I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0114

ESTIMATION IN CHANGE POINT MODELS FOR HAZARD FUNCTION WITH CENSORED DATA

I. GIJBELS AND Ü. GÜRLER

http://www.stat.ucl.ac.be

Estimation in change point models for hazard function with censored data

Irène Gijbels and Ülkü Gürler

Université Catholique de Louvain, Belgium

Bilkent University, Turkey

Abstract

The hazard function plays an important role in reliability or survival studies since it describes the instantaneous risk of failure of items at a time point, given that they have not failed before. In some real life applications, abrupt changes in the hazard function are observed due to overhauls, major operations or specific maintenance activities. In such situations it is of interest to detect the location where such a change occurs and estimate the size of the change. In this paper we consider the problem of estimating the change point in a hazard function when the observed variables are subject to random censoring. When the hazard function is piecewise constant and there is a single change point, we suggest an estimation procedure that is based on certain structural properties and on least squares ideas. A simulation study is carried out to compare the performance of this estimator with two estimators available in the literature: an estimator based on a functional of the Nelson-Aalen estimator and a maximum likelihood estimator. The proposed least squares estimator turns out to be less biased than the other two estimators, but has a larger variance. We illustrate the estimation method on some real data sets.

Key Words: Change point hazard model, cumulative hazard function, least squares estimation, random censoring.

This research is partially supported by NATO Collaborative Research Grant CRG 950271, and by 'Projet d'Actions de Recherche Concertées', No. 98/03–217, from the Belgian governement.

1 Introduction

Let X be a nonnegative random variable, denoting the lifetime of an item or in a more general context, the time to some event, such as time to failure of a component, time until sero-conversion in AIDS epidemic or the survival time after a certain operation. In the analysis of data related to the variables mentioned above, the hazard rate $\lambda(x)$, also defined as the instantaneous failure rate function plays an important role since it describes the immediate risk of failure of an item (or individual) at time x, given that it has not failed up to that time. In practical situations abrupt changes in the hazard function at unknown points can be encountered due to for example a major maintenance action in reliability or a new treatment in survival analysis.

In this paper we consider estimation methods for hazard functions with a jump discontinuity, when the observed data is subject to right censoring. In particular, suppose the interest is in the variable X but instead of observing an i.i.d. sample of X, we observe the random sample $(T_i, \delta_i), i = 1, 2, \ldots, n$, where $T_i = \min(X_i, C_i)$ and $\delta_i = 1$ if $X_i \leq C_i$ and zero otherwise. Here the random variable C is the censoring variable which is assumed to be independent of X. In the reliability context, censoring can occur due to the termination of the observation period so that the lifetime of an item may not be observed completely, or in medical research, incomplete data for the survival times of the patients can be observed due to drop outs from the study. Due to this censoring effect, the methods available for the analysis of i.i.d. data are clearly not appropriate in this set-up and other methods are needed. For continuous hazard rate functions, parametric and nonparametric estimation methods are available in the literature. For hazard functions with discontinuities, the literature is not so rich as will be discussed briefly below.

Suppose that the hazard function of X has a jump discontinuity at an unknown point τ, and that the interest is in estimating this location point based on the observed censored data. A general model is given by

$$\lambda(x) = \lambda_1(x) + \lambda_2(x) I_{\{\tau < x\}}, \tag{1.1}$$

where λ_1 and λ_2 are smooth functions and I_A denotes the indicator function on the set A. This problem belongs to the class of so-called change point problems. The literature on change point problems in the censored data situation is rather small. A nice overview of the existing literature is given in Müller and Wang (1994). In view of certain applications it might also be of interest to estimate rapid changes of the hazard function. This problem has been studied in Müller and Wang (1990).

The above model is rather general, and classical methods for detecting a jump discontinuity are based on differences of a smooth estimate for the left-hand and the right-hand limits. The location point τ is then estimated as the location where these absolute differences are maximal.

