Department of Applied Mathematics
Faculty of EEMCS

University of Twente
The Netherlands

P.O. Box 217
7500 AE Enschede
The Netherlands
Phone: +31-53-4893400
Fax: +31-53-4893114
Email: memo@math.utwente.nl
www.math.utwente.nl/publications

Memorandum No. 1751

Nonparametric conditional hazard rate estimation: A local linear approach

L. Spierdijk

February, 2005

ISSN 0169-2690

Nonparametric Conditional Hazard Rate Estimation: A Local Linear Approach

Laura Spierdijk*

18th February 2005

Abstract

Parametric and semiparametric methods often fail to capture the right shape of the conditional hazard rate in survival analysis. In this paper we propose a new and intuitive nonparametric estimator for the conditional hazard rate, based on local linear estimation techniques. This estimator can deal with both censored and uncensored data. We show that the local linear hazard rate estimator is consistent and asymptotically normal distributed. Moreover, we derive plug-in bandwidths based on normal and uniform reference distributions. We show that these bandwidths perform reasonably well, even when the underlying distributional assumptions are violated. We illustrate the use of the nonparametric local linear hazard rate estimator and the bandwidth selection method in several simulation experiments and in two applications to real-life data.

Keywords: conditional hazard rate estimation, local linear estimation, bandwidth selection, plug-in rates

AMS subject classification: 62N02, 62N01, 62G05,62G20

*Laura Spierdijk is affiliated to the Financial Engineering Laboratory and the Department of Applied Mathematics, University of Twente, P.O. Box 217, NL-7500 AE Enschede, The Netherlands. Phone: +31 53 489 3386. Fax: +31 53 489 3069. E-mail: L.Spierdijk@utwente.nl. The author would like to thank the participants of the statistics seminar at the University of Twente.

Introduction

Survival analysis is a widely used method in a variety of disciplines to assess the properties of durations between specific events. Important examples of durations are unemployment spells, life times, and durations between subsequent transactions in a financial security.

A useful tool in survival analysis is the so-called hazard rate, which reflects the instantaneous probability that a duration will end in the next time instant. An increasing hazard rate indicates that the probability that a spell will be completed is increasing with the duration of the event; this is called positive duration dependence. Similarly, a decreasing hazard rate reflects negative duration dependence.

Parametric, semiparametric, and nonparametric methods have been proposed to estimate hazard rates. Parametric methods impose an explicit parametric structure on the hazard rate, such as an exponential, Weibull, or lognormal distribution and have different degrees of flexibility with respect to duration dependence. For instance, the exponential distribution has a constant hazard rate, the Weibull hazard is either monotonically increasing or decreasing, and the lognormal hazard rate is non-monotonic. All parametric and semiparametric estimation techniques impose certain restrictions on the functional form of the hazard rate, which are often too restrictive. Nonparametric methods are more flexible and allow for hazard rate estimation without strong parametric assumptions. Surveys of nonparametric kernel rate estimation are provided by Singpurwalla and Wong (1983), as well as Hassani, Sarda, and Vieu (1986).

In practice, the hazard rate will often depend on certain covariates. For instance, the survival time of a patient will be affected by characteristics such as age and gender. A frequently used semiparametric method to estimate a conditional hazard rate is Cox's proportional hazards model. This model assumes that the conditional hazard rate is a multiplicative function of time (the so-called baseline hazard) and a vector of covariates. An attractive feature of this method is that can be estimated by means of Cox's partial likelihood method without specification of the baseline hazard. However, this semiparametric method imposes proportionality on the hazard rate. Unfortunately, in many cases the proportional hazards model is too restrictive. Often other semiparametric models such as the accelerated lifetime model (see Kiefer (1988)) are not flexible enough either. When parametric and semiparametric models fail, nonparametric hazard rate models are more appropriate.

The literature has paid quite some attention to nonparametric hazard rate estimation. McKeague and Utikal (1990), Li and Doss (1995), Van Keilegom and Veraverbeke (2001) and Liero (2004a, 2004b) study conditional hazard rate estimators based on kernel smoothing of a conditional Nelson-Aalen estimator for the cumulative conditional hazard function (see Nelson (1972) and Aalen (1978)), allowing for censoring. Nielsen, Linton (1995), Nielsen, Linton, and Bickel (1998), and Linton, Nielsen, and Van de Geer (2003) follow a similar approach in a multiplicative and additive hazard rate framework. Kooperberg, Stone, and Trong (1995) proceed in a different way and use a linear additive model based on linear splines and tensor products to estimate the conditional log-hazard rate.

This paper takes a different approach and introduces a new and intuitive nonparametric local linear conditional hazard rate estimator that is defined as the ratio of local linear estimators for the conditional density and survivor function. The resulting estimator is essentially a generalization of the unconditional kernel hazard rate estimator of Watson and Leadbetter (1964) and Murthy (1965). Moreover, the hazard rate estimator proposed in this paper estimator is based on local linear smoothing instead of classical Nadaraya-Watson kernel smoothing (see Nadaraya (1964) and Watson (1964)), resulting in better boundary behavior.

Surprisingly, the literature has not yet paid any attention to bandwidth selection for conditional hazard rate estimation. By contrast, various methods have been proposed for bandwidths selection in unconditional hazard rate estimation. Tanner and Wong (1984) propose maximum likelihood cross-validation for uncensored hazard rate estimation. Sarda and Vieu (1991) and Patil (1993a) extend this method to censored data. A very different approach is followed by Müller and Wang (1990, 1994), who replace the bias and variance terms in the mean squared error by estimates and minimize the resulting expression over a grid of bandwidths. Gonzàlez-Manteiga, Cao, and Marron (1996) propose a smoothed bootstrap approach for censored hazard rate estimation. With some effort these methods could be extended to the conditional case. As an alternative to this, this paper proposes plug-in bandwidths based on normal and uniform reference distributions. Rule-of-thumb reference bandwidths have proven their usefulness in kernel density estimation and we show that they are also useful in conditional kernel hazard rate estimation. They are easy to obtained and they perform reasonably well, even when the underlying distributional assumptions are violated.

The setup of this paper is as follows. Section I briefly reviews the literature on nonparametric kernel hazard rate estimation. Section II introduces the local linear conditional hazard rate estimator. The asymptotic bias and variance of the proposed local linear hazard rate estimator, as well as the optimal bandwidths that minimize the integrated mean squared error are derived in Section III. Plug-in bandwidths based on normal and uniform reference distributions are derived in Section IV. Section V extends the local linear hazard rate estimator to censored data. The results of various simulation studies are documented in Section VI. In Section VII the local linear hazard rate estimator is applied to two real-life data sets. Finally, Section VIII concludes.

I Literature review

The unconditional hazard rate is defined as the instantaneous probability that a duration Y will end in the next time instant. More precisely, the hazard rate $\lambda(\cdot)$ is defined as

$$\lambda(y) = \lim_{\Delta y \to 0} \frac{\mathbb{P}(Y \leq y + \Delta y \mid Y > y)}{\Delta y} \qquad [y > 0]. \tag{1}$$

It is not difficult to see that the hazard rate can be rewritten as the ratio of the density $f(\cdot)$ and the survivor function $S(\cdot) = 1 - F(\cdot)$ of Y; i.e.

$$\lambda(y) = \frac{f(y)}{S(y)}. \tag{2}$$

For more details we refer to Kalbfleisch and Prentice (1980), Kiefer (1988), and Lancaster (1990).

There exists an extensive literature on nonparametric hazard rate estimation. For a survey, see Singpurwalla and Wong (1983), Hassani, Sarda, and Vieu (1986), Gefeller and Michels (1992), and Padgett (1988). Roughly speaking, two different methods have been proposed to estimate the hazard rate in a nonparametric way. The first approach replaces $f(y)$ and $S(y)$ in expression (2) by estimators $\hat{f}(y)$ and $\hat{S}(y)$, resulting in the estimator

$$\hat{\lambda}(y) = \frac{\hat{f}(y)}{\hat{S}(y)}. \tag{3}$$

Watson and Leadbetter (1964) and Murthy (1965) propose the following estimators for $f(\cdot)$ and $S(\cdot)$. Assume that we observe (strictly stationary) durations $(Y_i)_{i=1}^n$. Let $w(\cdot)$ be a bounded and symmetric kernel, integrating to one and defined on the real line; e.g. the Gaussian density function. The kernel density estimator for $f(y)$ is defined as

$$\hat{f}(y) = \frac{1}{n}\sum_{i=1}^{n} w_h(y - Y_i), \tag{4}$$

with h a sequence of bandwidths such that $h \to 0$ and $nh \to \infty$ when $n \to \infty$ and

$$w_h(y) = \frac{1}{h} w\Big(\frac{y}{h}\Big). \tag{5}$$

The kernel estimator for the survivor function $S(\cdot)$ is constructed on the basis of the kernel density estimator in expression (4):

$$\begin{aligned} \hat{S}(y) = \quad = \quad & 1 - \hat{F}(y) \\ & = 1 - \int_{-\infty}^{y} \hat{f}(u)du \\ & = 1 - \frac{1}{n}\sum_{i=1}^{n} W_h(y - Y_i), \end{aligned} \tag{6}$$

where $W_h(y) = W(y/h)$ and $W(y) = \int_{-\infty}^{y} w(u)du$. Substitution of the estimators $\hat{f}(\cdot)$ and $\hat{S}(\cdot)$ in expression (3) yields a Nadaraya-Watson kernel estimator for the hazard rate. The same estimators are analyzed by Rice and Rosenblatt (1976), and Sethuraman and Singpurwalla (1981).

The second method is based on the relation between the cumulative hazard and the hazard rate, where the cumulative hazard is defined as

$$\Lambda(y) = \int_{0}^{y} \lambda(s)ds. \tag{7}$$

The relation between the cumulative hazard and the hazard rate suggest that $\lambda(\cdot)$ can be obtained by smoothing $\Lambda(\cdot)$ using a kernel; i.e.

$$\hat{\lambda}(y) = \int w_h(y-s)d\Lambda(s) \approx \frac{1}{h}\sum_{i=1}^{n} w_h(y - Y_{(i)})\frac{1}{n-i+1}, \tag{8}$$

with h a bandwidth such that $h \to 0$ as $n \to \infty$ and $Y_{(i)}$ the i-th order statistic of $(Y_i)_i^n$. This kernel estimation method was introduced by Watson and Leadbetter (1964). Interestingly, Rice and Rosenblatt (1976) compare the two classes of kernel hazard rate estimators. They show that they have the same asymptotic variance, but that their asymptotic bias is different. Ramlau-Hansen (1983), Yandell (1983), Tanner and Wong (1983), Blum and Susarla (1980), and Földes and Retjö (1981), and Lo, Mack, and Wang (1989) study similar estimators, allowing for censoring. Furthermore, Patil (1993a, 1993b), Sarda and Vieu (1991), and Patil, Wells, and Marron (1992), Gonzàlez-Manteiga, Cao, and Marron (1996) consider bandwidth selection for this type of hazard rate estimators.

Until so far, the focus has been on considered unconditional hazard rates. However, in practice, the hazard rate will often depend on certain covariates. For instance, the survival time of a patient will be affected by characteristics such as age and gender. The conditional hazard rate of Y given $X = x$ is defined as

$$\lambda(y \mid x) = \lim_{\Delta y \to 0} \frac{\mathbb{P}(Y \le y + \Delta y \mid Y > y, X = x)}{\Delta y} \qquad [y > 0]. \tag{9}$$

Now the hazard rate can be written as the ratio of the conditional density $f(\cdot \mid x)$ and survivor function $S(\cdot \mid x) = 1 - F(\cdot \mid x)$ of Y; i.e.

$$\lambda(y \mid x) = \frac{f(y \mid x)}{S(y \mid x)}. \tag{10}$$

The literature has paid quite some attention to nonparametric conditional hazard rate estimation. McKeague and Utikal (1990), Li and Doss (1995), Van Keilegom and Veraverbeke (2001) and Liero (2004a, 2004b) use martingale and counting process techniques to derive conditional hazard rate estimators based on kernel smoothing of a conditional Nelson-Aalen estimator for the cumulative conditional hazard function (see Nelson (1972) and Aalen (1978)), allowing for censoring. Their estimators are essentially conditional analogues of the unconditional hazard rate estimator given in expression (8). Nielsen and Linton (1995), Nielsen, Linton, and Bickel (1998), and Linton, Nielsen, and Van de Geer (2003) consider comparable nonparametric estimators in a multiplicative or additive hazard rate framework. Finally, a completely different method is followed by Kooperberg, Stone, and Trong (1995), who use linear splines and tensor products to estimate the conditional log-hazard rate.

II Nonparametric cnditional hazard rate estimation

The nonparametric kernel estimator for $\lambda(y \mid x)$ that we initially propose is a conditional analogue of the Nadaraya-Watson kernel hazard rate estimator of Watson and Leadbetter

(1964) and Murthy (1965). We assume that $(X_i, Y_i)_i$ is a strictly stationary process having the same marginal distribution as (X, Y), where X and Y are scalars. Extension to the case that X is d-dimensional is straightforward and therefore omitted. The kernels $k(\cdot)$ and $w(\cdot)$ are symmetric, integrate to one and are bounded with bounded support. Both b and h denote sequences of bandwidths such that $h, b \to 0$ as $n \to \infty$. Define

$$w_{h,i}(x) = \frac{1}{h} w\Big(\frac{x - X_i}{h}\Big); \tag{11}$$

$$k_b(y) = \frac{1}{b} k\Big(\frac{y}{b}\Big), \tag{12}$$

The conditional kernel density estimator is defined as

$$\hat{f}(y \mid x) = \frac{\sum_{i=1}^n w_{h,i}(x) k_b(y - Y_i)}{\sum_{i=1}^n w_{h,i}(x)}. \tag{13}$$

The estimator of the conditional survival function as the form

$$\hat{S}(y \mid x) = 1 - \hat{F}(y \mid x), \tag{14}$$

where

$$\hat{F}(y \mid x) = \frac{\sum_{i=1}^n w_{h,i}(x) I(Y_i \geq y)}{\sum_{i=1}^n w_{h,i}(x)}. \tag{15}$$

The estimator for the conditional hazard rate $\lambda(y \mid x)$ is the ratio of the estimators for the conditional density and survivor function; i.e.

$$\hat{\lambda}(y \mid x) = \frac{\hat{f}(y \mid x)}{\hat{S}(y \mid x)}. \tag{16}$$

The Nadaraya-Watson hazard rate estimator in expression (16) suffers from boundary effects near the endpoints of the support of the covariate X. To reduce this bias we propose a local linear hazard rate estimator. This estimator is based on local linear estimators for the density and survivor function. Local linear estimation is a special case of local polynomial regression and the resulting estimators have superior bias properties compared to Nadaraya-Watson estimators. For a general treatment of local polynomial estimation, see Fan, Yao, and Tong (1996).

