Aydemir, Biller:

Kernel smoothing of Aalen's linear regression model

Sonderforschungsbereich 386, Paper 101 (1997)

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

Kernel smoothing of Aalen's linear regression model

Sibel Aydemir[1], Clemens Biller[2]

Ludwig–Maximilians–Universität München,
[1]Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie (ayd@ibe.med.uni-muenchen.de),
[2]Institut für Statistik (biller@stat.uni-muenchen.de)

Summary

The linear regression model by Aalen for failure time analysis allows the inclusion of time–dependent covariates as well as the variation of covariate effects over time. For estimation Aalen considers cumulative hazard functions and derives estimates by applying counting process theory. Since often hazard functions themselves are of primary interest rather than cumulative hazard functions, in this paper we consider kernel estimation of the hazard functions, particularly in the presence of time–dependent covariates. Different kinds of bandwidths and kernel functions are discussed. A comparison of the considered methods is illustrated by data from the Stanford Heart Transplant Study.

Keywords: choice of bandwidth; survival analysis; kernel smoothing; linear hazard regression model; tail problem; time–dependent covariates; time–varying effects.

1 Introduction

In survival analysis the nonparametric linear regression model by Aalen (1989) is an alternative to Cox's proportional hazards model. The latter one is a standard tool for regression analysis of survival data, since it is easy to apply and to interpret. A drawback of the method is the assumption of proportional hazards, which is not always valid. For the situation of time–dependent covariates the so–called time–dependent Cox model is an extension of Cox's proportional hazards model, see e.g. Altman and De Stavola (1994) and Aydemir, Aydemir and Dirschedl (1996b). But a weakness both of the proportional hazards model and the extended model consists in the assumption of constant covariate effects. Unlike, Aalen's model allows both time–dependent covariates and the covariate effects to vary over time by introducing a time–varying hazard function for each covariate effect. For estimation Aalen considers cumulative hazard functions and derives estimates by applying counting process theory. Since often primary interest lies on the hazard functions themselves and not on the cumulative hazard functions, one has to look at the slope of the estimated cumulative function to get information about the covariate effects. However, for an easier interpretation of the results a direct estimation of the hazard functions would be desirable, such as the kernel estimate of the Nelson–Aalen–estimator in Ramlau-Hansen (1983) and Keiding and Andersen (1989). Huffer and McKeague (1991) suggested kernel estimates for the linear regression model of Aalen to derive weights for a weighted least squares estimate of cumulative hazard functions. For better illustration we give some results for survival times (in days) of patients with carcinoma of the oropharynx, analysed already by Aalen (1989 and 1993). Figure 1.1 (a) shows the estimated cumulative regression function $A^*(t)$ for the covariate *condition* of the patient at time of diagnosis (1=no disability, 2=restricted work, 3=requires assistance with self care, 4=confined to bed). The influence of the covariate *condition* on survival of patients has to be deduced from the slope of the function $A^*(t)$, which shows a very slight increase from the beginning to about time $t = 200$ and then a clearer increase to about $t = 400$. That means there is a positive effect of *condition* till time $t = 400$ with maximum between $t = 320$ and $t = 380$. Figure 1.1 (b) shows similar results from the direct approach with kernel estimation of the slope $\alpha^*(t)$. Here we immediately recognize a slightly positive effect at the beginning with a slight increase at about $t = 80$ and then the maximum at $t = 320$. After $t = 400$ the influence seems to disappear, since here the pointwise confidence band includes the

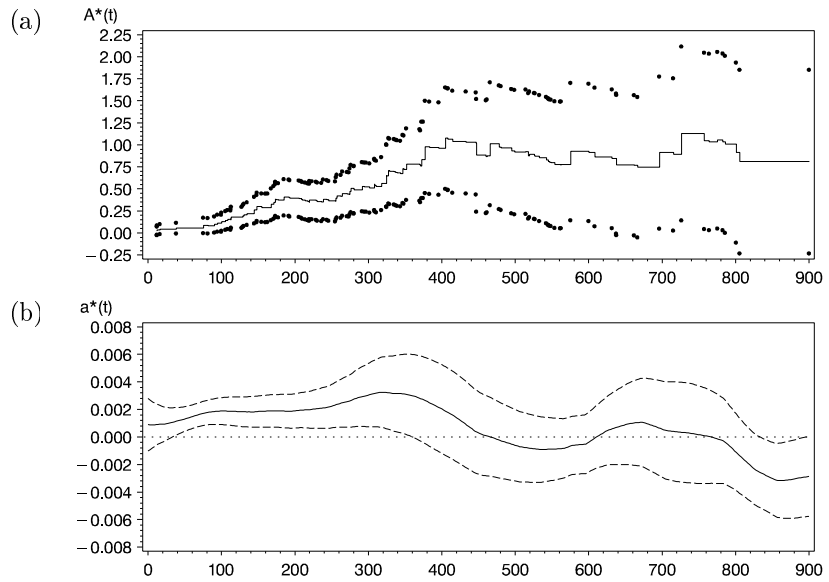

Figure 1.1: Estimated cumulative regression function (a) and kernel estimate of the slope (b), each with pointwise ±2 standard errors.

zero line.

In this paper we consider kernel estimation of the linear regression model corresponding to the suggestions of Aalen (1993), particularly in the presence of time–dependent covariates. For kernel smoothing of survival data there exist some problems. In survival analysis the density of the data often decreases as observation time increases. The use of a constant bandwidth as in Aalen (1993) therefore causes unexpected noise of the kernel estimates when data gets sparse. The kernel estimate in Figure 1.1 (b) shows such an effect from about time $t = 600$. To solve this problem we examine variable bandwidths increasing with time. Another problem of kernel smoothing in event history analysis is the so–called tail problem, which results from the fact, that the observation interval is bounded at the tails. Therefore we consider kernel functions dealing with the boundedness of the left tail.