In this paper we concentrate on the simple model

$$\lambda(x) = \beta + \theta I_{\{\tau<x\}}, \tag{1.2}$$

where β and θ are such that $\beta > 0$ and $\beta + \theta > 0$. It is easy to see that under this model the cumulative hazard function is given by

$$\Lambda(x) = \int_0^x \lambda(t)dt = \beta x + \theta(x-\tau)I_{\{\tau<x\}}. \tag{1.3}$$

Model (1.2) has been considered previously in the literature and three different approaches for inference under this model have been used. A first approach is based on maximum likelihood and is followed by Matthews and Farewell (1982), Nguyen, Rogers and Walker (1984), Henderson (1990) and Loader (1991). Testing for a constant hazard rate versus an alternative involving a change-point as in model (1.2), based on score statistics was considered by Matthews, Farewell and Pyke (1985). Finally, Chang, Chen and Hsiung (1994) exploit the structural properties under the simple model (1.2). Their estimation procedure relies on a certain functional of the estimated cumulative hazard function.

In this paper we propose an alternative approach for estimation of τ, the location of the jump point. This estimator is then used to obtain estimators for β and θ. Our approach relies on certain structural properties of the cumulative hazard function. With the estimated cumulative hazard we then fit an appropriate regression model to synthetically obtained data.

The paper is organized as follows. In Section 2 we describe in more detail the specific structural properties under model (1.2). We give a brief discussion on the estimation methods provided in Loader (1991) and Chang, Chen and Hsiung (1994), and introduce the new estimation approach based on least squares ideas. A detailed simulation study, presented in Section 3, reveals the specific merits of each of the three estimation methods discussed in Section 2. Illustrations with real data are also provided.

2 Estimation methods

In this section we briefly present the maximum likelihood estimators and the estimation procedure proposed by Chang, Chen and Hsiung (1994) in the first two sections. The alternative estimation procedure, based on a least squares idea will be explained in the last section.

For each of the procedures we will focus on the estimation of the unknown location point τ. Apart from estimating τ it is also of interest to estimate β and θ in model (1.2), i.e. the value of the hazard rate before the jump point and the size of the jump in τ, respectively. Once an estimator for τ is available estimates for β and θ are simply obtained by classical maximum likelihood estimators. These are described in (2.3), where we replace τ by its estimated value. As can be seen from these expressions, the estimation of β relies on that of τ. The estimator for θ will be influenced by the estimates of both τ and β. In our simulation study in Section 3 we will evaluate the performances of the estimators for τ, β and θ.

2.1 Maximum likelihood estimation

The density function of the variable of interest X can be obtained via its hazard function modeled in (1.2):

$$
\begin{aligned}
f(x) &= \lambda(x)\exp\left\{-\int_0^x \lambda(t)dt\right\} \\
&= \begin{cases} 0 & \text{if } \ x < 0 \\ \beta\exp(-\beta x) & \text{if } \ 0 \le x \le \tau \\ (\beta+\theta)\exp\{-\beta x - \theta(x-\tau)\} & \text{if } \ x > \tau. \end{cases}
\end{aligned}
$$

The expression for the survivor function $1 - F(x)$ is given by:

$$
1 - F(x) = \begin{cases} 1 & \text{if } \ x \le 0 \\ \exp(-\beta x) & \text{if } \ 0 < x \le \tau \\ \exp\{-\beta x - \theta(x-\tau)\} & \text{if } \ x > \tau. \end{cases}
$$

In case of non-informative censoring (meaning that the censoring distribution is independent of the parameters τ, β and θ), the appropriate likelihood based on the observations $T_1, \cdots, T_n$ is given by

$$L(\beta,\theta,\tau) = \prod_{i=1}^{n} f(T_i)^{\delta_i} \prod_{i=1}^{n} (1 - F(T_i))^{1-\delta_i}, \tag{2.1}$$

where the first product is over all uncensored observations, for which $\delta_i = 1$, and the second product is over all censored observations, for which $\delta_i = 0$. See Cox and Oakes (1984).