The local linear estimator for the conditional density proposed by Fan, Yao, and Tong (1996) is the intercept from a weighted least-squares regression of $k_b(y - Y_i)$ on $(x - X_i)$ with weights $k_h(x - X_i)$ and can be written as

$$\hat{f}(y \mid x) = \frac{\sum_{i=1}^n w^*_{h,i}(x) k_b(y - Y_i)}{\sum_{i=1}^n w^*_{h,i}(x)}, \tag{17}$$

where

$$w^*_{h,i}(x) = w_{h,i}(x)(s_{n,2}(x) - (x - X_i)s_{n,1}(x)); \tag{18}$$

$$s_{n,\ell}(x) = \sum_{i=1}^n w_{h,i}(x)(x - X_i)^\ell \qquad [\ell = 1, 2].$$

An obvious local linear estimator for the conditional survival function is

$$\hat{S}(y \mid x) = \frac{\sum_{i=1}^{n} w_{h,i}^{*}(x) I(Y_i \geq y)}{\sum_{i=1}^{n} w_{h,i}^{*}(x)}. \tag{19}$$

The local linear hazard rate estimator is obtained by substituting $\hat{f}(\cdot \mid \cdot)$ and $\hat{S}(\cdot \mid \cdot)$ in formula (16). Note that the local linear hazard rate estimator is not necessarily a true hazard rate, as it may become negative. This negativity problem is inherent to local linear estimation. The kernel hazard rate estimators studied by McKeague and Utikal (1990), Li and Doss (1995), Van Keilegom and Veraverbeke (2001), and Liero (2004a, 2004b) differ from our local linear hazard rate estimator in several ways. First, they are based on kernel smoothing of the Nelson-Aalen estimator for the cumulative conditional hazard function. Moreover, the kernel hazard rate estimator of McKeague and Utikal (1990) and Li and Doss (1995) can deal with time-varying covariates, whereas our estimator is only suitable for time-independent covariates. Nevertheless, our estimator has the advantage that it is very easy to calculate, since it is simply obtained as the ratio of local linear estimators for the conditional density and survivor function.

III Asymptotic bias and variance of hazard rate estimator

Before we turn to the asymptotic bias, variance, and mean squared error of the local linear hazard rate estimator defined in Section II, we introduce the following notation. Let

$$\begin{aligned} \mu_k &= \int t^2 k(t) dt; \\ \mu_2 &= \int t^2 w(t) dt; \\ \nu_k &= \int \{k(t)\}^2 dt; \\ \nu_0 &= \int \{w(t)\}^2 dt. \end{aligned}$$

In the Appendix the asymptotic distribution of $\hat{\lambda}(y \mid x)$ is derived and it is shown that its asymptotic bias equals, for $n \to \infty, h, b \to 0$ and $nhb \to \infty$,

$$\begin{aligned} \mathbb{E}\hat{\lambda}(y \mid x) - \lambda(y \mid x) &= \frac{1}{S(y \mid x)} \Big(\frac{h^2 \mu_2}{2} \frac{\partial^2 f(y \mid x)}{\partial x^2} + \frac{b^2 \mu_k}{2} \frac{\partial^2 f(y \mid x)}{\partial y^2} \\ &\quad - \lambda(y \mid x) \frac{h^2 \mu_2}{2} \frac{\partial^2 S(y \mid x)}{\partial x^2} \Big) + o(h^2 + b^2). \end{aligned} \tag{20}$$

Moreover, in the Appendix it is also proved that the asymptotic variance is equal to

$$\mathbb{V}\text{ar}\, \hat{\lambda}(y \mid x) = \frac{v_k v_0 \lambda(y \mid x)}{nhbS(y \mid x) f(x)} + o((nhb)^{-1}). \tag{21}$$

We notice that our local linear conditional hazard rate estimator has the same asymptotic variance as the kernel estimator studied by Van Keilegom and Veraverbeke (2001) and Liero

(2004a, 2004b). The asymptotic biases of the different kernel hazard rate estimators are not the same, however.

The asymptotic mean squared error (MSE) of the local linear kernel hazard rate estimator is given by

$$\begin{aligned}\text{MSE} &= \mathbb{E}(\hat{\lambda}(y \mid x) - \lambda(y \mid x))^2 \\ &= \mathbb{V}\text{ar}\,(\hat{\lambda}(y \mid x)) + (\mathbb{E}\hat{\lambda}(y \mid x) - \lambda(y \mid x))^2 \\ &\approx \Big(\frac{1}{S(y \mid x)}\Big(\frac{h^2\mu_2}{2}\frac{\partial^2 f(y \mid x)}{\partial x^2} + \frac{b^2\mu_k}{2}\frac{\partial^2 f(y \mid x)}{\partial y^2} \\ &\quad -\lambda(y \mid x)\frac{h^2\mu_2}{2}\frac{\partial^2 S(y \mid x)}{\partial x^2}\Big)\Big)^2 + \frac{v_k v_0 \lambda(y \mid x)}{nhbf(x)S(y \mid x)}. \qquad (22)\end{aligned}$$

This results a function of the form

$$\text{MSE}(h, b) \approx (A(x, y)h^2 + B(x, y)b^2)^2 + C(x, y)/(nhb), \qquad (23)$$

where

$$\begin{aligned}A(x, y) &= \frac{1}{S(y \mid x)}\frac{\mu_2}{2}\Big(\frac{\partial^2 f(y \mid x)}{\partial x^2} - \lambda(y \mid x)\frac{\partial^2 S(y \mid x)}{\partial x^2}\Big); \\ B(x, y) &= \frac{1}{S(y \mid x)}\frac{\mu_k}{2}\frac{\partial^2 f(y \mid x)}{\partial y^2}; \\ C(x, y) &= \frac{v_k v_0 \lambda(y \mid x)}{f(x)S(y \mid x)}.\end{aligned}$$

The integrated asymptotic mean squared error (IMSE) is defined as

$$\text{IMSE} = \int\int \mathbb{E}(\hat{\lambda}(y \mid x) - \lambda(y \mid x))^2 f(x)dxdy. \qquad (24)$$

This results in a function of the form

$$\text{IMSE}(h, b) \approx Ah^4 + Bb^4 + Dh^2b^2 + C/(nhb), \qquad (25)$$

where

$$\begin{aligned}A &= \int\int \{A(x, y)\}^2 f(x)dxdy; && (26) \\ B &= \int\int \{B(x, y)\}^2 f(x)dxdy; && (27) \\ C &= \int\int C(x, y)f(x)dxdy; && (28) \\ D &= 2\int\int A(x, y)B(x, y)f(x)dxdy. && (29)\end{aligned}$$

Differentiating expression (25) and setting it equal to zero yields the system of equations

$$\begin{aligned}\frac{\partial \text{IMSE}(h, b)}{\partial h} &= 4Ah^3 + 2Dhb^2 - C/(nbh^2); && (30) \\ \frac{\partial \text{IMSE}(h, b)}{\partial b} &= 4Bb^3 + 2Dh^2b - C/(nhb^2). && (31)\end{aligned}$$

Solving this system for h and b yields as optimal bandwidths

$$b^* = \frac{(n/C)^{-1/6}}{(4A(B/A)^{5/4} + 2D(B/A)^{3/4})^{1/6}}; \tag{32}$$

$$h^* = \frac{(n/C)^{-1/6}}{(4A(B/A)^{-1/4} + 2D(B/A)^{-3/4})^{1/6}}. \tag{33}$$

Hence, the optimal bandwidths are both of order $n^{-1/6}$ and the corresponding integrated asymptotic mean squared error is of order $n^{-2/3}$, like in bivariate density estimation and in kernel hazard rate estimation based on the approach of Van Keilegom and Veraverbeke (2001).

The optimal bandwidths in expression (32) can be used to obtain rule-of-thumb reference bandwidths in the spirit of Silverman (1986), assuming that the true conditional and marginal distributions are normal or of some other parametric form. This will be discussed in more detail in the next section.

IV Plug-in bandwidths

Various methods have been proposed for bandwidths selection in unconditional kernel hazard rate estimation. Tanner and Wong (1984) propose maximum likelihood cross-validation for uncensored hazard rate estimation. Sarda and Vieu (1991) and Patil (1993a) extend this method to censored data. A very different approach is followed by Müller and Wang (1990, 1994), who replace the bias and variance terms in the mean squared error by estimates and minimize the resulting expression over a grid of bandwidths. Gonzàlez-Manteiga, Cao, and Marron (1996) propose a smoothed bootstrap approach for censored hazard rate estimation. With some effort these methods could be extended to the conditional case. As an alternative to this, this section considers plug-in bandwidths based on normal and uniform reference distributions. Rule-of-thumb reference bandwidths, see Silverman (1986) and Ruppert, Sheather, and Wand (1985), have proven their usefulness in density estimation. They are easy to obtain and they perform reasonably well, even when the underlying distributional assumptions are violated.

To derive rule-of-thumb reference bandwidths for the local linear hazard rate estimator, we have to make some parametric assumptions regarding the conditional and marginal distributions involved. As parametric framework, we take a simple accelerated lifetime model (see Kiefer (1988)) of the form

$$\log Y_i = a + bX_i + \varepsilon_i, \tag{34}$$

with coefficients a, b and $(\varepsilon_i)_i$ a sequence of iid random variables with a $\mathcal{N}(0, (c + dx)^2)$ distribution and ε_i independent of X_i. The marginal distribution of X_i is assumed to be $\mathcal{N}(\mu, \sigma^2)$ or $U(\ell, u)$ distributed. Under these assumptions the conditional density, survivor function, and hazard rate equal

$$f(y \mid x) \quad = \quad \frac{1}{y(c+dx)}\phi\Big(\frac{\log(y)-a-bx}{c+dx}\Big); \tag{35}$$

$$S(y \mid x) \quad = \quad 1-\Phi\Big(\frac{\log(y)-a-bx}{c+dx}\Big); \tag{36}$$

$$\lambda(y \mid x) \quad = \quad \frac{f(y \mid x)}{S(y \mid x)}. \tag{37}$$

To obtain plug-in bandwidths, we need to estimate the unknown coefficients a, b, c, μ, and σ from the data using maximum likelihood. Subsequently, the resulting constants A, B, C and D in expressions (26), (27), (28), and (29) have to be calculated. Unfortunately, analytical calculation of these constants seems impossible, but they can be obtained numerically using, for instance, a Gauss-Seidel quadrature. Finally, we calculate the optimal bandwidths using formulas (32) and (33).

Note that durations are always nonnegative. Therefore, we do not want the kernel hazard rate to assign positive mass to negative values. This is ensured by calculating optimal bandwidths for the hazard rate corresponding to the logarithmic durations. Subsequently, we transform the hazard rate of the logarithmic durations back to the ordinary hazard rate by means of the formula

$$\lambda_{Y|X}(y \mid x) = \frac{1}{y}\lambda_{\log Y|X}(\log y \mid x). \tag{38}$$

Clearly, there are several other methods to deal with the bounded domain problem. We mention reflection and boundary kernels (see Fan and Yao (2003)). We leave the implementation of these methods as a topic for further research.

V Extension to censoring

Until now, this paper has only considerd kernel hazard rate estimation in the absence of censoring. Clearly, in many situations data are randomly censored. Among others, Yandell (1983), Tanner and Wong (1983), and Blum and Susarla (1980) discuss unconditional kernel hazard rate estimation with censored data. This section extends their approach to the conditional case.

Due to censoring the durations of interest $(Y_i)_i$ are not completely observed. We assume a random-right censoring model with censoring times $(C_i)_i$. As in the uncensored case, $(X_i)_i$ are the covariates. The observed data are the triples $(T_i, X_i, \delta_i)_i$, where $T_i = \min(Y_i, C_i)$ and $\delta_i = I(Y_i \leq C_i)$. We assume that $(Y_i, X_i, C_i, \delta_i)$ is a strictly stationary process having the same marginal distributions as X, Y, C, and δ, where X, Y and C are scalars. Moreover, we make an assumption that is common in the censored data literature and assume that the random variables $Y \mid X$ and $C \mid X$ are independent, with conditional survivor functions $S(\cdot \mid x)$ and $G(\cdot \mid x)$ and conditional densities $f(\cdot \mid x)$ and $g(\cdot \mid x)$. Let $H(\cdot \mid x)$ denote the conditional survivor function of T and let $h(\cdot, \cdot \mid x)$ denote the conditional density of (T, δ) and write $r(y \mid x) = h(y, 1 \mid x)$. For $y > 0$, the conditional hazard rate in the presence of censoring is defined as

$$\lambda(y \mid x) \quad = \quad \lim_{\Delta y \to 0} \frac{\mathbb{P}(Y \leq y + \Delta y; \mid Y > y, X = x)}{\Delta y}. \tag{39}$$

From the independence of Y and C it follows that

$$H(\cdot \mid x) = S(\cdot \mid x) G(\cdot \mid x). \tag{40}$$

As a consequence,

$$r(\cdot \mid x) = f(\cdot \mid x) G(\cdot \mid x). \tag{41}$$

By combining (40) and (41) it is readily seen that definition (39) is equivalent to

$$\lambda(y \mid x) \quad = \quad \lim_{\Delta y \to 0} \frac{\mathbb{P}(T \leq y + \Delta y; \delta_i = 1 \mid T > y, X = x)}{\Delta y}. \tag{42}$$

Note that this definition is in terms of the observed data, which is a more convenient formulation for estimation purposes. The local linear hazard rate estimator in the presence of censoring is directly based on definition (42) and is naturally defined as

$$\hat{\lambda}(y \mid x) = \frac{\hat{r}(y \mid x)}{\hat{H}(y \mid x)}. \tag{43}$$

In definition (43) the kernel estimator for the survivor function $\hat{H}(y \mid x)$ is taken to be

$$\hat{H}(y \mid x) = \frac{\sum_{i=1}^{n} w_{h,i}^{*}(x) I(T_i \geq y)}{\sum_{i=1}^{n} w_{h,i}^{*}(x)}, \tag{44}$$

where

$$w_{h,i}^{*}(x) \quad = \quad w_{h,i}(x)(s_{n,2}(x) - (x - X_i) s_{n,1}(x)); \tag{45}$$

$$w_{h,i}(x) \quad = \quad \frac{1}{h} w\Big(\frac{x - X_i}{h}\Big); \tag{46}$$

$$s_{n,\ell}(x) \quad = \quad \sum_{i=1}^{n} w_{h,i}(x)(x - X_i)^{\ell} \qquad [\ell = 1, 2]. \tag{47}$$

To account for censoring we take as kernel density estimator

$$\hat{r}(y \mid x) = \frac{\sum_{i=1}^{n} w_{h,i}^{*}(x) k_b(y - T_i) \delta_i}{\sum_{i=1}^{n} w_{h,i}^{*}(x)}. \tag{48}$$

In the presence of censoring, we can derive plug-in bandwidths following the approach outlined in Section IV. We make the same distributional assumptions regarding $f(\cdot \mid \cdot)$ and $S(\cdot \mid \cdot)$ as before (see expressions (35) and (36)). We also have to make certain assumptions on the distribution of the censoring times. We simply assume that they follow a normal distribution with mean μ_c and variance σ_c^2; i.e.

$$g(y) = \psi\Big(\frac{y-\mu_c}{\sigma_c}\Big), \quad G(y) = 1 - \Psi\Big(\frac{y-\mu_c}{\sigma_c}\Big). \tag{49}$$

Note that the parameters of all distributions under consideration are easily obtained using (censored) maximum likelihood. Theorem A.5 leads optimal bandwidths that have the same form as without censoring, see equations (32) and (33). Only the functions $A(x,y), B(x,y), C(x,y)$ and $D(x,y)$ are different than in the uncensored case. We now find

$$\begin{aligned}
A(x,y) &= \frac{1}{H(y \mid x)}\frac{\mu_2}{2}\Big(C_1(x,y) - \lambda(y \mid x)C_3(x,y)\Big);\\
B(x,y) &= \frac{1}{H(y \mid x)}\frac{\mu_k}{2}C_2(x,y);\\
C(x,y) &= \frac{v_k v_0 \lambda(y \mid x)}{f(x)H(y \mid x)},
\end{aligned}$$

with $C_i(x,y)$ $(i = 1,2,3)$ as defined in expressions (A.40), (A.41), and (A.45).