The paper is organized as follows: In Section 2 we introduce the linear regression model following Aalen (1989). For an application to time–dependent covariates see Aydemir, Aydemir and Dirschedl (1996a). Section 3 presents the kernel estimation of the linear regression model. Here we also discuss different choices of bandwidths and the tail problem. In Section 4 we discuss and compare the considered methods using data of the Stanford Heart Transplant Study. Concluding remarks are given in Section 5.

2 The linear regression model

Interest lies in the occurrence of a certain event, for example the death of a patient, so we observe n individuals over some period of time. For some individuals the event is observed within the observation period, while for the others we only know, that the duration time to the event exceeds a certain point of time, but the exact value is not known. This feature is denoted as right censoring. For each individual additionally we observe r covariates whose values possibly vary over time.

Let $\lambda_i(t)$ denote the intensity of the occurrence of the event at time t for individual i, i.e., $\lambda_i(t)dt$ is the probability that the event occurs in interval $[t, t+dt]$ given that no event occurred before. The n–dimensional vector $\lambda(t)$ of intensities $\lambda_i(t)$, $i = 1, \ldots, n$, is modelled in dependence of the possibly time–varying covariates in the linear form

$$\lambda(t) = Y(t)\alpha(t).$$

The first component of the $(r+1)$–dimensional vector $\alpha(t) = (\alpha_0(t), \alpha_1(t), \ldots, \alpha_r(t))'$ is defined as the baseline effect, while the regression function $\alpha_j(t)$ measures the influence of covariate $j = 1, \ldots, r$. The $n \times (r+1)$ design matrix $Y(t)$ contains for each individual i an intercept and the values of the covariates measured at time t. If individual i is under risk at time t, i.e., the event has not occurred and the individual is not censored, then row i of $Y(t)$ is defined as

$$Z^i(t) = (1, Z_1^i(t), \ldots, Z_r^i(t)),$$

where $Z_j^i(t)$, $j = 1, \ldots, r$, denote the covariate values of individual i at time t. Otherwise, if individual i is not under risk at time t, row i of matrix $Y(t)$ consists only of zeros.

Two additional assumptions regarding the structure of the matrix $Y(t)$ are required. First, the sample functions are left continuous, what means that the value of $Y(t)$ is known immediately before an event time. Second, the value of $Y(t)$ depends only on the past, not on the future.

For non–parametric statistical analysis, Aalen (1989) considers cumulative regression functions $A(t) = (A_0(t), A_1(t), \ldots, A_r(t))'$, with

$$A_j(t) = \int_0^t \alpha_j(s)ds,\ j = 0, 1, \ldots, r,$$

instead of the regression functions $\alpha_j(t)$ themselves. Let $T_1 < T_2 < \ldots$ denote the ordered observed event times, i.e., the censored observations are not considered. We assume that there are no tied event times, otherwise we add a random number between 0 and 1 to each event time, see Aalen (1989), Section 6. The estimator of the cumulative regression functions is given by

$$A^*(t) = \sum_{T_k \le t} X(T_k) I_k, \tag{2.1}$$

where summation takes place over all observed event times T_k less or equal to time t. The matrix $X(t)$ is a generalized inverse of $Y(t)$, where usually the least squares inverse $X(t) = (Y(t)'Y(t))^{-1}Y(t)'$ is used. I_k is a n–dimensional vector of zeros except for a one corresponding to the individual who experiences an event at time T_k.

The estimator $A^*(t)$, only defined as long as $Y(t)$ has full rank, is asymptotically normal with covariance matrix

$$\Omega^*(t) = \mathrm{cov}(A^*(t)) = \sum_{T_k \le t} X(T_k) I_k^D X(T_k)', \tag{2.2}$$

where I_k^D is a $(n \times n)$ diagonal matrix with I_k as diagonal.

Plotting the estimated cumulative regression function $A_j^*(t)$, $j \ge 1$, against time, the slope of this function describes the influence of covariate j over time.

Within the framework of the linear regression model it is also possible to examine whether a covariate has any influence on the survival of individuals. The test can be formulated by the following null hypothesis for covariate $j \in \{1, \ldots, r\}$:

$$H_j : \alpha_j(t) = 0 \quad \text{for all } t. \tag{2.3}$$

With the diagonal weight matrix

$$K(t) = \left\{ \mathrm{diag}\left[\left(Y(t)'Y(t) \right)^{-1} \right] \right\}^{-1}$$

Aalen (1989) suggests to use the jth element U_j of the vector test statistic

$$U = \sum_{T_k} K(T_k) X(T_k) I_k \tag{2.4}$$

as a test statistic for the null hypothesis H_j, which only can be tested over the time interval where $Y(t)$ has full rank. Aalen (1989) points out, that these test statistics are only suitable for alternatives where departure of the regression function from zero is either only in the positive direction or only in the negative direction. The test statistic U is asymptotically multivariate normal distributed with estimated covariance matrix

$$V = \widehat{\mathrm{cov}}(U) = \sum_{T_k} K(T_k) X(T_k) I_k^D X(T_k)' K(T_k).$$

If only the influence of one covariate is to be tested with null hypothesis H_j, we use the test statistic

$$z = U_j V_{jj}^{-1/2}, \tag{2.5}$$

which is asymptotically standard normal distributed under the null hypothesis.

For further details on inference and goodness of fit in the linear regression model see Aalen (1989).