Let $X(t)$ denote the number of deaths observed up to time t:

$$X(t) = \sum_{i=1}^{n} I(T_i \leq t)\delta_i$$

and let n_u be the number of uncensored observations.

With the above expression for the density f and the expression for the survivor function $1 - F$ under model (1.2), the likelihood function can be written as

$$\begin{aligned} L(\beta,\theta,\tau) &= \beta^{X(\tau)}(\beta+\theta)^{n_u - X(\tau)} \\ &\quad \times \exp\left\{ -\beta \sum_{i=1}^{n} \min(T_i,\tau) - (\beta+\theta) \sum_{i=1}^{n} (T_i - \tau) I(T_i > \tau) \right\}, \end{aligned}$$

and the log-likelihood function is given by

$$\begin{aligned} \log L(\beta,\theta,\tau) &= X(\tau)\log\beta + (n_u - X(\tau))\log(\beta+\theta) \\ &\quad -\beta \sum_{i=1}^{n} \min(T_i,\tau) - (\beta+\theta) \sum_{i=1}^{n} (T_i - \tau) I(T_i > \tau). \end{aligned} \tag{2.2}$$

For fixed τ the maximum likelihood estimates of the parameters β and θ are given by

$$\begin{aligned} \widehat{\beta} &= \frac{X(\tau)}{\sum_{i=1}^{n} \min(T_i,\tau)} \\ \widehat{\theta} &= \frac{n_u - X(\tau)}{\sum_{i=1}^{n} (T_i - \tau) I(T_i > \tau)} - \widehat{\beta}. \end{aligned} \tag{2.3}$$

Substitution of these estimates for β and θ into the log-likelihood function in (2.2) leads to the objective function

$$\log L(\tau) = -n_u + X(\tau) \cdot \log \frac{X(\tau)}{\sum_{i=1}^{n} \min(T_i,\tau)}$$

$$+(n_u - X(\tau))\log \frac{n_u - X(\tau)}{\sum_{i=1}^{n}(T_i - \tau)I(T_i > \tau)}. \tag{2.4}$$

Now, maximizing this objective function over $\tau \in [\tau_0, \tau_1]$, with $0 \leq \tau_0 < \tau_1 < \infty$ and τ_1 strictly smaller than the largest uncensored observation, leads to the maximum likelihood estimator of τ, denoted by $\widehat{\tau}_{\text{MLE}}$. Distribution theory for $\widehat{\tau}_{\text{MLE}}$ has been studied by Yao (1986). He showed that $\widehat{\tau}_{\text{MLE}} - \tau = O_P(n^{-1})$ under certain conditions.

Note that this maximum likelihood estimation of τ involves the choice of an interval $[\tau_0, \tau_1]$. This interval can be chosen quite wide in practice, with the only important requirement that the largest uncensored observation should be falling above τ_1.

2.2 Estimation method of Chang, Chen and Hsiung

Chang, Chen and Hsiung (1994) assume that the unknown location point τ belongs to a certain known interval $[\tau_0, \tau_1]$ with $0 < \tau_0 \leq \tau \leq \tau_1 < \infty$. Their method is based on the following functional of the cumulative hazard function $\Lambda(\cdot)$:

$$Y_{\text{CCH}}(t) = \left[\frac{\Lambda(T) - \Lambda(t)}{T - t} - \frac{\Lambda(t) - \Lambda(0)}{t}\right] g\left(t(T - t)\right), \tag{2.5}$$

for $0 < t < T$, where $T > \tau_1$ and $g(x) = x^p, 0 \leq p \leq 1$. When $\theta > 0$, it can be shown that $Y_{\text{CCH}}(t)$ is increasing on the interval $[0, \tau]$ and decreasing on the interval $[\tau, T]$. This motivated Chang, Chen and Hsiung (1994) to consider the estimator

$$\widehat{\tau}_{\text{CCH}} = \inf\{t \in [\tau_0, \tau_1] : Y_n(t\pm) = \sup_{u \in [\tau_0, \tau_1]} Y_n(u)\}, \tag{2.6}$$