VI Simulation studies

This section assesses the performance of the local linear hazard rate estimator and the plug-in bandwidths by means of several simulation studies.

We proceed as follows. We simulate m samples, each consisting of n pairs (X_i, Y_i). For each sample, we calculate the plug-in bandwidths based on normal and uniform reference distributions as outlined in Section II and Section V. Subsequently, the conditional hazard rate is calculated using the plug-in bandwidths. We obtain the constants A, B, C and D by numerical integration based on a Gauss-Seidel quadrature with 32 points.

The unknown parameters of the reference distribution are estimated using maximum likelihood. In case of a normal reference distribution, the integration interval in expressions (26), (27), (28), and (29) is taken of the form $I = (\mu - 2\sigma, \mu + 2\sigma) \times (-\infty, +\infty)$, covering about 95% of the data. [1]We take the integration interval of this particular form, since the bandwidths tend to infinity when x is also integrated over the interval $(-\infty, \infty)$. With the uniform reference distribution $U(\ell, u)$ the integration domain has the form $I = (\ell, u) \times (-\infty, +\infty)$.

For numerical examples, the error between the true and the estimated conditional hazard rate is measured by means of the average squared error, defined as

$$I(\mathbf{X}, \mathbf{Y}; \mathbf{x}, \mathbf{y}, h, b) = \frac{\sum_{k=1}^{N}\sum_{\ell=1}^{M} I(\hat{S}(y_k \mid x_\ell) > 0.05)\Big(\lambda(y_k \mid x_\ell) - \hat{\lambda}(y_k \mid x_\ell)\Big)^2}{\sum_{k=1}^{N}\sum_{\ell=1}^{M} I(\hat{S}(y_k \mid x_\ell) > 0.05)}. \tag{50}$$

[1]We have also tried the interval $I = (\mu - 3\sigma, \mu + 3\sigma) \times (-\infty, +\infty)$, covering about 97.5% of the data, but this resulted in very variable bandwidths.

Here $\mathbf{X} = \{X_1, \ldots, X_n\}$ and $\mathbf{Y} = \{Y_1, \ldots, Y_n\}$ is an iid sample with conditional hazard rate $\lambda(\cdot \mid \cdot)$ and $\mathbf{x} = \{x_1, \ldots, x_M\}$ and $\mathbf{y} = \{y_1, \ldots, y_N\}$ are equally spaced values over the sample spaces of X and Y, respectively. Furthermore, $\hat{\lambda}(\cdot \mid \cdot)$ is the local local linear hazard rate estimator as defined in Section II. We only evaluate the local linear hazard rate estimator at points (x, y) where it shows stable behavior. Therefore, we focus on points (x, y) such that $\hat{S}(y \mid x) > 0.05$. Note that this only excludes about 5% of the data; mainly points at the right tail of Y. This also excludes points at which $S(x, y)$ is negative, which may occur since we use local linear estimation. The average squared error is averaged across samples to obtain the mean average squared error

$$\text{MASE}(\mathbf{x}, \mathbf{y}, h, b) = \frac{1}{m} \sum_{i=1}^{m} I(\mathbf{X}^{(\mathbf{i})}, \mathbf{Y}^{(\mathbf{i})}; \mathbf{x}, \mathbf{y}, h, b). \tag{51}$$

We take the Gaussian kernel in all simulations. Although local polynomial estimators with Epanechnikov weights are asymptotically minimax integrated mean squared error efficient, we prefer Gaussian kernels since these lead to finite unconditional variance (see Seifert and Gasser (1996)). Clearly, we could also use local polynomial ridge regression with Epanechnikov kernels to ensure finite unconditional variance, but this is beyond the scope of our paper.

Example VI.1 As a first simulation experiment, we simulate $m = 100$ samples of each $n = 100$ observations from an accelerated lifetime model of the form

$$\log Y_i = X_i + \varepsilon_i, \tag{52}$$

with $(X_i)_i$ iid with $X_i \sim U(0, 2)$, $(\varepsilon_i)_i$ iid with $\varepsilon_i \sim \mathcal{N}(0, 1)$ and ε_i independent of X_i. Figures 1, 2, and 3 display boxplots of the estimated plug-in bandwidths and corresponding mean squared average errors under different distributional assumptions. The optimal bandwidths are also displayed. The mean average squared errors are calculated with $M = 10$ over the interval $(0, 2)$ and with $N = 40$ over the interval $(0, 20]$. Figure 4 shows the estimate of the conditional hazard rate based on one simulated sample, together with the true hazard rate.

Example VI.2 In this second simulation experiment we consider an accelerated lifetime model with nonnormal disturbances and a nonlinear relation between the durations and the covariate. We simulate $m = 100$ samples of each $n = 100$ observations from an accelerated lifetime model of the form

$$\log Y_i = |X_i| + \varepsilon_i, \tag{53}$$

with $(X_i)_i$ iid with $X_i \sim \mathcal{N}(0, 1)$, $(\varepsilon_i)_i$ iid with a logistic distribution[2] (with mean zero and variance $11\pi^2/30$) and ε_i independent of X_i. The shape of the hazard rate corresponding to the chosen logistic distribution is monotonically decreasing. Figures 5, 6, and 7 display boxplots of the estimated plug-in bandwidths and corresponding mean average squared errors under

[2] We simulated from a logistic distribution using the inversion method.

various distributional assumptions. The mean average squared errors are calculated with $M = 10$ over the interval $(0, 2)$ and with $N = 60$ over the interval $(0, 110]$. Figure 8 shows the estimate of the conditional hazard rate based on one simulated sample, together with the true hazard rate.

Example VI.3 To evaluate the local linear hazard rate estimator in the presence of censoring, we extend Example VI.1 by introducing random right censoring into the model. We simulate $m = 100$ samples of each $n = 100$ observations from an accelerated lifetime model of the form

$$\log Y_i = X_i + \varepsilon_i, \tag{54}$$

with $(X_i)_i$ iid with $X_i \sim U(0, 2)$, $(\varepsilon_i)_i$ iid with $\varepsilon_i \sim \mathcal{N}(0, 1)$ and ε_i independent of X_i. The iid censoring times $C_1, \ldots, C_n$ follow a $\mathcal{N}(\mu, 0.36)$ distribution independent of $Y_1, \ldots, Y_n$. Note that the value of μ determines the degree of censoring. We choose μ in such a way that the expected censoring rate is 10%. Figures 9, 10, and 11 show boxplots of the estimated plug-in bandwidths and corresponding mean squared errors for uniform and normal reference distributions. Figure 12 displays the estimate of the conditional hazard rate based on one simulated sample, together with the true hazard rate.

Example VI.4 We also extend Example VI.2 by allowing for censoring. We simulate $m = 100$ samples of each $n = 100$ observations from an accelerated lifetime model of the form

$$\log Y_i = |X_i| + \varepsilon_i, \tag{55}$$

with $(X_i)_i$ iid with $X_i \sim \mathcal{N}(0, 1)$, $(\varepsilon_i)_i$ iid with a logistic distribution (with mean zero and variance $11\pi^2/30$) and ε_i independent of X_i. The censoring times $C_1, \ldots, C_n$ are simulated from a normal $\mathcal{N}(\mu, 0.36)$ distribution, independent of $Y_1, \ldots, Y_n$. As in Example VI.3, the value of μ determines the degree of censoring. Again we take an expected censoring rate of 10%. Figures 13, 14, and 15 display boxplots of the estimated bandwidths and corresponding mean squared errors under different distributional assumptions. Figure 16 shows the estimate of the conditional hazard rate based on one simulated sample, together with the true hazard rate.

Table I reports average bandwidths and mean average squared errors for all simulation studies. It is no surprise that, when the reference distributions coincide with the true distributions of the dependent variable and the covariates, the estimated bandwidths are close to the optimal ones. But even when the assumption of normality or uniformity is violated, the normal and uniform reference rules lead to reasonable local linear hazard rate estimates. This also holds in the presence of censoring. Hence, plug-in bandwidths can play a useful rule in kernel hazard rate estimation. As in kernel density estimation, they are quite robust.
Censoring affects the quality of the local linear hazard rate estimator. Obviously, we find higher (lower) mean squared errors for higher (lower) censoring rates.

Interestingly, Table I shows that the true optimal bandwidths h in case of censoring are very close to the optimal values of h in the absence of censoring. In contrast, the optimal bandwidths b in the presence of censoring differ substantially from the optimal bandwidths when there is no censoring.

VII Empirical applications

In this section we apply the local linear hazard rate estimator and the plug-in reference bandwidths to two real-life data sets.

Example VII.1 Old Faithful data
In this example we analyze the waiting time between the starts of consecutive eruptions and the duration of the subsequent eruption of the Old Faithful geyser in Yellowstone National Park, Wyoming. The waiting time, as well as the eruption lengths are in minutes. The Old Faithful data set contains 299 observations, measured continuously from August 1 until August 15, 1985. The geyser data are available from Splus and have been studied in detail by Azzalini and Bowman (1990). The average eruption length equals about 3.5 minutes and the average waiting time is approximately 72.3 minutes.
Bashtannyk and Hyndman (2001) apply nonparametric kernel hazard rate estimation to this data and show that the conditional density of duration given a waiting time longer than 70 minutes is bimodal, whereas it is unimodal for a waiting time shorter than 70 minutes.
Although the conditional hazard rate is equivalent to the conditional density, the first provides direct information on the duration dependence. Therefore, it is interesting to study the hazard rate as well.
To estimate the conditional hazard rate, we first calculate plug-in bandwidths based on normal and uniform reference distributions. We find that the differences between the estimated bandwidths based on normal and uniform reference distributions are small.
Figure 17 displays the estimated conditional hazard rate $y \longrightarrow \hat{\lambda}(y \mid x)$ for various values of x based on the normal reference rule with $d \neq 0$. Figure 17 shows that the duration dependence is nonmonotonic: first, the conditional hazard rate is very flat, then it increases, subsequently it decreases and it finally increases again. Stated differently, for eruption lengths up to 1.2 minutes, the probability that the eruption will end within the next few seconds is negligibly small. For eruption lengths of 1.2−2 minutes, this probability increases with the length of the eruption. For eruption lengths between 2−3 minutes the probability that the eruption will end decreases over time and for eruptions lengths longer than 3 minutes this probability increases with the length of the eruption.
Moreover, Figure 17 shows that the probability that an eruption will end increases with the waiting time between the eruptions. Roughly speaking, when the previous eruption has taken place long time ago, the probability that the consecutive eruption will end is larger than in

case of a recent eruption.

It is not obvious how to detect the aforementioned bimodality in the kernel hazard rate, since there is no clear change in the shape of the hazard rate around waiting time 70 minutes. Yet the increases and decreases in the conditional hazard rate become much steeper when the waiting time increases. Conditional on a waiting time of 70 minutes, the hazard rate already has a strong bump.

Example VII.2 Kidney transplant data

This data set contains 863 survival times (in days) of patients following a kidney transplant (see Klein and Moeschberger (2004)). The data provide information on the gender (male/female), race (black/white), and age (in years) of each patient. Moreover, it is also known whether a survival time was censored or not. Totally, about 84% of the observations is censored. Hence, the survival times are characterized by very strong censoring.

We focus on white males and estimate the local linear conditional hazard rate estimator for this group of 432 patients and a censoring rate of 83%. The survival times vary between 1 and 3,434 days (= 9.4 years). We take age as the conditioning variable. The average age of patients in the group under consideration is slightly less than 44 years.

We find only small differences between the various plug-in bandwidths. Figure 19 (based on the normal reference rule with $d = 0$) shows that the hazard rate is nonmonotonic: for short durations, say up to 50 days, it is increasing. For longer durations the hazard rate is decreasing. Hence, just after the kidney transplant, the probability that the male patient will die increases with time. When the patient has survived up to a certain turning point, the probability that he will die decreases with time. Moreover, the hazard rate is higher for older patients. This means that the survival probability is higher the younger the patient.

For the subpopulation of white females (280 observations, 86% censored) the results are slightly different. The survival times of the patients in this group vary between 1 and 3,420 days (= 9.4 years). The average age of the patients is almost 41 years. Figure 20 (based on the normal reference rule with $d = 0$) shows that the hazard rate is almost monotonically decreasing and that the impact of age is smaller than in case of male patients.

Comparison to parametric models

The motivation for using nonparametric instead of parametric is given by the inability of parametric models to capture the right hazard shape. Therefore, an important question that arises is to what extent the nonparametric local linear hazard rate estimates differ from hazard estimates based on accelerated lifetime or proportional hazard models. To answer this question, we make a comparison between parametric and nonparametric estimators. We focus on two different aspects. Does the parametric model capture well the relation between the covariates and the durations? And does the parametric model capture well the duration dependence that is present in the data?

The proportional hazards model assumes that the conditional hazard rate is a multiplicative function of time and a vector of covariates. This means that a change in the value of the covariate only leads to a vertical shift in the position of the hazard rate. From Figure 17 it is readily seen that proportionality is strongly violated. The same holds for the conditional hazard rates based on the kidney transplant data, see Figures 19 and 20.

Regarding the Old Faithful data, Figure 18 displays the local linear hazard rate estimate together with the hazard rate estimate based on an accelerated lifetime model with normally distributed errors. Furthermore, Figure 21 compares the local linear hazard rate estimate and the normal accelerated lifetime hazard estimate for the kidney transplant data on male white patients. In both cases, the differences between the parametric and nonparametric estimates are large. Clearly, the estimators based on the accelerated failure time models could be improved by taking an error term distribution that captures more accurately the duration dependence present in the data. In case of the kidney data, a log Weibull distribution would do a better job. However, in case of the geyser data the choice of the error term distribution is less obvious because of the nonstandard shape of the hazard rate.

The difficulty of finding an appropriate parametric model illustrates the power of the non-parametric approach. At the same time, we see how the nonparametric approach can play a role in selecting an appropriate (semi-) parametric model.

VIII Conclusions

Parametric and nonparametric methods often fail to capture the right shape of the conditional hazard rate in survival analysis. This paper has proposed a new and intuitive nonparametric estimator for the conditional hazard rate, based on local linear estimation techniques. This estimator is consistent and asymptotically normal distributed and has more favorable bias properties than a Nadaraya-Watson equivalent. It can deal effectively with both censored and uncensored data.

