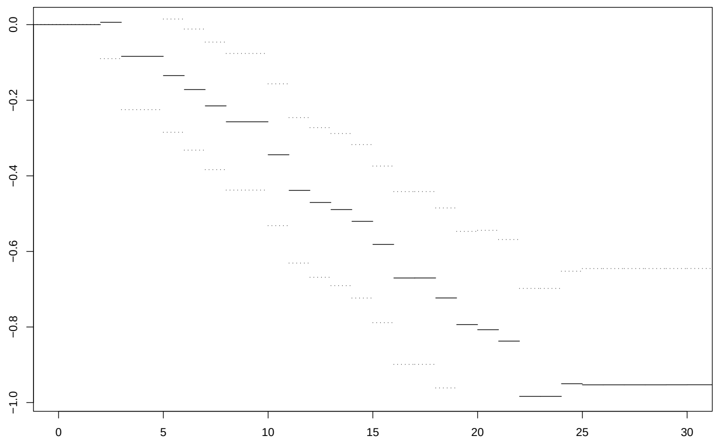

Figure 3: Estimated integrated intensity $\hat{B}(t)$ of *iart_1*

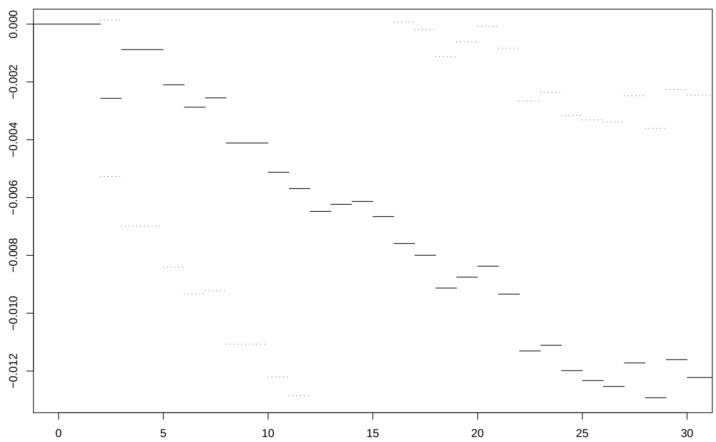

Figure 4: Estimated integrated intensity $\hat{B}(t)$ of *vsu_ktg_mean*

References

[1] O.O. Aalen. *A Model for Nonparametric Regression Analysis of Counting Processes*, volume 2 of *Lecture Notes in Statistics*, pages 1–25. Springer-Verlag, 1980.

[2] P.K. Andersen and R.D. Gill. Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4):1100–1120, 1982.

[3] P.K. Andersen, Ø.B., R.D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, 1993.

[4] W.S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *J. Am. Statist. Assoc.*, 74(368):829–836, 1979.

[5] D.R. Cox. Regression models and life tables. *Journal of the Royal Statistical Society*, Ser. B(34):187–200, 1972.

[6] C. Czado and F. Rudolph. Application of survival analysis methods to long term care insurance. *Insurance Mathematics and Economics*, 31(3):395–413, 2002.

[7] T.R. Fleming and D.P. Harrington. *Counting processes and survival analysis*. Wiley & Sons, 1991.

[8] A. Gandy and U. Jensen. A nonparametric approach to software reliability. *Appl. Stoch. Models Bus. Ind.*, 20(1):3–15, 2004.

[9] A. Gandy and U. Jensen. On goodness of fit tests for Aalen's additive risk model. Technical report, University of Hohenheim, 2004. www.uni-hohenheim.de/˜jensen/spp/preprints.html.

[10] R.D. Gill. Understanding Cox's regression model: A martingale approach. *J. Am. Statist. Assoc.*, 79:441–447, 1984.

[11] J.M. Krall, V.A. Uthoff, and J.B. Harley. A step-up procedure for selecting variables associated with survival. *Biometrics*, 31:49–57, 1975.

[12] T.H. Scheike and M.J. Zhang. An additive multiplicative Cox-Aalen regression model. *Scand. Journal Stat.*, 29:75–88, 2002.

[13] T.M. Therneau, P.M. Grambsch, and T.R. Fleming. Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160, 1990.