with $Y_n(t\pm)$ denoting the right-hand and left-hand limits, and where $Y_n(t)$ is the empirical version of (2.5) obtained by replacing the unknown cumulative hazard function by the Nelson-Aalen type estimator. Let $Z_1 < Z_2 < \ldots < Z_n$ be the order statistics corresponding to $T_1, T_2, \ldots, T_n$. Then the empirical cumulative hazard is given by

$$\Lambda_n(t) = \sum_{i: Z_i \leq t} \frac{\delta_i}{n - i + 1}. \tag{2.7}$$

See Aalen (1978). When $\theta < 0$, the function $Y_{\text{CCH}}(t)$ in (2.5) is decreasing on $[0, \tau]$ and increasing on $[\tau, T]$ and an estimator similar to $\widehat{\tau}$ in (2.6) can be obtained:

$$\widehat{\tau}_{\text{CCH}} = \inf\{t \in [\tau_0, \tau_1] : Y_n(t\pm) = \inf_{u \in [\tau_0, \tau_1]} Y_n(u)\},$$

Chang, Chen and Hsiung (1994) show that $\widehat{\tau}_{\mathrm{CCH}}$ defined in (2.6) is a consistent estimator of τ , with $\widehat{\tau}_{\mathrm{CCH}} - \tau = O_P(n^{-1})$.

Note that the estimation procedure of Chang, Chen and Hsiung (1994) involves the choice of the interval $[\tau_0, \tau_1]$ and the choice of the truncation point T. Although we use the same notation for the interval over which the candidate estimate for τ is searched for, the choice of this interval is far more restrictive in practice than in the maximum likelihood procedure, as will be seen from the simulations.

2.3 The least squares estimator

The estimation procedure of the previous section is based on the particular functional $Y_{\mathrm{CCH}}(t)$ as defined in (2.5). This functional is basically considering the differences of slopes of the cumulative hazard function. We base our least squares estimation procedure on the following function

$$Y(x) = \frac{\Lambda(x)}{x}, \tag{2.8}$$

which is the slope of the line joining the points $(0, \Lambda(0))$ and $(x, \Lambda(x))$ lying on the curve of the cumulative hazard function $\Lambda(\cdot)$. In a more intuitive sense, $Y(x)$ can be viewed as the *average hazard* at time x.

Under model (1.2) we have

$$Y(x) = \beta + \theta \left(1 - \frac{\tau}{x}\right) I_{\{\tau < x\}}. \tag{2.9}$$

Hence, under the assumed model, the function $Y(\cdot)$ has a simple structure: it remains constant up to time τ and from τ on starts increasing (in case $\theta > 0$) or decreasing (in case $\theta < 0$) as a function $1/x$.

Denote by $Y_n(x)$ the empirical version of $Y(x)$ obtained by replacing the unknown cumulative hazard function by the Nelson-Aalen estimator given in (2.7). The least squares procedure now consists of fitting, via least squares, a constant line up to τ and from τ on a function of the form $\beta + \theta(1 - \tau/x)$, through the 'data points' $(x_i, Y_n(x_i))$, $i = 1, \cdots, n_g$, where x_i's are grid points to be chosen. The splitting point τ which would give the best least squares fit would give the estimator for τ.