Moreover, this paper has derived plug-in bandwidths based on normal and uniform reference distributions. Rule-of-thumb reference bandwidths have proven their usefulness in kernel density estimation, even when the underlying distributional assumptions are violated. These rule-of-thumb bandwidths are easy to obtain and perform reasonably well, even when the underlying distributional assumptions are violated.

The nonparametric local linear hazard rate estimator proposed in this paper can be extended in several ways. For example, as shown by Müller and Wang (1994), the use of boundary kernels in hazard rate estimation will reduce the boundary bias in the right endpoints of the domain. Moreover, in line with Liero (2004a, 2004b) statistical tests based on the L^2 distance between a parametric and a nonparametric hazard rate estimator can be used effectively to test the null hypothesis that the conditional hazard rate has a specific parametric form.

References

Aalen, O., 1978, Nonparametric Inference for a Family of Counting Processes, *Annals of Statistics*, 4, pp. 701-726.

Azzalini, A. and A.W. Bowman, 1990, A Look at Some Data on the Old Faithful Geyser, *Applied Statistics*, 39, pp. 357-365.

Bashtannyk, D.M. and R.J. Hyndman, 2001, Bandwidth Selection for Kernel Conditional Density Estimation, *Computational Statistics & Data Analysis*, 36, pp. 279-298.

Hyndman, R.J., D.M. Bashtannyk, and G.K. Grunwald, 1996, Estimating and Visualizing Conditional Density Estimators, *Journal of Computational and Graphical Statistics*, 5, pp. 315-336.

Blum, J.R. and V. Susarla, 1980, Maximal Deviation Theory of Density and Failure Rate Function Estimates based on Censored Data. In P.R. Krishnaiah (ed), *Multivariate Analysis-V*, Amsterdam, North-Holland, pp. 213-222.

Fan, J., Q. Yao, and H. Tong, 1996, Estimation of Conditional Densities and Sensitivity Measures in Nonlinear Dynamical Systems, *Biometrika*, 83, pp. 189-206.

Fan J. and Q. Yao, 2003, Nonlinear Time Series. Nonparametric and Parametric Methods, Springer Verlag.

Földes, A. and L. Retjö, 1981, Strong Uniform Consistency for Nonparametric Survival Estimators from Randomly Censored Data, *Annals of Statistics*, 9, pp. 122-129.

Gefeller, O. and P. Michels, 1992, A Review on Smoothing Methods for the Estimation of the Hazard Rate based on Kernel Functions, in Dodge, Y. and J. Whittaker (eds), *Computational Statistics*, Physica-Verlag, Switzerland, pp. 459-464.

Gonzalez-Manteiga, W., R. Cao, and J.S. Marron, 1996, Bootstrap Selection of the Smoothing Parameter in Nonparametric Hazard Rate Estimation, *Journal of the American Statistical Association*, 91, pp. 1130-1140.

Hassani, S., P. Sarda, and P. Vieu, 1986, Approche Non-Paramétrique en Théorie de la Fiabilité: Revue Bibliographique, *Revue de Statistique Appliquée*, 35, pp. 27-41.

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

Kiefer, N.M., 1988, Economic Duration Data and Hazard Functions, *Journal of Economic Literature*, 26, pp. 646-679.

Klein, J.P. and M.L. Moeschberger, 1997, Survival Analysis: Techniques for Censored and Truncated Data, Springer Verlag.

Kooperberg, C., C.J. Stone, and Y.K. Truong, 1995, Hazard Regression, *Journal of the American Statistical Association*, 90, pp. 78-94.

Lee, M.-j., 1996, Method of Moments and Semiparametric Econometrics for Limited Dependent Variable Models, Springer.

Li, G. and H. Doss, 1995, An Approach to Nonparametric Regression for Life History Data Using Local Linear Fitting, *Annals of Statistics*, 23, pp. 787-823.

Liero, H., 2004a, Goodness of Fit Tests of L_2 Type, Institute of Mathematics and University of Potsdam, mimeo.

Liero, H., 2004b, Testing the Hazard Rate Part I, Institute of Mathematics and University of Potsdam, mimeo.

Linton, O.B., J.P. Nielsen, and S. van de Geer, 2003, Estimating Multiplicative and Additive Hazard Functions by Kernel Methods, *Annals of Statistics*, 31, pp. 464-492.

Lo, S.-H., Y.P. Mack, and J.L. Wang, 1989, Density and Hazard Rate Estimation for Censored Data via Strong Representation of the Kaplan-Meier Estimator, *Probability Theory Related Fields*, 80, pp. 461-472.

Masry, E. and J. Fan, 1997, Local Polynomial Estimation of Regression Functions for Mixing Processes, *Scandinavian Journal of Statistics*, 24, pp. 165-179.

McKeague, I.W. and K.J. Utikal, 1990, Inference for a Nonlinear Counting Process Regression Model, *Annals of Statistics*, 18, pp. 1172-1185.

Müller, H.G and J.L. Wang, 1990, Locally Adaptive Hazard Smoothing, *Probability Theory and Related Fields*, 85, pp. 523-538.

Müller, H.G and J.L. Wang, 1994, Hazard Rate Estimation Under Random Censoring with Varying Kernels and Bandwidths, *Biometrics*, 50, pp. 61-76.

Murthy, V.K., 1965, Estimation of Jumps, Reliability and Hazard Rate, *Annals of Mathematical Statistics*, 36, pp. 1032-1040.

Nadaraya, E.A., 1964, On Estimating Regressions, *Theory of Probability and its Applications*, 10, pp. 186-190.

Nelson, W., 1972, Theory and Applications of Hazard Plotting for Censored Failure Data, *Technometrics*, 14, pp. 945-965.

Nielsen, J.P., O.B. Linton, 1995, Kernel Estimation in a Nonparametric Marker Dependent Hazard Model , *Annals of Statistics*, 23, pp. 1735-1748.

Nielsen, J.P., O.B. Linton, and P.J. Bickel, 1998, On a Semiparametric Survival Model with Flexible Covariate Effect, *Annals of Statistics*, 26, pp. 215-241.

Padgett, W.J., 1988, Nonparametric Estimation of Density and Hazard Rate Functions when Samples are Censored, P.R. Krishnaiah and C.R. Rao (eds), *Handbook of Statistics*, 7, pp. 313-331. Elsevier Science Publishers.

Patil, P.N., 1993, Bandwidth Choice for Nonparametric Hazard Rate Estimation, *Journal of Statistical Planning and Inference*, 35, pp. 15-30.

Patil, P.N., 1993,b On the Least Squares Cross-Validation Bandwidth in Hazard Rate Estimation, *Annals of Statistics*, 21, pp. 1792-1810.

Patil, P., M.T. Wells, and J.S. Marron, 1994, Some Heuristics of Kernel Based Estimators of Ratio Functions, *Journal of Nonparametric Statistics*, 4, pp. 203-209.

Ramlau-Hansen, H., 1983, Smoothing Counting Process Intensities by Means of Kernel Functions, *Annals of Statistics*, 11, pp. 453-466.

Rice, J. and M. Rosenblatt, 1976, Estimation of the Log Survivor Function and Hazard Function, *Sankhya Series A*, 38, pp. 60-78.

Ruppert, D., S.J. Sheather, and M.P. Wand, 1985, An Effective Bandwidth Selector for Local Least Squares Regression, *Journal of the American Statistical Association*, 90, pp. 1257-1270.

Seifert, B. and T. Gasser, 1996, Finite-Sample Variance of Local Polynomials: Analysis and Solutions, *Journal of the American Statistical Association*, 91, pp. 267-275.

Sarda, P. and P. Vieu, 1991, Smoothing Parameter Selection in Hazard Estimation, *Statistics and Probability Letters*, 11, pp. 429-434.

Silverman, B.W., 1986, Density Estimation for Statistics and Data Analysis, Monographs of Statistics and Applied Probability 26, Chapman & Hall/CRC.

Singpurwalla, N.D. and M.-Y. Wong, 1983, Estimation of the Failure Rate − A Survey of Nonparametric Models. Part I: Non-Bayesian Methods, *Communications in Statistics − Theory and Methods*, 12, pp. 559-588.

Sethuraman, J. and N.D. Singpurwalla, 1981, Large Sample Estimates and Uniform Confidence Bounds for the Failure Rate Function based on a Naive Estimator, *Annals of Statistics*, 9, pp. 628-632.

Tanner, M.A. and W.H. Wong, 1983, The Estimation of the Hazard Function from Randomly

Censored Data by the Kernel Method, *Annals of Statistics*, 11, pp. 989-993.

Therneau, T.M. and P.M. Grambsch, 2000, Modeling Survival Data: Extending the Cox Model, Springer Verlag.

Van Keilegom, I. and N. Veraverbeke, 2001, Hazard Rate Estimation in Nonparametric Regression with Censored Data, 53, pp. 730-745.

Watson, G.S., 1964, Smooth Regression Analysis, *Sankhya Series A*, 26, pp. 359-372.

Watson, G.S. and M.R. Leadbetter, 1964, Hazard Rate Analysis. I, *Biometrika*, 51, pp. 175-184.

Yandell, B.S., 1983, Nonparametric Inferences for Rates with Censored Survival Data, *Annals of Statistics*, 11, pp. 1119-1135.

	optimal	uniform $d \neq 0$	uniform $d = 0$	normal $d \neq 0$	normal $d = 0$
normal ALM					
h	0.54	0.56	0.57	0.57	0.60
b	0.45	0.46	0.46	0.53	0.51
MASE	6.83E-03	7.19E-03	6.63E-03	6.14E-03	7.11E-03
logistic ALM					
h	0.93	1.56	1.63	1.32	1.29
b	0.79	0.87	0.91	0.91	0.95
MASE	1.18E-03	1.09E-03	1.03E-03	9.39E-04	9.94E-04
censored normal ALM					
h	0.55	0.58	0.63	0.63	0.64
b	0.39	0.39	0.38	0.44	0.42
MASE	9.27E-03	9.17E-03	9.55E-03	8.57E-03	7.70E-03
censored logistic ALM					
h	0.93	2.02	2.58	1.94	2.17
b	0.46	0.47	0.45	0.45	0.44
MASE	5.73E-03	6.27E-03	5.69E-03	5.23E-03	4.63E-03

Table I: Optimal and estimated bandwidths and corresponding mean average squared errors

This table applies to the kernel estimators of the hazard rate corresponding to several accelerated lifetime models (ALM). It reports optimal and estimated bandwidths, as well as mean average squared errors. The bandwidths are the averages over 100 samples. The mean average squared errors are averages of the average squared errors over 100 samples.

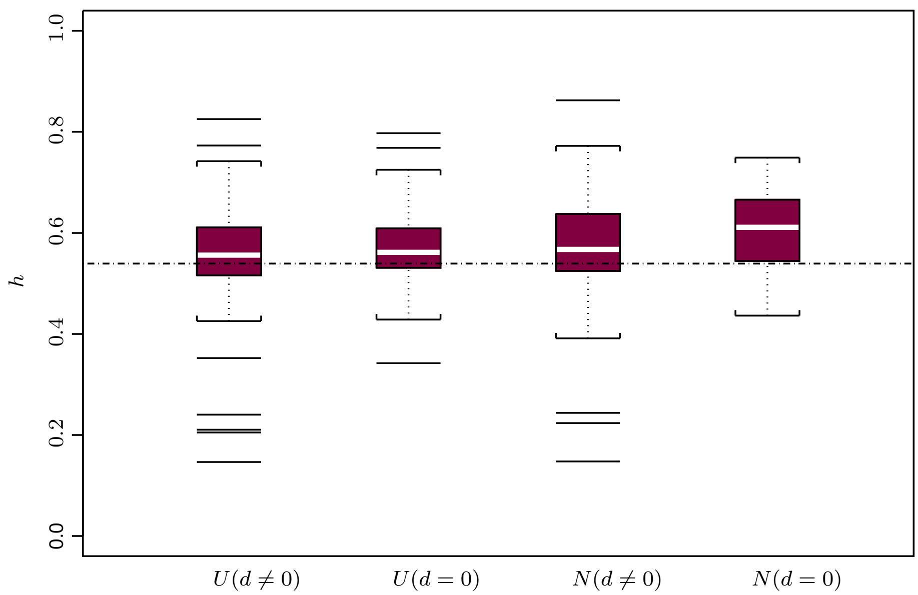

Figure 1: Values of h for each of the 100 samples from a normal accelerated lifetime model

These boxplots show the estimated values of h based on different reference distributions. The dashed line shows the optimal value of h.

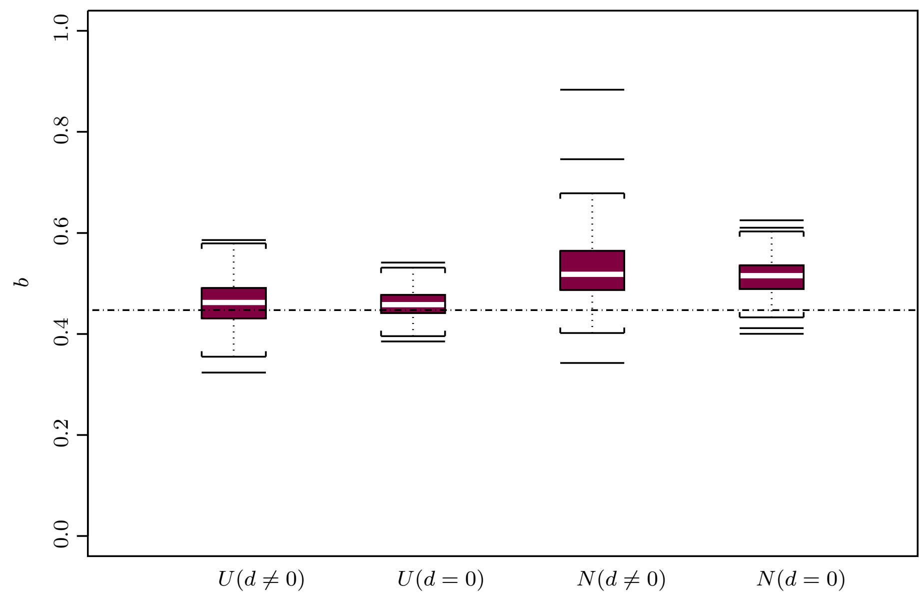

Figure 2: Values of b for each of the 100 samples from a normal accelerated lifetime model

These boxplots show the estimated values of h based on different reference distributions. The dashed line shows the optimal value of b.