3 Kernel smoothing

In this section we derive a kernel estimate for $\alpha(t)$ based on the estimate $A^*(t)$ of the cumulative regression functions. Similar to Ramlau-Hansen (1983), we define the kernel estimate for the intensity vector $\alpha(t)$ as

$$\alpha^*(t) = \frac{1}{b} \int_0^\infty K\left(\frac{t-s}{b} \right) dA^*(s), \tag{3.1}$$

where the positive parameter b denotes the bandwidth and K is a kernel function which satisfies the condition $\int_{-\infty}^{\infty} K(x)dx = 1$, for example the Epanechnikov kernel

$$K(x) = 0.75(1 - x^2) \quad \text{for } |x| \le 1. \tag{3.2}$$

Since the jump times of $A^*(t)$ are the event times $T_1 < T_2 < \ldots$, the kernel estimate (3.1) may be written as

$$\alpha^*(t) = \frac{1}{b} \sum_{T_k} K\left(\frac{t - T_k}{b} \right) dA^*(T_k).$$

With

$$\begin{aligned} dA^*(T_k) &= A^*(T_k) - A^*(T_{k-1}) \\ &= \sum_{T_j \le T_k} X(T_j) I_j - \sum_{T_j \le T_{k-1}} X(T_j) I_j \\ &= X(T_k) I_k \end{aligned}$$

we get the kernel estimate

$$\alpha^*(t) = \frac{1}{b} \sum_{T_k} K\left(\frac{t - T_k}{b}\right) X(T_k) I_k . \tag{3.3}$$

An estimator for the covariance matrix is given by the diagonal of the weighted sum of the terms of the covariance matrix (2.2) of $A^*(t)$,

$$\mathrm{cov}(\alpha^*(t)) = \mathrm{diag}\left\{ \frac{1}{b^2} \sum_{T_k} K^2\left(\frac{t - T_k}{b}\right) X(T_k) I_k^D X(T_k)' \right\}.$$

This result holds, since the increments of the cumulative regression estimator are uncorrelated (a consequence of the martingale property, see Aalen, 1980 and 1993).

For kernel smoothing in event history analysis there exist two problems: the choice of the bandwidth b and the tail problem. Both are explained in detail below.

Choice of bandwidth

In survival analysis the amount of data decreases as observation time increases. Hence also the number of the data used for each time t to compute $\alpha^*(t)$ decreases as t increases, when we use a constant bandwidth b. This results in unexpected noise of the estimates $\alpha^*(t)$ at the end of the observation period. Here we propose two methods dealing with this problem.

The first method, described in Fahrmeir and Tutz (1996), Section 9.4.2, uses the size n_t of the set of individuals at risk at time t, which decreases with time. With a constant b_0 to be chosen the bandwidth is defined as

$$b_1(t) = \frac{b_0 n}{n_t}, \tag{3.4}$$

and increases as n_t decreases with time.

The second method, the kth nearest neighbour method (see Silverman, 1986), controls the degree of smoothing by the distance $d_k(t)$ of t to the k–nearest uncensored observed event time T_j. Here the bandwidth, also increasing with time, is defined as

$$b_2(t) = d_k(t), \tag{3.5}$$

where we have to choose the integer k. Here for each t the same number k of observations is used for smoothing of $\alpha(t)$.

The constant part of the bandwidth, i.e., b itself, b_0 or k (depending on the used bandwidth), still has to be chosen. There exist data driven methods with certain optimality criterions (see e.g. Keiding and Andersen, 1989), but these methods may cause oversmoothing and do not work well for each data. Therefore in the example below we choose the constants b, b_0 and k subjectively.

Tail problem

In kernel smoothing generally symmetric kernels $K(x)$ are used, that integrate to one over their support $[-1, 1]$, as the Epanechnikov kernel mentioned above. See Figure 3.1 for the shape of the Epanechnikov kernel (solid line). Due to this symmetric definition all observations s with the same absolute distance from t get the same weight $K((t-s)/b)$ in building the integral for the estimate (3.1). For $t \in [b, T_{(n)} - b]$, with $T_{(n)}$ the maximum of the observed event times T_k, integration in (3.1) takes place over all s from the interval $[t - b, t + b]$, or, considering the term $x = (t-s)/b$, over the whole support $[-1, 1]$ of the kernel $K(x)$. For $t < b$ (and similarly for $t > T_{(n)} - b$) the integral is not over the whole support $[-1, 1]$, but only over the interval $[-1, q]$, with $q = t/b < 1$. For $t < b$ (and $t > T_{(n)} - b$) the estimates $\alpha^*(t)$ therefore have less weight and are nearer to zero than for $t \in [b, T_{(n)} - b]$.

To deal with that problem, Keiding and Andersen (1989) define a smooth family of nonsymmetric kernels $K_q(x)$ with support $[-1, q]$ and use these kernels for $t < b$ instead of the symmetric kernel $K(x)$. Following Gasser and Müller (1979) $K(x)$ is multiplied by a linear function, i.e.,

$$K_q(x) = K(x)\,(\alpha_q + \beta_q x), \tag{3.6}$$

requiring that the new kernel $K_q(x)$ has integral one and mean zero over $[-1, q]$:

$$\int_{-1}^{q} K_q(x)dx = 1, \qquad \int_{-1}^{q} xK_q(x)dx = 0.$$

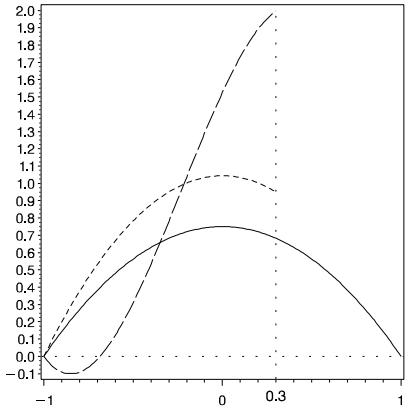

Figure 3.1: Epanechnikov kernel $K(x)$ (——), unsymmetric kernel $K_q(x)$ (– – –) and kernel $K_{I(q)}(x)$ (- - - -) for $q = 0.3$.