We now describe in detail the least squares estimation procedure. Let τ_0, τ_1 be the wide interval over which the estimated value of τ is sought for. Of course this interval should be such that it contains the true value of τ. We start by considering a fixed partition of this interval via n_g grid points denoted by $\tau_0 = x_1 < x_2 < .. < x_{n_g} = \tau_1$. Since our method consists of carrying out a least squares fit through the points $(x_i, Y_n(x_i))$, we need to make sure that we do have sufficient 'data points' for the fitting. To the left-hand side we fit a constant line, and hence this can be carried out even when there is only one 'data point'. We have to be more careful for the fitting of the function to the right-hand side, since this typically requires more 'data points'. In order to take care of this we include the option of carrying out the least squares fit on a subset of the initial grid points $x_1, x_2, \cdots, x_{n_g}$. Recall that n_u denotes the number of uncensored observations in the sample, and denote by $T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(n_u)}$ the ordered uncensored observations. We will search for the estimator of the location point τ over a possible reduced set of fixed grid points, such that we have a guaranteed portion of the actual data falling to the left of the first grid point in this reduced set, and to the right of the last grid point in the reduced set. Let α_1, α_2 be these proportions in the left and right tail respectively. The choice $\alpha_1 = 0$ is completely justified, but for α_2 one preferably would take a small positive value. Define $m_\ell = [\alpha_1 \times n_u]$ and $m_u = [(1-\alpha_2) \times n_u]$ and get $T_{(m_\ell)}$ and $T_{(m_u)}$. We then consider

$$
\begin{aligned}
x_\ell \quad &= \text{the smallest grid point in the initial grid that exceeds } T_{(m_\ell)} \\
&= \min\{x_j : j = 0, \cdots, n_g \text{ such that } x_j > T_{(m_\ell)}\} \\
x_u \quad &= \text{the largest grid point in the initial grid that does not exceed } T_{(m_u)} \\
&= \max\{x_j : j = 0, \cdots, n_g \text{ such that } x_j < T_{(m_u)}\}
\end{aligned}
$$

and search the candidate for τ, via least squares fitting, on the interval $I_{\alpha_1,\alpha_2} = [x_\ell, x_u]$. So, the set of grid points over which we search the candidate estimator changes with the sample, but the grid points itself are fixed from the start. In the simulation study we briefly illustrate the performance of the least squares procedure for various values of α_1 and α_2.

More precisely for each point $x_i \in I_{\alpha_1,\alpha_2}$ we carry out the following least squares fit:

minimize

$$L(x_i, \beta, \theta) = \sum_{j=1}^{i} [Y(x_j) - \beta]^2 + \sum_{j=i+1}^{n_g} [Y(x_j) - \beta - \theta(1 - x_{ij})]^2,$$

with respect to β and θ, where $x_{ij} = x_i/x_j$ for $x_j > x_i$, $j = 1, \cdots, n_g$.

Introducing the notation

$$\bar{Y}_{n_g} = \frac{1}{n_g} \sum_{i=1}^{n_g} Y_n(x_j) \qquad (2.10)$$

it is easily seen that for this fixed point x_i, the least squares estimates of β and θ are given by

$$\begin{aligned}
\widehat{\theta}(x_i) &= \frac{\sum_{j=i+1}^{n_g} Y_n(x_j)(1-x_{ij}) - \bar{Y}_{n_g} \sum_{j=i+1}^{n_g} (1-x_{ij})}{\sum_{j=i+1}^{n_g} (1-x_{ij})^2 - {n_g}^{-1}[\sum_{j=i+1}^{n_g} (1-x_{ij})]^2} \\
\widehat{\beta}(x_i) &= \bar{Y}_{n_g} - {n_g}^{-1}\widehat{\theta}(x_i) \textstyle\sum_{j=i+1}^{n_g}(1-x_{ij}).
\end{aligned} \tag{2.11}$$

These estimated values of θ and β are then used to find the fitted value of $Y(x_j)$, for any grid point x_j in the initial grid. Referring to (2.9), for the fixed point x_i, this fitted value is given by

$$Y_{n,x_i}(x_j) = \widehat{\beta}(x_i) + \widehat{\theta}(x_i)(1-x_{ij})I_{\{x_i<x_j\}}, \tag{2.12}$$

for all the grid points x_j.