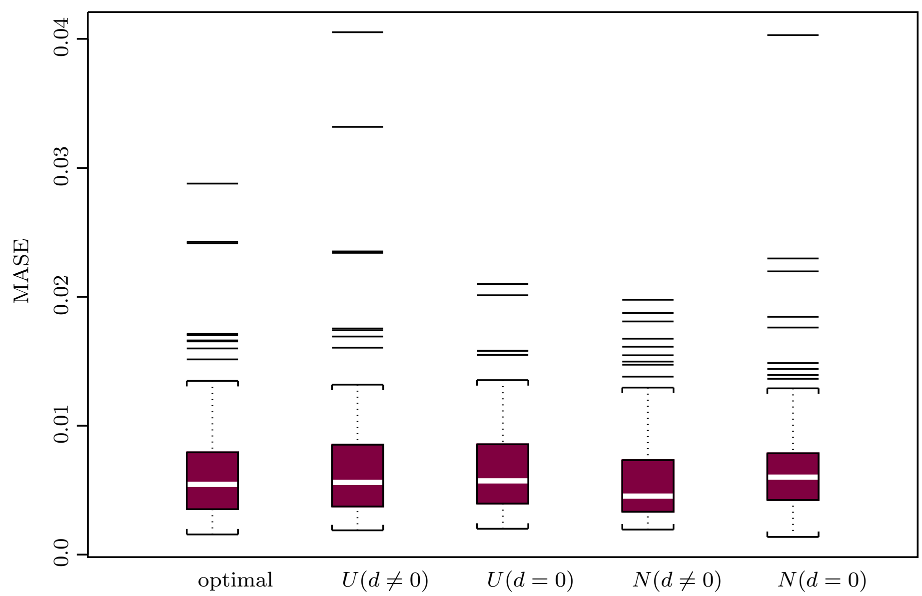

Figure 3: Estimated MASE values for each of the 100 samples from a normal accelerated lifetime model

These boxplots display the estimated MASE values based on different reference distributions. The first boxplot shows the MASE value for 100 samples based on the optimal bandwidths h and b.

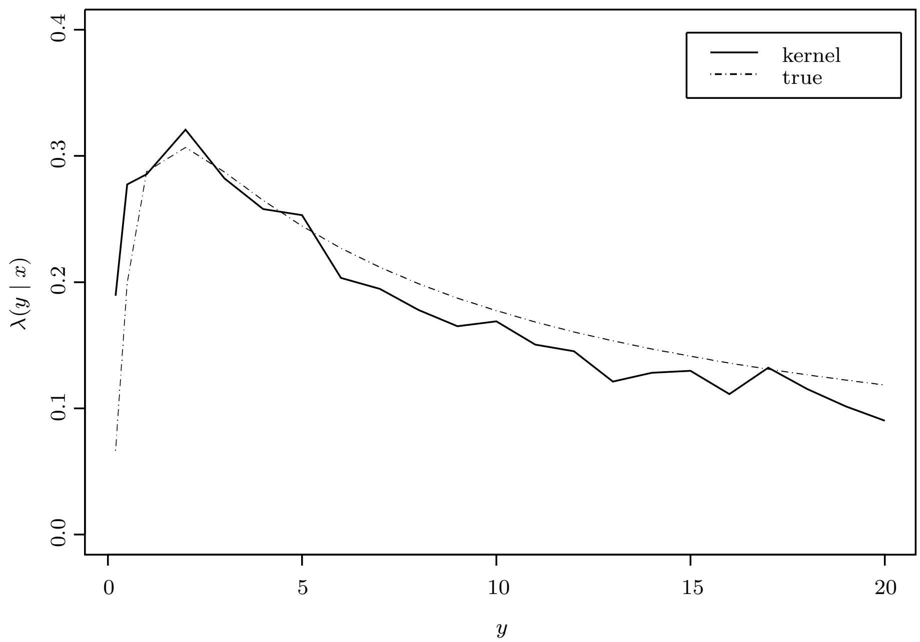

Figure 4: Kernel hazard rate estimate and true hazard rate

This plot shows the local linear hazard rate estimate $y \longrightarrow \hat{\lambda}(y \mid x)$ for $x = 1$, based on the uniform reference rule (with $d = 0$) for one simulated sample from the normal accelerated lifetime model. The dashed line represents the true hazard rate.

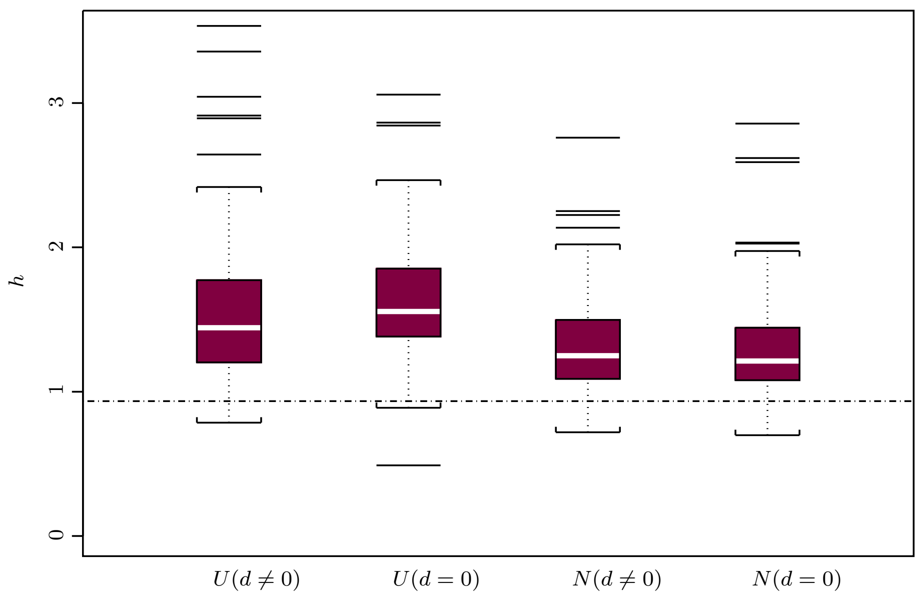

Figure 5: Values of h for each of the 100 samples from a logistic accelerated lifetime model

These boxplots show the estimated values of h based on different reference distributions. The dashed line shows the optimal value of h.

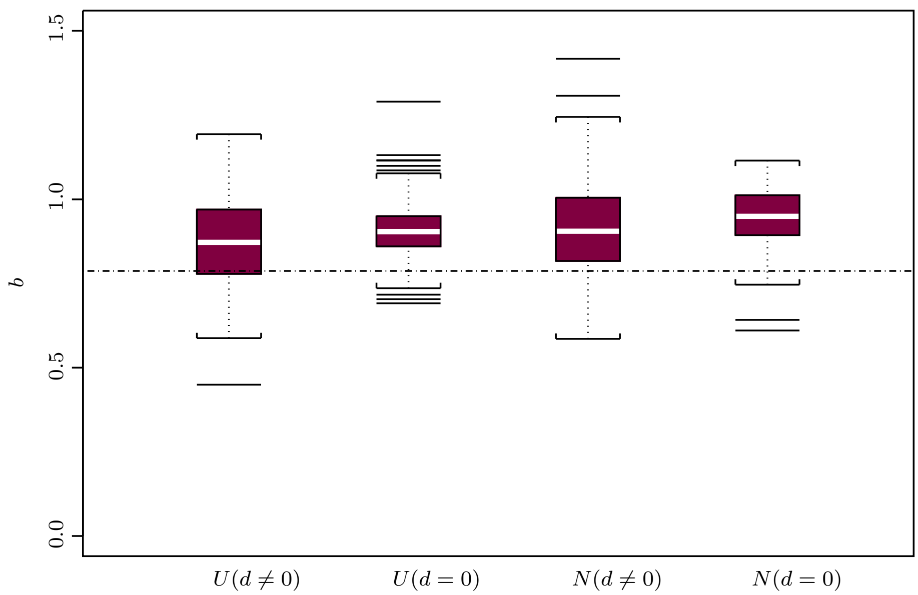

Figure 6: Values of b for each of the 100 samples from a logistic accelerated lifetime model

These boxplots show the estimated values of h based on different reference distributions. The dashed line shows the optimal value of b.

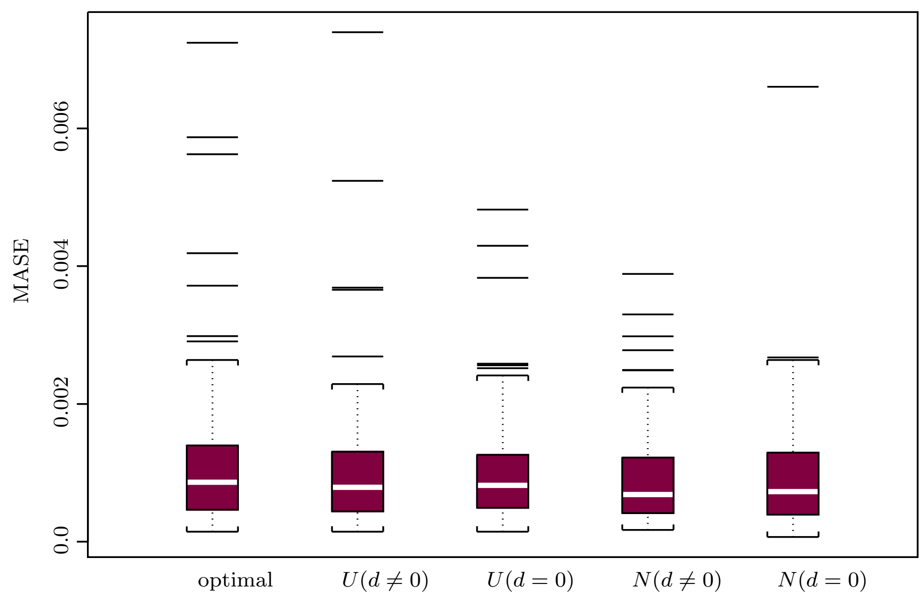

Figure 7: Estimated MASE values for each of the 100 samples from a logistic accelerated lifetime model

These boxplots display the estimated MASE values based on different reference distributions. The first boxplot shows the MASE value for 100 samples based on the optimal bandwidths h and b.

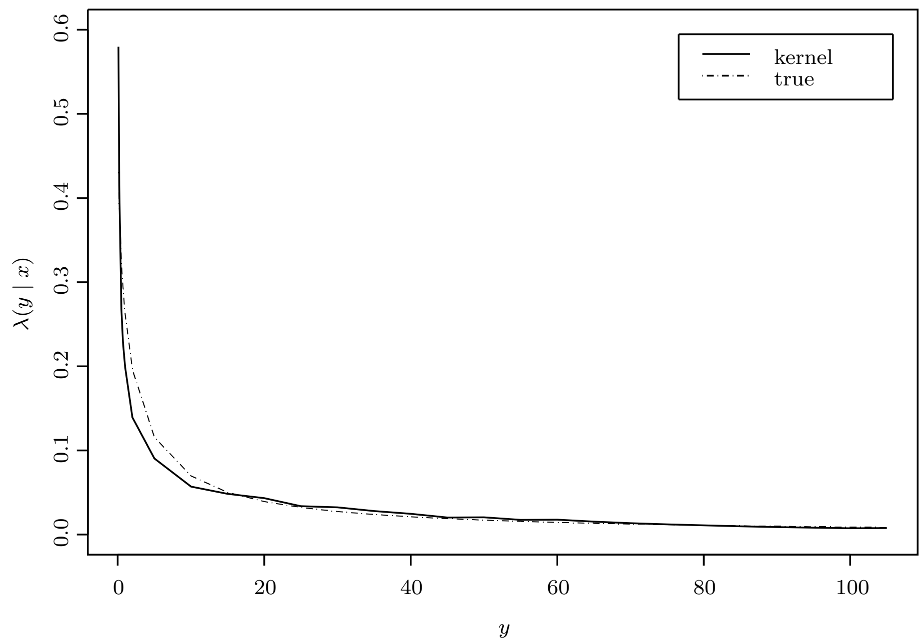

Figure 8: Kernel hazard rate estimate and true hazard rate

This plot shows the local linear hazard rate estimate $y \longrightarrow \hat{\lambda}(y \mid x)$ for $x = 1$, based on the normal reference rule (with $d = 0$) for one simulated sample from the logistic accelerated lifetime model. The dashed line represents the true hazard rate.

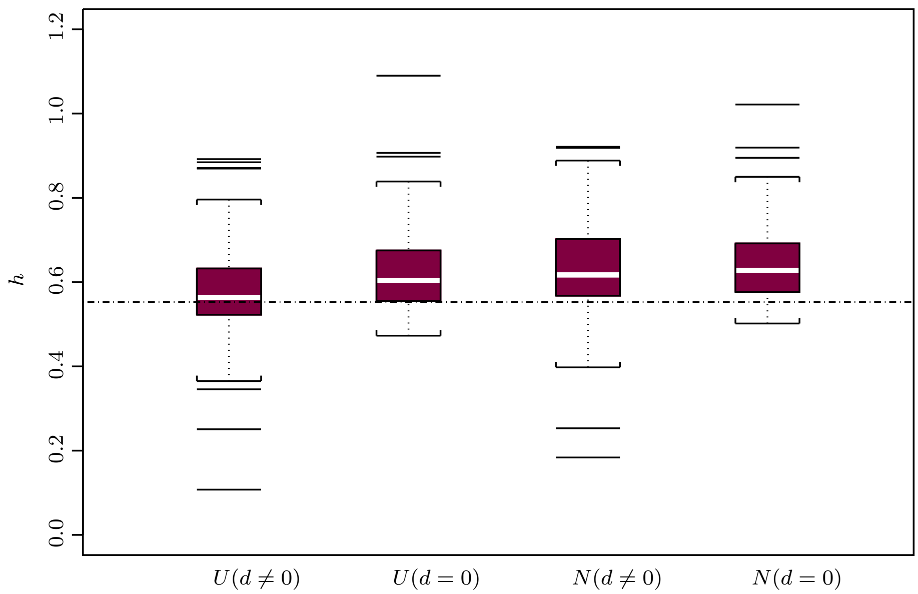

Figure 9: Values of h for each of the 100 samples from a normal accelerated lifetime model with censoring

These boxplots show the estimated values of h based on different reference distributions. The dashed line shows the optimal value of h.

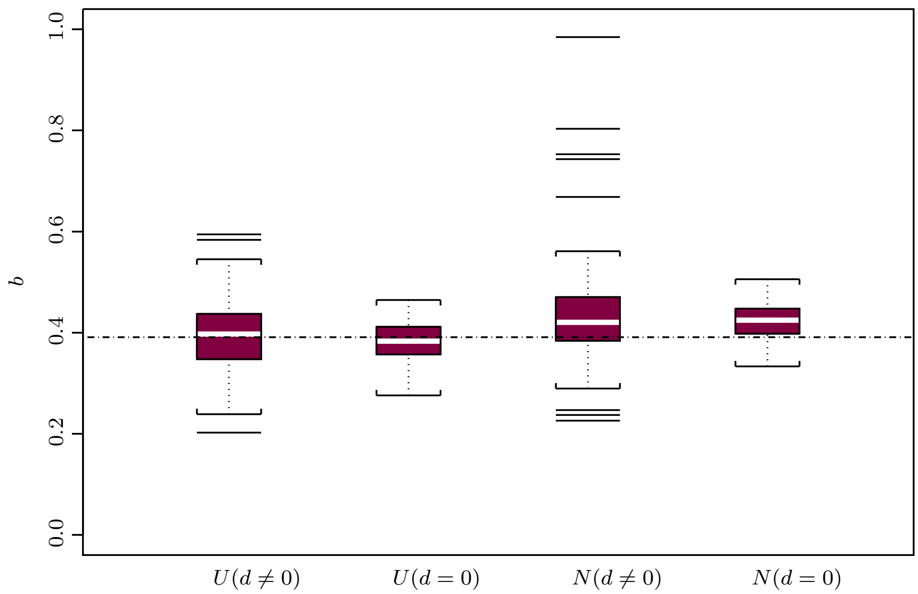

Figure 10: Values of b for each of the 100 samples from a normal accelerated lifetime model with censoring

These boxplots show the estimated values of h based on different reference distributions. The dashed line shows the optimal value of b.