With the Epanechnikov kernel $K(x)$ Andersen, Borgan, Gill and Keiding (1993) derive from these equations the coefficients

$$\alpha_q = \left(\frac{2}{15} + \frac{q^3}{3} - \frac{q^5}{5}\right)\gamma_q, \qquad \beta_q = \frac{(1-q^2)^2\gamma_q}{4}$$

with

$$\gamma_q = \frac{4}{3}\left\{\left(\frac{2}{15} + \frac{q^3}{3} - \frac{q^5}{5}\right)\left(\frac{2}{3} + q - \frac{q^3}{3}\right) - \frac{1}{16}(1-q^2)^4\right\}^{-1}.$$

They point out that $K_q(-x)$ is identical to the "Optimal 1" kernel quoted by Gasser and Müller (1979). Figure 3.1 shows the unsymmetric kernel $K_q(x)$ (long dashes) in comparsion to the Epanechnikov kernel $K(x)$ (solid line) for $q = 0.3$. The third kernel function $K_{I(q)}(x)$ is defined below. From the picture we recognize that for small t (i.e., q near zero) the unsymmetric kernel $K_q(x)$ gives too much weight to observations s near zero (i.e., $x = (t-s)/b$ near q). This often leads to too big estimates of $\alpha(t)$ for small t, as we show in the example in Section 4. It also is unsatisfactory that observations s near to t have lower weights than observations which are more distant to t but which are near to 0.

For that reason for $t < b$ we intuitively define another kernel with support $[-1, q]$ and integral one over that interval. However, we do not

prove any optimality for that kernel. With

$$I(q) = \int_{-1}^{q} K(x)dx$$

we define

$$K_{I(q)}(x) = \frac{K(x)}{I(q)} \quad \text{for } x \in [-1, q] \tag{3.7}$$

and $K_{I(q)}(x) = 0$ otherwise. As we see in Figure 3.1 the kernel $K_{I(q)}(x)$ (short dashes) gives maximum weight to observations near and around t and is symmetric in the interval $[-q, q]$.

The kernels $K_q(x)$ and $K_{I(q)}(x)$ are only defined at the left tail ($t < b$). At the right tail ($t > T_{(n)} - b$) we do not consider the tail problem, since data here generally is too sparse for reasonable estimates.

4 Example: Heart Transplant Data

To discuss and compare the methods we use data of the Stanford Heart Transplant Study (Kalbfleisch and Prentice, 1980). The data includes the survival times of 103 potential heart transplant recipients. Within the observation period 69 of the patients received a new heart and 75 died. Besides the right censored survival times (in days) the following covariates were observed:

age	of the patient in years
previous *surgery*	1 = yes, 0 = no
transplant status	1 = transplanted, 0 = not transplanted
waiting time	to transplant in days
year of *acceptance*	in the study

All covariates except the transplant were observed at the baseline time $t = 0$. The transplant, which can be observed only for some patients during the observation period, is a time–varying covariate with only one possible switch of the value, and the waiting time to transplant differs from patient to patient. For transplanted patients additionally three mismatch variables and a mismatch score were observed, measuring the degree to which donor and recipient are mismatched for tissue type. In our analyses we will not take into consideration these mismatch variables, but we include the three time–constant covariates *age*, *surgery*, *acceptance* and the time–varying covariate *transplant*, as a combination of the covariates transplant status and waiting time to transplant.

The significance tests for the covariates with null hypotheses (2.3) and test statistics (2.5) yield following results:

covariate	test statistic
age	2.24
surgery	-4.26
acceptance	-1.78
transplant	-0.30

The covariate *surgery* shows the strongest (and negative) influence on the intensity. That means, a previous surgery has a positive effect on survival of patients. Also the covariate *age* has a remarkable and positive influence, whereas the effect of *acceptance* on the intensity is only modest. The results for the time–varying covariate *transplant* indicate that a transplant has no influence on the survival of patients.

Since data get sparse with increasing time, the estimated cumulative regression functions $A^*(t)$ and the kernel estimates $\alpha^*(t)$ have no interpretable effects at the end of the observation period. Therefore time axes of all figures below are shown only up to time $t = 350$. At this time the risk set of the Heart Transplant Data contains only 28 patients.

We first discuss the different kernel functions considered in Section 3 to solve the tail problem. Then we compare the different choices for the bandwidth.

Tail problem

To compare the different kernel functions (3.2), (3.6) and (3.7) Figures 4.1 (for covariate *age*) and 4.2 (for covariate *surgery*) show the cumulative regression function $A^*(t)$ following (2.1) and three kernel estimates $\alpha^*(t)$ following (3.3) using the kernel functions $K(x)$, $K_q(x)$ and $K_{I(q)}(x)$. For the kernel estimates we (subjectively) chose the constant bandwidth $b = 30$, that means, for the tail problem only the part of the graphics with time less than 30 ($t < 30$) is of interest. But for further discussions below we show the time axes up to time $t = 350$.

After small ups and downs the estimate $A^*(t)$ of the covariate *age* in Figure 4.1 (a) shows a clear increase from time $t = 60$ to $t = 110$. Then we see a very slight increase up to $t = 200$. The ups and downs afterwards are due to too less data and are therefore not interpretable. The plot of $A^*(t)$ indicates that at the beginning the effect of *age* (or the slope of $A^*(t)$) is nearly zero, while from $t = 60$ to $t = 110$ the effect increases and has its heighest value at about $t = 80$. Afterwards it gets smaller and tends to zero.

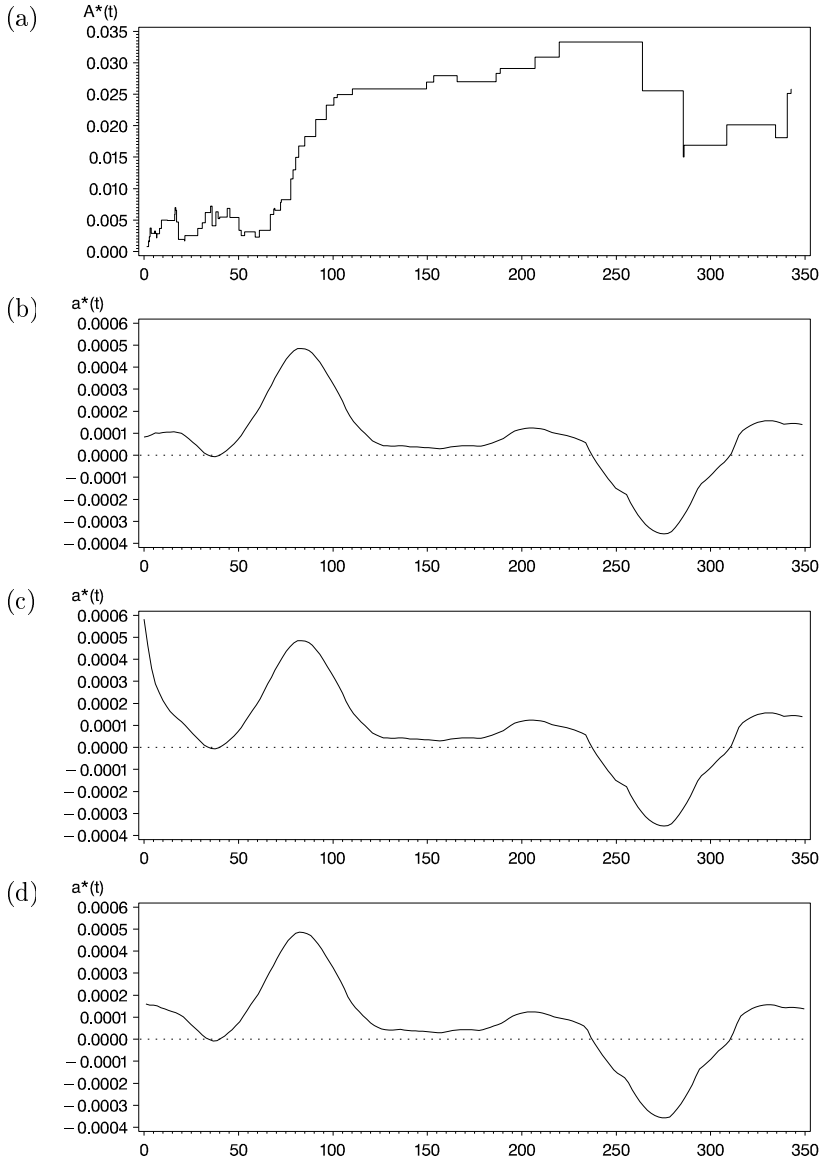

Figure 4.1: Covariate *age*: cumulative function $A^*(t)$ (a), and kernel estimates $\alpha^*(t)$ with kernel functions $K(x)$ (b), $K_q(x)$ (c), $K_{I(q)}(x)$ (d).

We recognize similar results from the kernel estimates $\alpha^*(t)$ in Figure 4.1 (b) to (d). At $t = 40$ the effect is zero, then it increases with maximum between $t = 70$ and $t = 80$, where the slope of $A^*(t)$ has it's maximum too. After that the kernel estimates decrease to zero till about $t = 180$. The ups and downs afterwards again are due to too less data.

For $t < b\,(= 30)$ the three kernel estimates differ. The estimate with kernel function $K(x)$ (Figure 4.1 (b)) is only slightly positive at the beginning. This is in agreement with the results of the estimated cumulative regression function $A^*(t)$. But this may also follow from too small weights of the kernel function $K(x)$ for values $t < b$ (see discussion of the tail problem in Section 3). Therefore let us look at the estimate using the unsymmetric kernel functions $K_q(x)$ (Figure 4.1 (c)) dealing with the tail problem. Here the kernel estimate has its maximum at time $t = 1$, indicating a strong positive effect of *age* at the beginning of the observation period, and then it decreases to zero at $t = 40$. This result is contradictory to the estimated cumulative regression function $A^*(t)$, where there is no interpretable effect at the beginning. Hence, we think that there is a strong overestimation with the unsymmetric kernel function $K_q(x)$ for t near one. Figure 4.1 (d) shows the kernel estimate using the intuitively defined kernel function $K_{I(q)}(x)$. Here at the beginning the effect $\alpha^*(t)$ is small, but bigger than the (presumably underestimated) effect using the Epanechnikov kernel $K(x)$ in Figure 4.1 (b).

In Figure 4.2 we see, that the choice of the kernel function has similar effects on the kernel estimates of the covariate *surgery*. The estimated cumulative regression function $A^*(t)$ (Figure 4.2 (a)) decreases from the beginning to about $t = 100$, with maximal decrease between $t = 70$ and $t = 80$ and a strong decrease at the beginning. From $t = 100$ on the function stays at the reached level with small ups and downs (which result from the data getting sparse). That means, from beginning till $t = 100$ there is a negative effect of *surgery*, with minima at the beginning and between $t = 70$ and $t = 80$, while there is no effect from $t = 100$ on. For $t \geq b\,(= 30)$ the kernel estimates correspond with the estimate $A^*(t)$. For $t < b$ the kernel estimates could be interpreted as follows: with the kernel $K(x)$ again there is some underestimation at the beginning (Figure 4.2 (b)), since there should be a (local) minimum; the use of the kernel $K_q(x)$ in Figure 4.2 (c) yields a strong overestimation, since such a big negative value at $t = 1$ seems not to be justified; in Figure 4.2 (d) there is presumably also an (only very small) overestimation at $t = 1$ using the kernel $K_{I(q)}(x)$, but this kernel estimate represents the results of the estimated cumulative regression function best of the three kernel estimates.