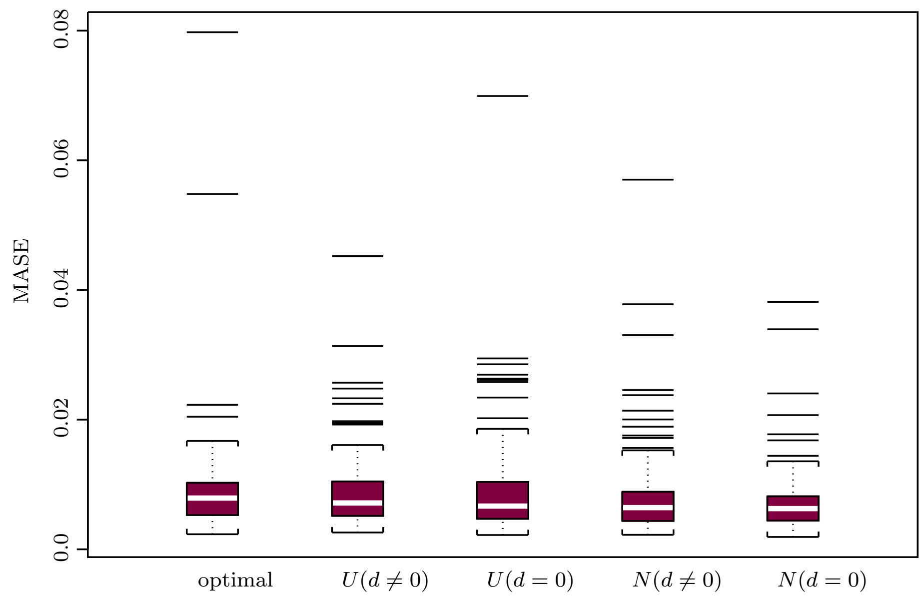

Figure 11: Estimated MASE values for each of the 100 samples from a normal accelerated lifetime model with censoring

These boxplots display the estimated MASE values based on different reference distributions. The first boxplot shows the MASE value for 100 samples based on the optimal bandwidths h and b.

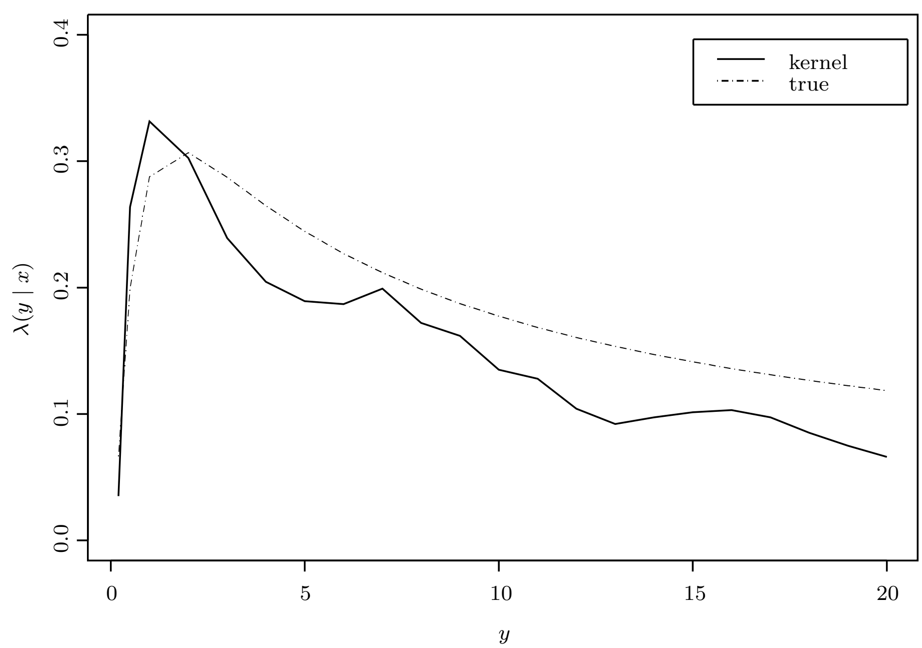

Figure 12: Kernel hazard rate estimate and true hazard rate

This plot shows the local linear hazard rate estimate $y \longrightarrow \hat{\lambda}(y \mid x)$ for $x = 1$, based on the normal reference rule (with $d = 0$) for one simulated sample from the normal accelerated lifetime model with censoring. The dashed line represents the true hazard rate.

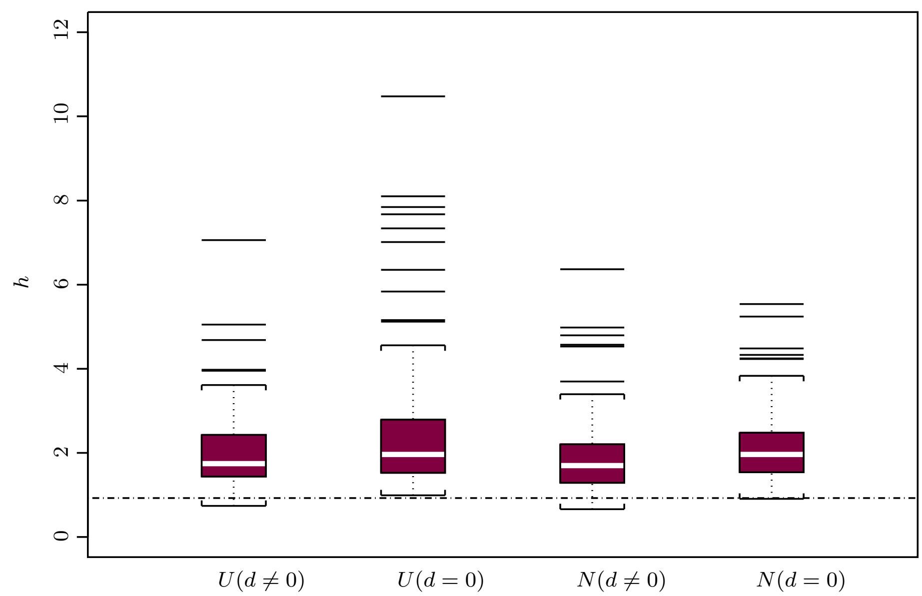

Figure 13: Values of h for each of the 100 samples from a logistic accelerated lifetime model with censoring

These boxplots show the estimated values of h based on different reference distributions. The dashed line shows the optimal value of h.

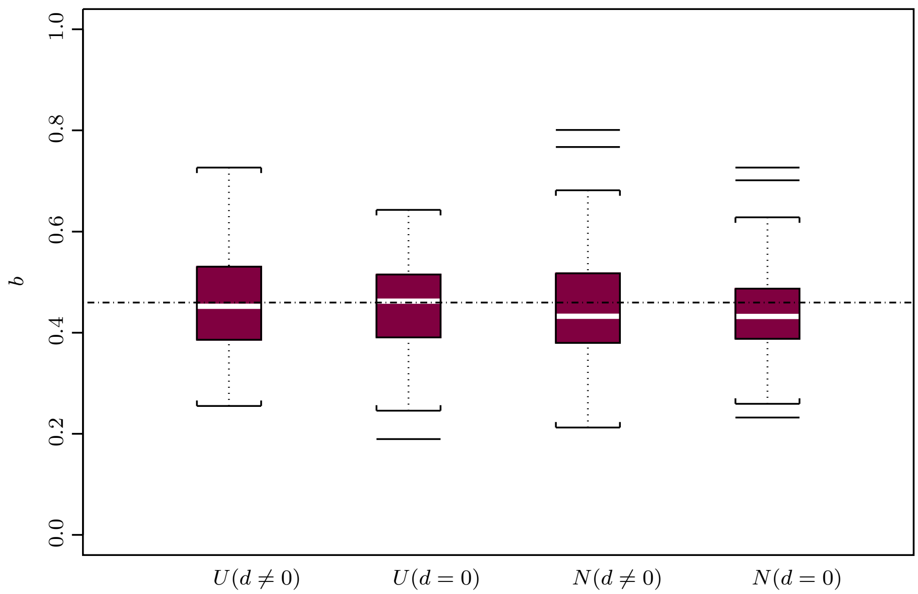

Figure 14: Values of b for each of the 100 samples from a logistic accelerated lifetime model with censoring

These boxplots show the estimated values of h based on different reference distributions. The dashed line shows the optimal value of b.

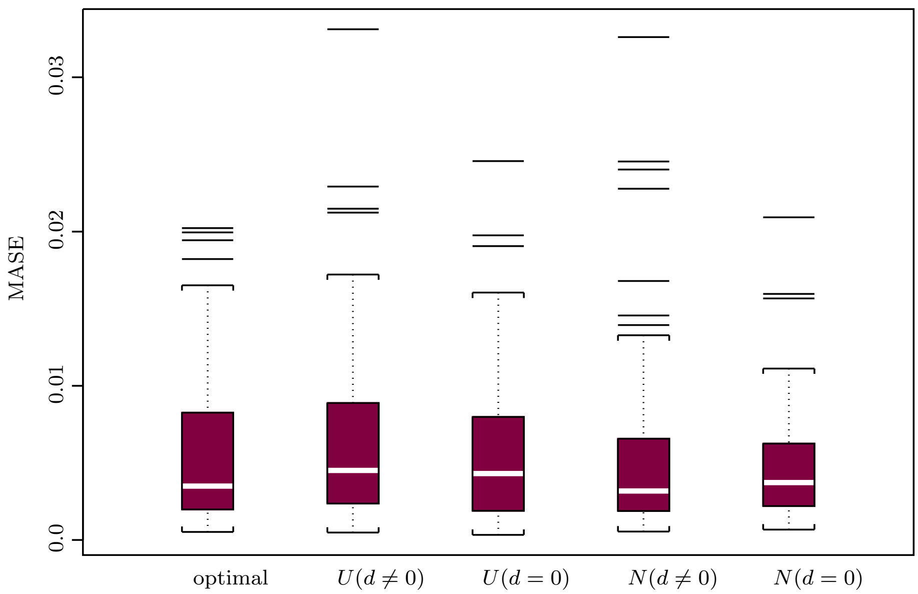

Figure 15: Estimated MASE values for each of the 100 samples from a logistic accelerated lifetime model with censoring

These boxplots display the estimated MASE values based on different reference distributions. The first boxplot shows the MASE value for 100 samples based on the optimal bandwidths h and b.

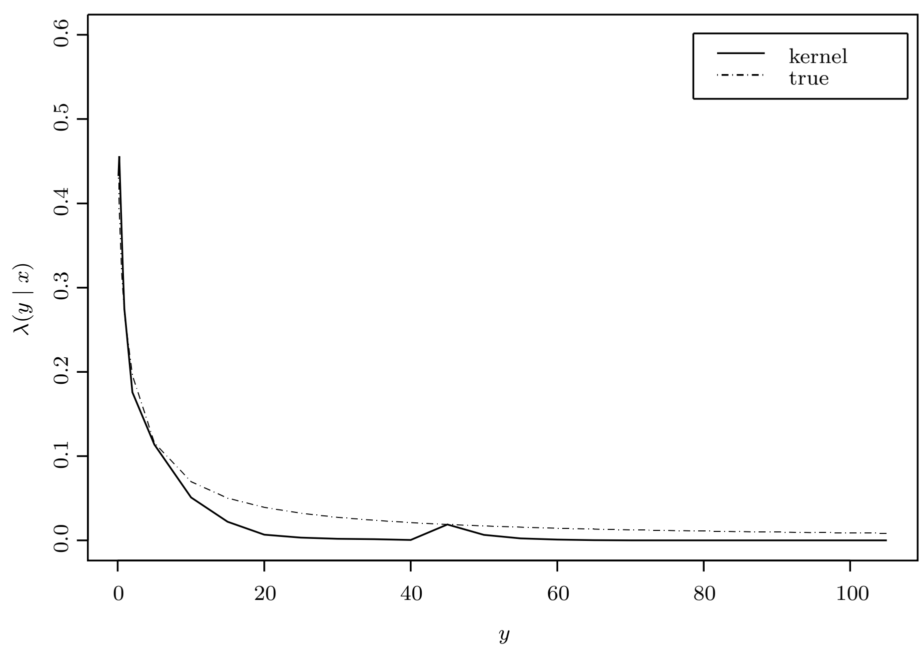

Figure 16: Kernel hazard rate estimate and true hazard rate

This plot shows the local linear hazard rate estimate $y \longrightarrow \hat{\lambda}(y \mid x)$ for $x = 1$, based on the logistic reference rule (with $d = 0$) for one simulated sample from the normal accelerated lifetime model with censoring. The dashed line represents the true hazard rate.

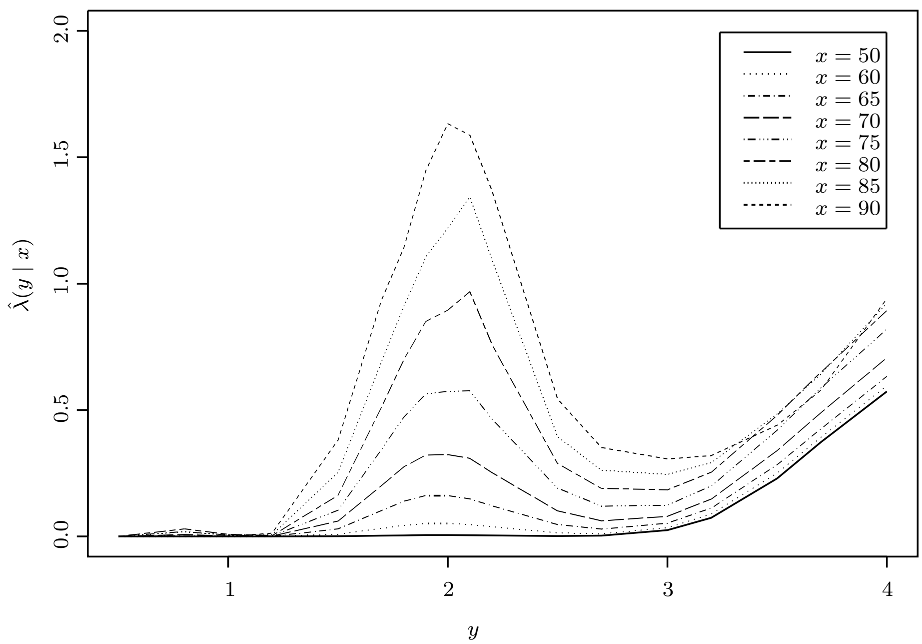

Figure 17: Estimated hazard rate for Old Faithful geyser data

This figure displays the estimated conditional hazard rate $y \longrightarrow \hat{\lambda}(y \mid x)$ for several values of x. The bandwidths used to estimate the conditional hazard rate are obtained using the normal reference rule with $d \neq 0$.