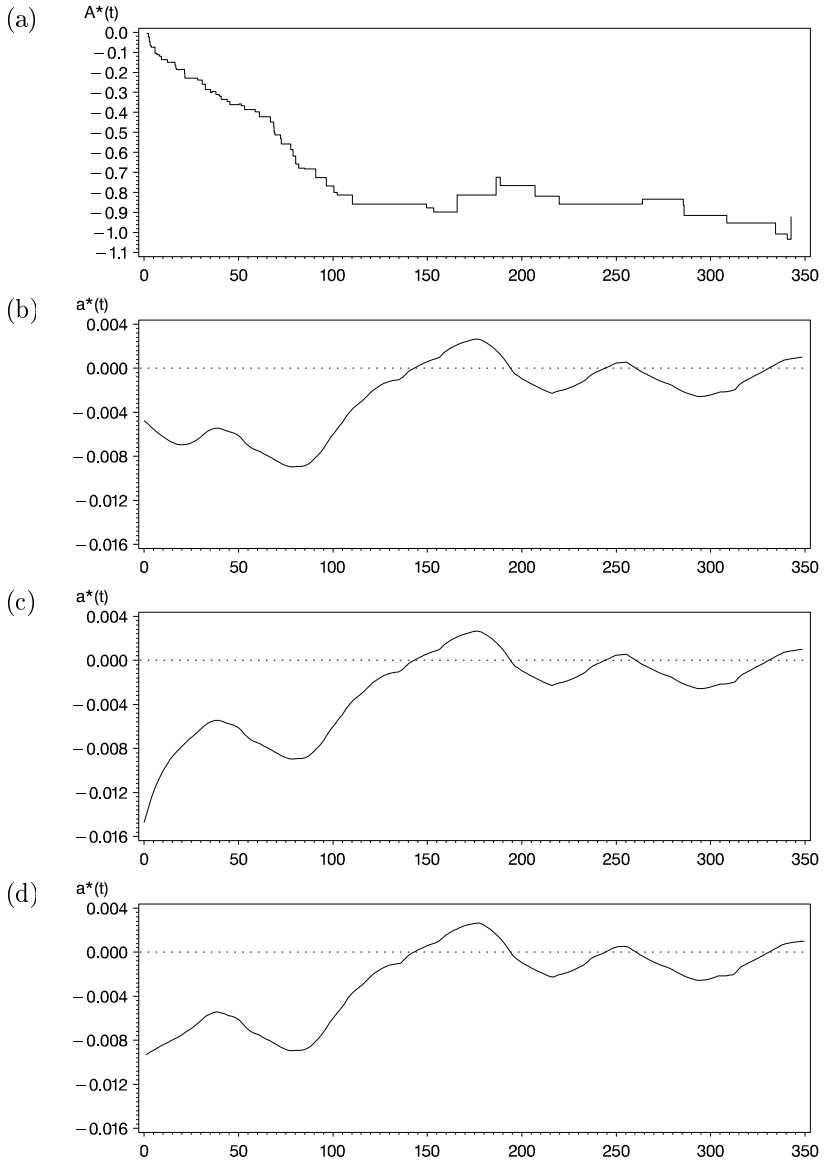

Figure 4.2: Covariate *surgery*: cumulative function $A^*(t)$ (a), and kernel estimates $\alpha^*(t)$ with kernel functions $K(x)$ (b), $K_q(x)$ (c), $K_{I(q)}(x)$ (d).

Similar interpretations about choosing the kernel function result for the kernel estimates of the covariates *acceptance* and *transplant*, so we omit them here.

As a summary we may point out the following: though the kernel function $K_q(x)$ has some theoretical optimality, for smoothing the covariate effects in this example it causes strong overfitting. In contrast to this, the use of the Epanechnikov kernel $K(x)$ results in some underestimation, but in our example it represents the data in a better way than the kernel $K_q(x)$. The best results come from the intuitively defined kernel function $K_{I(q)}(x)$. Hence, for the remainder of the example we use this kernel function for all further estimations.

Choice of bandwidth

Now we want to discuss the effect of the different bandwidths on the kernel estimates, i.e., the constant bandwidth b, the bandwidth $b_1(t)$, depending on the size of the risk set at time t (see (3.4)), and the kth nearest neighbour bandwidth $b_2(t)$ (see (3.5)). The kernel estimates using bandwidths $b_1(t)$ and $b_2(t)$ are denoted by $\alpha_1^*(t)$ and $\alpha_2^*(t)$, respectively. For $b_1(t)$ we chose the constant $b_0 = 22$ and for $b_2(t)$ the integer $k = 30$, since these choices gave (subjectively) the best results.

The kernel estimate $\alpha_1^*(t)$ with bandwidth $b_1(t)$ of covariate *age* in Figure 4.3 (a) has almost the same shape as the estimate with constant bandwidth b in Figure 4.1 (d). From $t = 200$ on there are uninterpretable ups and downs around zero when data gets sparse, too. Unlike, the kernel estimate $\alpha_2^*(t)$ of *age* using bandwidth $b_2(t)$ in Figure 4.3 (b) is very smooth and tends to zero when data gets sparse. For times $t \in [100, 200]$ the positive but decreasing estimate $\alpha_2^*(t)$ also represents the slight increase of the cumulative function A^* (Figure 4.1 (a)) in a better way than with bandwidths b and $b_1(t)$ in Figures 4.1 (d) and 4.3 (a).