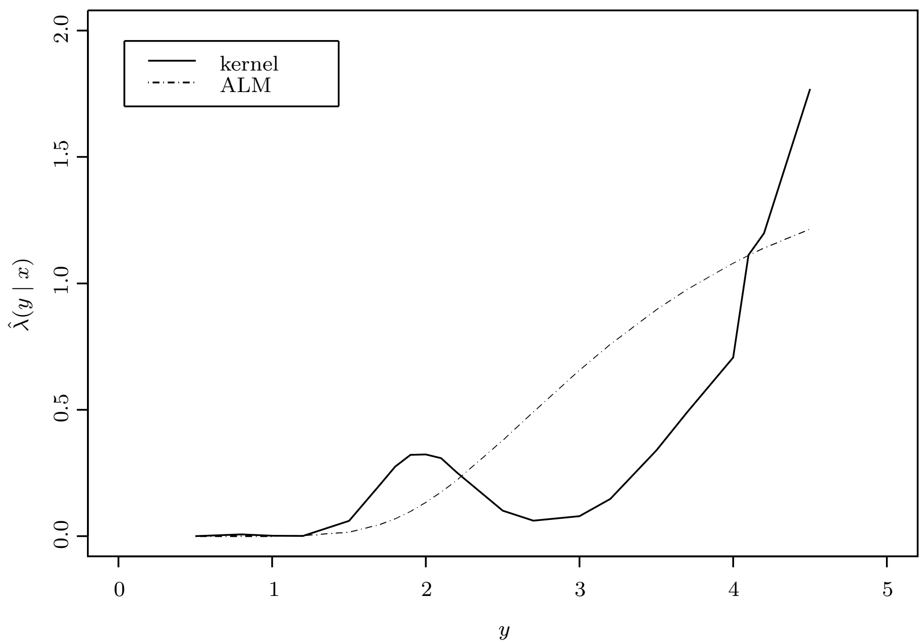

Figure 18: Estimated hazard rate for Old Faithful geyser data: parametric versus nonparametric estimate

This figure displays two estimates of the conditional hazard rate $y \longrightarrow \hat{\lambda}(y \mid x)$ for $x = 70$, based on the Old Faithful geyser data. The solid line represents the local linear hazard rate estimate (normal reference rule, $d \neq 0$) and the dashed line corresponds to the parametric hazard estimate based on a normal accelerated lifetime model (ALM).

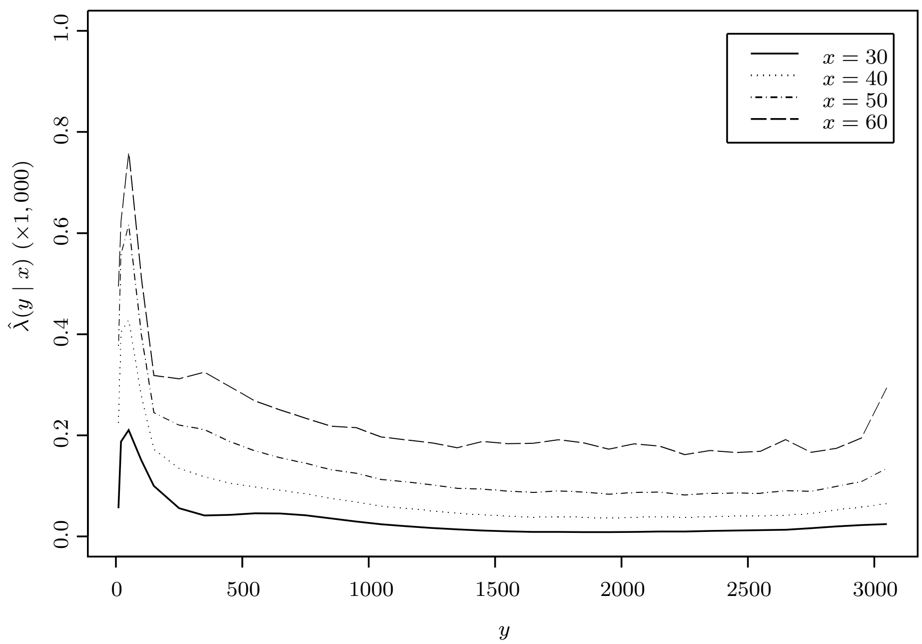

Figure 19: Estimated hazard rate for kidney transplant data (white males)

This figure displays the estimated conditional hazard rate $y \longrightarrow \hat{\lambda}(y \mid x)$ for several values of x. The bandwidths used to estimate the conditional hazard rate are obtained using the normal reference rule with $d = 0$.

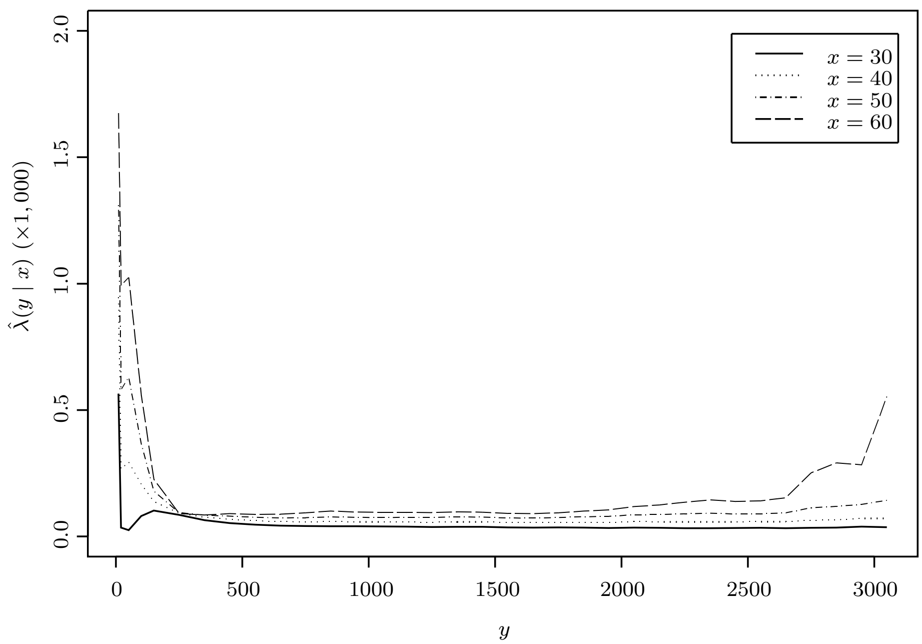

Figure 20: Estimated hazard rate for kidney transplant data (white females)

This figure displays the estimated conditional hazard rate $y \longrightarrow \hat{\lambda}(y \mid x)$ for several values of x. The bandwidths used to estimate the conditional hazard rate are obtained using the normal reference rule with $d = 0$.

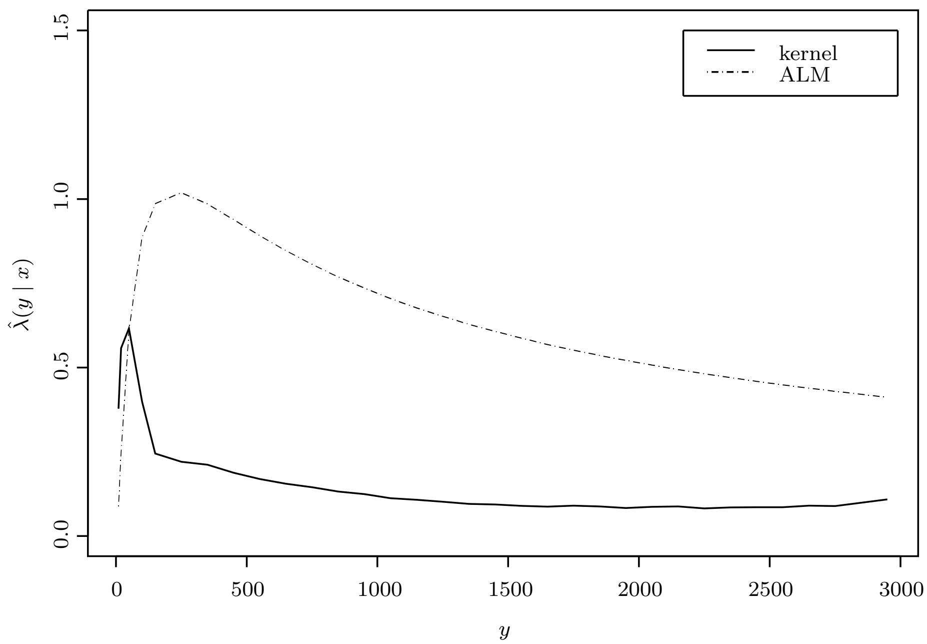

Figure 21: Estimated hazard rate for kidney transplant data: parametric versus nonparametric estimate

This figure displays two estimates of the conditional hazard rate $y \longrightarrow \hat{\lambda}(y \mid x)$ for $x = 40$, based on the survival times of white male patients following a kidney transplant. The solid line represents the local linear hazard rate estimate (normal reference rule, $d \neq 0$) and the dashed line corresponds to the parametric hazard estimate based on a normal accelerated lifetime model (ALM).

A Proofs

Uncensored data

We assume that $(X_i, Y_i)_i$ is a strictly stationary process having the same marginal distribution as (X, Y), where X and Y are scalars. Extension to the case that X is d-dimensional is straightforward and omitted. The kernels $k(\cdot)$ and $w(\cdot)$ are symmetric probability density functions. Both b and h denote sequences of bandwidths such that $b, h \to 0$ and $nbh \to \infty$ as $n \to \infty$.

The local linear estimator for the conditional density proposed by Fan, Yao, and Tong (1996) is defined as

$$\hat{f}(y \mid x) = \frac{\sum_{i=1}^{n} w_{h,i}^{*}(x) k_b(y - Y_i)}{\sum_{i=1}^{n} w_{h,i}^{*}(x)}, \tag{A.1}$$

where

$$w_{h,i}^{*}(x) = w_{h,i}(x)(s_{n,2}(x) - (x - X_i)s_{n,1}(x)); \tag{A.2}$$

$$w_{h,i}(x) = \frac{1}{h} w\Big(\frac{x - X_i}{h}\Big); \tag{A.3}$$

$$s_{n,\ell}(x) = \sum_{i=1}^{n} w_{h,i}(x)(x - X_i)^{\ell} \qquad [\ell = 1, 2]. \tag{A.4}$$

and

$$k_b(y) = \frac{1}{b} k\Big(\frac{y}{b}\Big).$$

Moreover, the local linear estimator for the conditional survival function is defined as

$$\hat{S}(y \mid x) = \frac{\sum_{i=1}^{n} w_{h,i}^{*}(x) I(Y_i \geq y)}{\sum_{i=1}^{n} w_{h,i}^{*}(x)}. \tag{A.5}$$

The local linear hazard rate estimator is defined as the ratio of the kernel density (A.1) and the kernel survivor function (A.5); i.e.

$$\hat{\lambda}(y \mid x) = \frac{\hat{f}(y \mid x)}{\hat{S}(y \mid x)}. \tag{A.6}$$

Before deriving the asymptotic bias and variance of the above local linear conditional hazard rate estimator, we introduce some notation. The marginal density of X_i is denoted by $f(\cdot)$. Moreover, let

$$\mu_k = \int t^2 k(t) dt;$$

$$\mu_2 = \int t^2 w(t) dt;$$

$$\nu_k = \int \{k(t)\}^2 dt;$$

$$\nu_0 = \int \{w(t)\}^2 dt;$$

To prove the asymptotic normality of the local linear hazard rate estimator, we need the following theorem.

Theorem A.1 Under conditions (1)-(6) the local linear density estimator $\hat{f}(y \mid x)$ as defined in equation (A.1) satisfies, for $n \to \infty, h, b \to 0$, and $nhb \to \infty$,

$$\sqrt{nhb}(\hat{f}(y \mid x) - \mathbb{E}\hat{f}(y \mid x)) \xrightarrow{d} \mathcal{N}(0, \sigma_{xy}^2), \tag{A.7}$$

where

$$\sigma_{xy}^2 = \frac{v_k v_0 f(y \mid x)}{f(x)}. \tag{A.8}$$

Moreover, the asymptotic bias of $\hat{f}(y \mid x)$ is given by

$$\mathbb{E}\hat{f}(y \mid x) - f(y \mid x) = \frac{h^2\mu_2}{2}\frac{\partial^2 f(y \mid x)}{\partial x^2} + \frac{b^2\mu_k}{2}\frac{\partial^2 f(y \mid x)}{\partial y^2} + o(h^2 + b^2). \tag{A.9}$$

Proof: See Fan, Yao, and Tong (1996). ■

Another theorem that we need to prove the asymptotic normality of the local linear hazard rate estimator is directly based on Masry and Fan (1997).

Theorem A.2 Under conditions (a)-(g) and for $n \to \infty, h \to 0$ and $nh \to \infty$, the local linear estimator for the survivor function $\hat{S}(y \mid x)$ as defined by equations (A.1) and (A.5) satisfies

$$\sqrt{nh}(\hat{S}(y \mid x) - \mathbb{E}\hat{S}(y \mid x)) \xrightarrow{d} \mathcal{N}(0, \nu_{xy}^2), \tag{A.10}$$

where

$$\nu_{xy}^2 = \frac{v_0 S(y \mid x)(1 - S(y \mid x))}{f(x)}. \tag{A.11}$$

Furthermore, the asymptotic bias equals

$$\mathbb{E}\hat{S}(y \mid x) - S(y \mid x) \quad = \quad \frac{h^2\mu_2}{2}\frac{\partial^2 S(y \mid x)}{\partial x^2} + o(h^2); \tag{A.12}$$

Proof: The theorem follow directly from Masry and Fan (1997), applied to the local linear regression of $Y = I(Y \geq y)$ on X. ■

Theorem A.3 Assume that $S(y \mid x) > 0$. Under conditions (1)-(4), (6), and (d)-(g) and for $n \to \infty, h, b \to 0$ and $nhb \to \infty$ the local linear hazard rate estimator as defined in equation (A.6) is asymptotically normally distributed; i.e.

$$\sqrt{nhb}(\hat{\lambda}(y \mid x) - \mathbb{E}\hat{\lambda}(y \mid x)) \xrightarrow{d} \mathcal{N}(0, \tau_{xy}^2), \tag{A.13}$$

where

$$\tau_{xy}^2 = \frac{v_k v_0 \lambda(y \mid x)}{f(x)S(y \mid x)}. \tag{A.14}$$

Moreover, the asymptotic bias of $\hat{\lambda}(y \mid x)$ is given by

$$\begin{aligned}\mathbb{E}\hat{\lambda}(y \mid x) - \lambda(y \mid x) &= \frac{1}{S(y \mid x)}\Big(\frac{h^2\mu_2}{2}\frac{\partial^2 f(y \mid x)}{\partial x^2} + \frac{b^2\mu_k}{2}\frac{\partial^2 f(y \mid x)}{\partial y^2} \\ &\quad -\lambda(y \mid x)\frac{h^2\mu_2}{2}\frac{\partial^2 S(y \mid x)}{\partial x^2}\Big) + o(h^2 + b^2).\end{aligned} \tag{A.15}$$

Proof: Throughout we assume that $n \to \infty, h, b \to 0$ and $nhb \to \infty$. We first prove the asymptotic normality. First note that $\hat{S}(y \mid x)$ is consistent for $S(y \mid x)$, since $\mathbb{E}\hat{S}(y \mid x) \to S(y \mid x)$ and $\mathbb{V}\text{ar}\, \hat{S}(y \mid x) \to 0$ as $n \to \infty, h, b \to 0$ and $nhb \to \infty$ (see Lee (1996), p. 130). Therefore, using Theorem A.1 and Slutsky's lemma, it follows that

$$\sqrt{nhb}\Big(\frac{\hat{f}(y \mid x)}{\hat{S}(y \mid x)} - \frac{\mathbb{E}\hat{f}(y \mid x)}{\hat{S}(y \mid x)}\Big) \xrightarrow{d} \mathcal{N}(0, \tau_{xy}^2), \tag{A.16}$$

where

$$\tau_{xy}^2 = \frac{v_k v_0 \lambda(y \mid x)}{f(x)S(y \mid x)}. \tag{A.17}$$

However, our focus is on the asymptotic distribution of

$$\sqrt{nhb}\Big(\frac{\hat{f}(y \mid x)}{\hat{S}(y \mid x)} - \frac{\mathbb{E}\hat{f}(y \mid x)}{\mathbb{E}\hat{S}(y \mid x)}\Big). \tag{A.18}$$

Notice that we can write

$$\begin{aligned}\sqrt{nhb}\Big(\frac{\hat{f}(y \mid x)}{\hat{S}(y \mid x)} - \frac{\mathbb{E}\hat{f}(y \mid x)}{\mathbb{E}\hat{S}(y \mid x)}\Big) &= \underbrace{\sqrt{nhb}\Big(\frac{\hat{f}(y \mid x)}{\hat{S}(y \mid x)} - \frac{\mathbb{E}\hat{f}(y \mid x)}{\hat{S}(y \mid x)}\Big)}_{(1)} \\ &\quad + \underbrace{\sqrt{nhb}\Big(\frac{\mathbb{E}\hat{f}(y \mid x)}{\hat{S}(y \mid x)} - \frac{\mathbb{E}\hat{f}(y \mid x)}{\mathbb{E}\hat{S}(y \mid x)}\Big)}_{(2)}.\end{aligned} \tag{A.19}$$

The asymptotic distribution of part (1) of equation (A.19) has already been derived and is given in equation (A.16). Part (2) of equation (A.19) can be rewritten as

$$\frac{\sqrt{nhb}(\mathbb{E}\hat{S}(y \mid x) - \hat{S}(y \mid x))\mathbb{E}\hat{f}(y \mid x)}{\hat{S}(y \mid x)\mathbb{E}\hat{S}(y \mid x)}. \tag{A.20}$$

Let

$$\hat{Z}(y \mid x) = \sqrt{nhb}(\mathbb{E}\hat{S}(y \mid x) - \hat{S}(y \mid x)). \tag{A.21}$$

Note that $\mathbb{E}\hat{Z} = 0$ and that $\mathbb{V}\text{ar}\, Z(y \mid x) \to 0$ for $n \to \infty, h, b \to 0$ and $nhb \to \infty$. Therefore (see Lee (1996)),

$$Z(y \mid x) \xrightarrow{p} 0, \tag{A.22}$$

and thus

$$\frac{\sqrt{nhb}(\mathbb{E}\hat{S}(y \mid x) - S(y \mid x))\mathbb{E}\hat{f}(y \mid x)}{\hat{S}(y \mid x)\mathbb{E}\hat{S}(y \mid x)} \xrightarrow{p} 0. \tag{A.23}$$

Using Slutsky's lemma again, it follows that

$$\sqrt{nhb}\Big(\frac{\hat{f}(y \mid x)}{\hat{S}(y \mid x)} - \frac{\mathbb{E}\hat{f}(y \mid x)}{S(y \mid x)}\Big) \xrightarrow{d} \mathcal{N}(0, \tau_{xy}^2). \tag{A.24}$$

Now we turn to the asymptotic bias. Using Theorems A.1 and A.2, we write

$$\begin{aligned} \mathbb{E}\hat{\lambda}(y \mid x) &= \frac{\mathbb{E}\hat{f}(y \mid x)}{\mathbb{E}\hat{S}(y \mid x)} \\ &= \frac{f(y \mid x) + \frac{h^2\mu_2}{2}\frac{\partial^2 f(y|x)}{\partial x^2} + \frac{b^2\mu_k}{2}\frac{\partial^2 f(y|x)}{\partial y^2}}{S(y \mid x) + \frac{h^2\mu_2}{2}\frac{\partial^2 S(y|x)}{\partial x^2}} + o(h^2 + b^2). \end{aligned} \tag{A.25}$$

Using the result

$$\frac{1}{s+\delta} = \frac{1}{s} + \frac{\delta}{s^2} + o(\delta), \tag{A.26}$$

taken from Hyndman, Bashtannyk, and Grunwald (1996), we obtain

$$\begin{aligned} \mathbb{E}\hat{\lambda}(y \mid x) - \lambda(y \mid x) &= \frac{1}{S(y \mid x)}\Big(\frac{h^2\mu_2}{2}\frac{\partial^2 f(y \mid x)}{\partial x^2} + \frac{b^2\mu_k}{2}\frac{\partial^2 f(y \mid x)}{\partial y^2} \\ &\quad -\lambda(y \mid x)\frac{h^2\mu_2}{2}\frac{\partial^2 S(y \mid x)}{\partial x^2}\Big) + o(h^2 + b^2). \end{aligned} \tag{A.27}$$

■

Censored data

We consider the situation that, due to censoring, the durations of interest $(Y_i)_i$ are not completely observed. We assume a random-right censoring model with censoring times $(C_i)_i$. As in the uncensored case, $(X_i)_i$ are the covariates. The observed data are the triples $(X_i, T_i, \delta_i)_i$, where $T_i = \min(Y_i, C_i)$ and $\delta_i = I(Y_i \leq C_i)$. We assume that $(X_i, Y_i, C_i, \delta_i)$ is a strictly stationary process having the same marginal distributions as X, Y, C, and δ, where X, Y and C are scalars. Moreover, we assume that the random variables $Y \mid X$ and $C \mid X$ are independent, with conditional survivor functions $S(\cdot \mid x)$ and $G(\cdot \mid x)$ and conditional densities $f(\cdot \mid x)$ and $g(\cdot \mid x)$. Let $H(\cdot \mid x)$ denote the survivor distribution function of T and let $h(\cdot, \cdot \mid x)$ denote the conditional density of (T, δ) and write $r(y \mid x) = h(y, 1 \mid x)$. The object of interest is the conditional hazard rate

$$\lambda_{Y|X}(y \mid x) = \lambda(y \mid x) = \frac{f(y \mid x)}{S(y \mid x)}. \tag{A.28}$$

From the independence of Y and C it follows that

$$H(\cdot \mid x) = S(\cdot \mid x)G(\cdot \mid x). \tag{A.29}$$

As a consequence,

$$r(\cdot \mid x) = f(\cdot \mid x)G(\cdot \mid x). \tag{A.30}$$

The local linear hazard rate estimator in the presence of censoring is directly based on definition (42) and is naturally defined as

$$\hat{\lambda}(y \mid x) = \frac{\hat{r}(y \mid x)}{\hat{H}(y \mid x)}. \tag{A.31}$$

In definition (A.31) the kernel estimator for the survivor function $\hat{H}(y \mid x)$ is defined as before; i.e.

$$\hat{H}(y \mid x) = \frac{\sum_{i=1}^{n} w_{h,i}^{*}(x) I(T_i \geq y)}{\sum_{i=1}^{n} w_{h,i}^{*}(x)}, \tag{A.32}$$

where

$$w_{h,i}^{*}(x) = w_{h,i}(x)(s_{n,2} - (x - X_i)s_{n,1}); \tag{A.33}$$

$$w_{h,i}(x) = \frac{1}{h} w\Big(\frac{x - X_i}{h}\Big); \tag{A.34}$$

$$s_{n,\ell} = \sum_{i=1}^{n} w_{h,i}(x)(x - X_i)^{\ell} \qquad [\ell = 1, 2]. \tag{A.35}$$

To account for censoring we take as kernel density estimator

$$\hat{r}(y \mid x) = \frac{\sum_{i=1}^{n} w_{h,i}^{*}(x) k_b(y - T_i)\delta_i}{\sum_{i=1}^{n} w_{h,i}^{*}(x)}. \tag{A.36}$$

For the censored kernel density estimator (A.36) we can formulate the following theorem.

Theorem A.4 Under conditions (A)-(F) and for $nhb \to \infty$, $n, b \to 0$ as $n \to \infty$

$$\sqrt{nhb}(\hat{r}(y \mid x) - \mathbb{E}\hat{r}(y \mid x)) \xrightarrow{d} \mathcal{N}(0, \nu_{xy}^2), \tag{A.37}$$

where

$$\nu_{xy}^2 = \frac{v_k v_0 f(y \mid x) G(y \mid x)}{f(x)}. \tag{A.38}$$

Moreover, the asymptotic bias of $\hat{r}(y \mid x)$ is given by

$$\mathbb{E}\hat{r}(y \mid x) - r(y \mid x) = \frac{h^2\mu_2}{2}C_1(x,y) + \frac{b^2\mu_k}{2}C_2(x,y) + o(h^2+b^2). \tag{A.39}$$

Here

$$C_1(x,y) = 2\frac{\partial f(y \mid x)}{\partial x}\frac{\partial G(y \mid x)}{\partial x} + f(y \mid x)\frac{\partial^2 G(y \mid x)}{\partial x^2} + \frac{\partial^2 f(y \mid x)}{\partial x^2}G(y \mid x); \tag{A.40}$$

$$C_2(x,y) = 2\frac{\partial f(y \mid x)}{\partial y}\frac{\partial G(y \mid x)}{\partial y} + f(y \mid x)\frac{\partial^2 G(y \mid x)}{\partial y^2} + \frac{\partial^2 f(y \mid x)}{\partial y^2}G(y \mid x). \tag{A.41}$$

Proof: The proof is very similar to the proof of Theorem A.1 as given by See Fan, Yao, and Tong (1996) and is therefore omitted. ■

We can now formulate a convergence theorem for the censored local linear hazard rate estimator.

Theorem A.5 Assume that $S(y \mid x) > 0$ and $G(y \mid x) > 0$. Under conditions (A)-(D), (F), and (d)-(g), the censored local linear hazard rate estimator $\hat{\lambda}(y \mid x)$ satisfies

$$\sqrt{nhb}(\hat{\lambda}(y \mid x) - \mathbb{E}\hat{\lambda}(y \mid x)) \xrightarrow{d} \mathcal{N}(0, \xi^2_{xy}), \tag{A.42}$$

where

$$\xi^2_{xy} = \frac{v_k v_0 \lambda(y \mid x)}{f(x)H(y \mid x)}. \tag{A.43}$$

Moreover, the asymptotic bias of $\hat{\lambda}(y \mid x)$ is given by

$$\begin{aligned}\mathbb{E}\hat{\lambda}(y \mid x) - \hat{\lambda}(y \mid x) = {} & \frac{1}{H(y \mid x)}\Big(\frac{h^2\mu_2}{2}C_1(x,y) + \frac{b^2\mu_k}{2}C_2(x,y) \\ & -\lambda(y \mid x)\frac{h^2\mu_2}{2}C_3(x,y)\Big) + o(h^2+b^2).\end{aligned} \tag{A.44}$$

Here

$$C_3(x,y) = 2\frac{\partial S(y \mid x)}{\partial x}\frac{\partial G(y \mid x)}{\partial x} + S(y \mid x)\frac{\partial^2 G(y \mid x)}{\partial x^2} + \frac{\partial^2 S(y \mid x)}{\partial x^2}G(y \mid x). \tag{A.45}$$

Proof:
Note that only the numerator of the local linear hazard rate estimator changes relative to the uncensored case. Hence, the proof of the asymptotic normality of the censored local linear hazard rate estimator is along the same as lines as for uncensored data and therefore omitted.
■

Assumptions

The following assumptions are taken from Fan, Yao, and Tong (1996).

(1) The kernel functions $k(\cdot)$ and $w(\cdot)$ are symmetric and bounded with bounded supports.

(2) The process $(X_i, Y_i)_i$ is ρ-mixing with $\sum_\ell \rho(\ell) < \infty$.

(3) There exists a sequence of positive integers $s_n = o((nhb)^{0.5})$ and $(n/(hb))^{0.5}\rho(s_n) \to 0$.

(4) The function $f(y \mid x)$ has bounded and continuous third order derivatives at (x, y).

(5) $f(\cdot)$ is continuous in x and $f(x) > 0$.

(6) The joint density of the distinct elements of $(X_0, Y_0, X_\ell, Y_\ell)$ for $\ell > 0$ is bounded by a constant that is independent of ℓ.

The conditions below are from Masry and Fan (1997).

(a) The kernel $w(\cdot)$ is bounded with bounded support.

(b) The process $(X_i, Y_i)_i$ is ρ-mixing with $\sum_\ell \rho(\ell) < \infty$.

(c) There exists a sequence of positive integers $s_n = o((nh)^{0.5})$ and $(n/h)^{0.5}\rho(s_n) \to 0$.

(d) $\mathbb{E}Y_0 \leq \infty$.

(e) The conditional density $f_{X_0,X_\ell|Y_0,Y_\ell}(x_0, x_\ell \mid y_0, y_\ell) \leq A < \infty$ for $\ell > 0$.

(f) The function $S(\cdot \mid \cdot)$ has bounded and continuous third order derivatives at (x, y).

(g) $f(\cdot)$ is continuous in x and $f(x) > 0$.

The conditions below apply to censored data.

(A) The kernel functions $k(\cdot)$ and $w(\cdot)$ are symmetric and bounded with bounded support.

(B) The process $(X_i, Y_i)_i$ is ρ-mixing with $\sum_\ell \rho(\ell) < \infty$.

(C) There exists a sequence of positive integers $s_n = o((nhb)^{0.5})$ and $(n/(hb))^{0.5}\rho(s_n) \to 0$.

(D) The functions $f(\cdot \mid \cdot)$ and $G(\cdot \mid \cdot)$ have bounded and continuous third order derivatives at (x, y).

(E) $f(\cdot)$ is continuous in x.

(F) The joint density of the distinct elements of $(X_0, Y_0, \delta_0, X_\ell, Y_\ell, \delta_\ell)$ for $\ell > 0$ is bounded by a constant that is independent of ℓ.

We notice that conditions (1), (a), and (A) have been chosen for the sake of simplicity and are not the weakest conditions possible. In particular, the Gaussian kernel is permitted.