Similar effects result for the kernel estimates $\alpha_1^*(t)$ and $\alpha_2^*(t)$ of covariate *surgery*. Figure 4.4 (b) shows a better smoothing by the bandwidth $b_2(t)$ in contrast to the bandwidths b (Figure 4.2 (d)) and $b_1(t)$ (Figure 4.4 (a)).

Covariates acceptance and transplant

Due to the results above we use the bandwidth $b_2(t)$ (with $k = 30$) and the kernel function $K_{I(q)}(x)$ for estimating the effects of the remaining covariates *acceptance* and *transplant*.

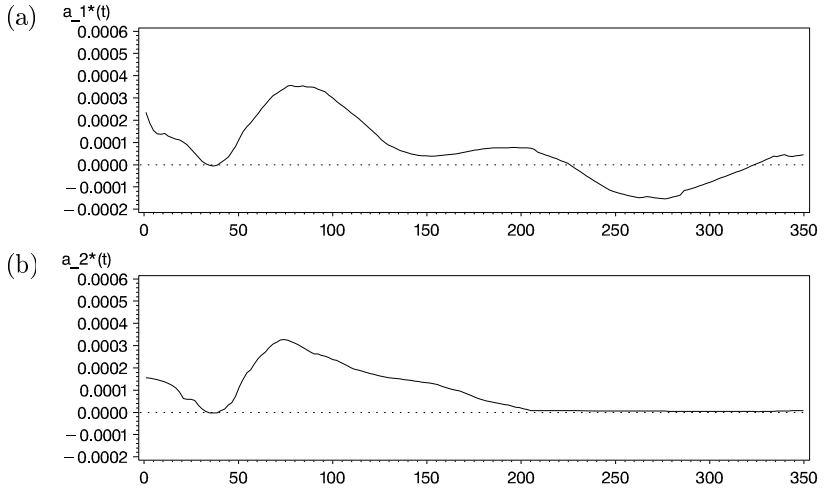

Figure 4.3: Covariate *age*: kernel estimates $\alpha_1^*(t)$ (a) and $\alpha_2^*(t)$ (b).

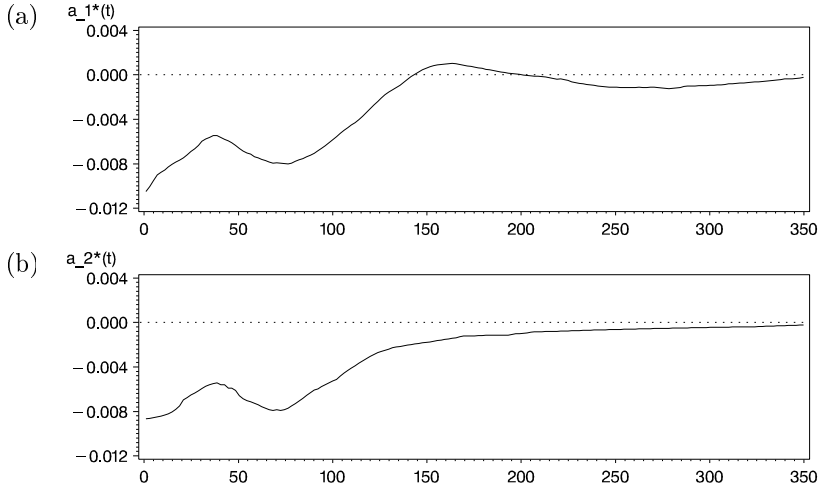

Figure 4.4: Covariate *surgery*: kernel estimates $\alpha_1^*(t)$ (a) and $\alpha_2^*(t)$ (b).

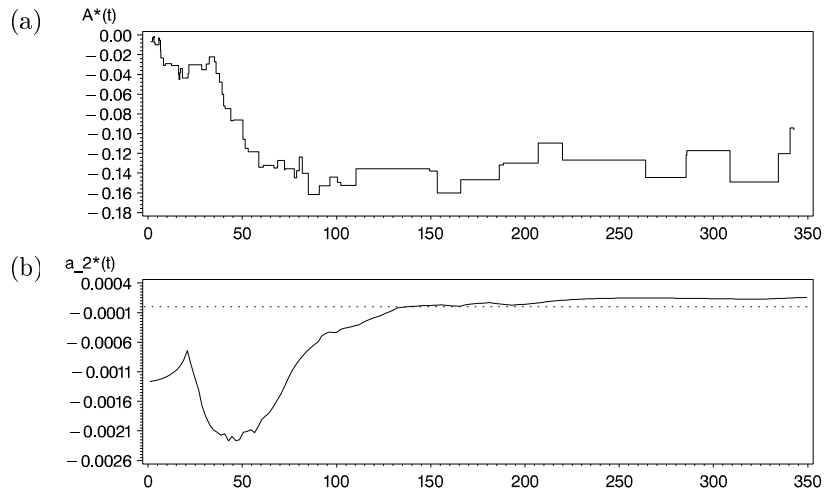

Figure 4.5: Covariate *acceptance*: cumulative function $A^*(t)$ (a) and kernel estimate $\alpha_2^*(t)$ (b).

In Figure 4.5 (a) the estimated cumulative regression function of *acceptance* decreases from the beginning to about $t = 100$ and stays at the same level with small ups and downs (which again result from data getting sparse). Part (b) of the Figure shows an equivalent result from the kernel estimate $\alpha_2^*(t)$ of the slope $\alpha(t)$. The smoothed version indicates a negative effect till time $t = 140$, which can not be recognized from the cumulative function.

The estimates of the non–significant covariate *transplant* in Figure 4.6 differ fundamentally from the estimates of the other covariates. In Figures 4.1 to 4.5 the departures of the regression functions from zero are either only in the positive direction (*age*) or only in the negative direction (*surgery* and *acceptance*). These results are in accordance to the definition of the test statistics (2.4) and (2.5) (see also Aalen, 1989, Section 3.2), that are only suitable for alternatives covering either the positive or the negative direction. However, the estimate $A^*(t)$ of *transplant* (Figure 4.6 (a)) varies around zero with no visible trend to the positive or the negative direction, that means, also the slope of $A^*(t)$ varies between positive and negative values, which is in accordance with keeping the null hypothesis (2.3) for the covariate *transplant*. Equivalent results are shown in Figure

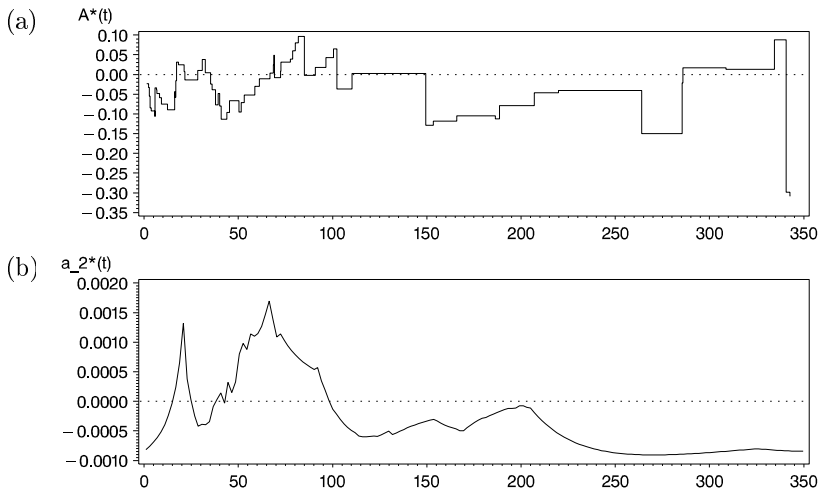

Figure 4.6: Covariate *transplant*: cumulative function $A^*(t)$ (a) and kernel estimate $\alpha_2^*(t)$ (b).

4.6 (b) by the kernel estimate $\alpha_2^*(t)$ that varies between negative and positive values.

5 Concluding remarks

The results of our example in Section 4 indicate that kernel estimates of the regression function $\alpha(t)$ correspond to the estimate of the cumulative regression function $A(t)$, if we interpret the slope of the plot of $A^*(t)$. Kernel estimation is therefore a useful alternative to the estimation of the cumulative regression functions, since viewing the development of the regression function $\alpha^*(t)$ over time is easier and more direct than using the indirect way of looking at the slope of the estimate $A^*(t)$. But for kernel smoothing in survival analysis there exist two problems, namely the choice of the bandwidth and the tail problem. In the example a better smoothing is given by the kth nearest neighbour bandwidth $b_2(t) = d_k(t)$, where we have to choose the integer k in a way that handles the trade off between smoothness and fit to the data. An oversmoothed estimate conceals the details, while the opposite yields a jagged and rough curve with a very difficult interpretation. In the example we chose the smoothing pa-

rameters subjectively, but, as pointed out above, there exists methods for an automatic choice. The unsymmetric kernel function $K_q(x)$, proposed by Keiding and Andersen (1989) to solve the tail problem, fulfils some optimality criterions, in our example, however, it causes strong overestimation for small time t. Therefore other methods should be used to deal with the tails, as we did applying the kernel function $K_{I(q)}(x)$. Hall and Wehrly (1991), for example, propose a method based on reflection of the data set at the endpoints of the design interval.

The methods presented in this paper are implemented in SAS–IML macros and are available from the authors.

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

References

Aalen, O. O. (1980). A model for non–parametric regression analysis of counting processes, *Lecture Notes in Statistics*, Vol. 2, Springer Verlag, New York, pp. 1–25.

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

Aalen, O. O. (1993). Further results on the non–parametric linear regression model in survival analysis, *Statistics in Medicine* **12**, 1569–1588.

Altman, D. G. and De Stavola, B. (1994). Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates, *Statistics in Medicine* **13**, 301–341.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer–Verlag, New York.

Aydemir, S., Aydemir, U. and Dirschedl, P. (1996a). Das lineare Regressionsmodell von Aalen zur Analyse von Überlebenszeiten unter Berücksichtigung zeitveränderlicher Kovariablen, *Discussion Paper 25*, Sonderforschungsbereich 386, Ludwig-Maximilians-Universität München.

Aydemir, S., Aydemir, U. and Dirschedl, P. (1996b). Survivalanalysen mit Berücksichtigung der zeitlichen Kovariablenentwicklung in klinischen Studien, *Discussion Paper 44*, Sonderforschungsbereich 386, Ludwig-Maximilians-Universität München.

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

Fahrmeir, L. and Tutz, G. (1996). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2. edn, Springer–Verlag, New York.

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions, *in* T. Gasser and M. Rosenblatt (eds), *Smoothing Techniques for Curve Estimation*, Vol. 757, Lecture Notes in Mathematics, Springer–Verlag, Berlin, pp. 23–68.

Hall, P. and Wehrly, T. E. (1991). A Geometrical Method for Removing Edge Effects from Kernel–Type Nonparametric Regression Estimators, *J. A. Statist. Assoc.* **86**(415), 665–672.

Härdle, W. (1991). *Smoothing techniques: with implementation in S*, Springer–Verlag, New York.

Huffer, F. W. and McKeague, I. W. (1991). Weighted Least Squares Estimation for Aalen's Additive Risk Model, *J. A. Statist. Assoc.* **86**(413), 114–129.

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

Keiding, N. and Andersen, P. K. (1989). Nonparametric Estimation of Transition Intensities and Transition Probabilities: a Case Study of a Two–state Markov Process, *Applied Statistics* **38**(2), 319–329.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions, *Ann. Statist.* **11**, 453–466.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.