The residual sum of squares resulting from estimating the unknown τ by the grid point x_i is then

$$\mathrm{RSS}(x_i) = \sum_{j=1}^{n_g}[Y_n(x_j) - Y_{n,x_i}(x_j)]^2. \tag{2.13}$$

The estimator of τ is now defined as

$$\widehat{\tau}_{\mathrm{LS},\alpha_1,\alpha_2} = \mathrm{argmin}_{x_i \in I_{\alpha_1,\alpha_2}} \mathrm{RSS}(x_i) \tag{2.14}$$

3 Performances of the estimators

3.1 Simulation comparison

The objective of this section is to investigate the performance of the least squares estimators (LSE) for τ, β and θ via simulations. This performance will be compared with that of the maximum likelihood estimator (MLE) described in Section 2.1 and the estimator studied by Chang, Chen and Hsiung (1994) from Section 2.2, referred to as the CCH estimator.

We simulated data from model (1.2) defined on the interval $[0,1]$, with $\tau=1$, $\beta=1$ and for the jump size θ we considered two values: $\theta=1$ and $\theta=0.7$, where the latter value of θ indicates a more difficult estimation task. For the censoring variable C we assumed a Uniform distribution on the interval $[0.4,1]$, which results in a censoring proportion of about 20%.

Since each of the three procedures involves some choices of intervals and/or other parameters, we simulated data for a number of set-ups. The most detailed simulation study was done for the newly-proposed LSE estimator. We carried out an extensive simulation study for each of these estimation procedures, but only a subset of them are reported here for space considerations. Other simulations led to similar conclusions. Simulations were done for various sample sizes, and the results we report on are based on 1000 simulations for each set-up.

Tables 1, 2 and 3 summarize the results obtained for the MLE, CCH and the LSE respectively for $\theta = 1$. Each table reports on the estimated value and its associated bias and standard deviation for each of the three parameters.

From the above tables we can draw some conclusions on the performances of each of the three estimation procedures:

* The choice of the interval $[\tau_0, \tau_1]$ as well as the choice of the upper limit T are quite important in the CCH procedure (Table 2 only reports on the choice of the upper limit, but see also Table 5 in this respect).

* Both the MLE and the LSE estimator are rather insensitive to the choice of the interval $[\tau_0, \tau_1]$.

* The choice of the 'left-out' proportions α_1 and α_2 are of minor importance in the LSE procedure. A small positive value for α_2 is recommendable.

Based on the tables we can also draw the following overall conclusions regarding a comparison of the estimation procedures:

◇ The CCH method has overall the smallest standard deviation, followed by the MLE. The LSE has typically a somewhat larger variance, which is not surprising since it involves in fact two estimation steps, namely the nonparametric estimation of the cumulative hazard function, and the least squares fitting.
  For estimation of β and θ though, the variances of the three estimators are quite comparable.

◇ The least squares procedure clearly outperforms the two other estimation methods in terms of bias, having a far smaller bias for estimation of τ, β and θ. The biggest gain is noted for the estimation of β and θ.

◇ The maximum likelihood estimator for θ performs quite badly (recall that the estimator of θ is influenced by the estimators of τ and β).

Tables 4, 5 and 6 report similarly on the performances of the three procedures when $\theta = 0.7$, representing a more difficult estimation task. Similar conclusions as the ones drawn from Tables 1—3 can be formulated here, with an even more pronounced gain in terms of bias for the LSE estimator.

For an additional evaluation we also produced kernel density estimates for the three parameters. In particular, we simulated 1000 samples of sample size $n = 300$ from model (1.2) with $\tau = \beta = 1$ and $\theta = 0.7$. The estimators for τ are searched for in the interval $[0.25, 1.75]$. The upper limit in the CCH method was chosen to be $T = 4.0$, and for the proportions in the least squares procedure we took $\alpha_1 = 0.0$ and $\alpha_2 = 0.03$. For each of the 1000 simulated samples, we obtained the estimated values for τ, β and θ. Then a kernel density estimate is obtained for each parameter estimator, using a Gaussian kernel and a bandwidth $h = 1.06Sn^{-1/5}$, where S^2 is the variance of the 1000 estimated parameter values. Figures 1—3 show these estimated distributions for the estimates of τ, β and θ respectively. The displayed density functions confirm our conclusions from the tables. Figure 1 clearly shows that he LSE estimator is far less biased, at a price of having a considerably bigger variance. The advantage of the LSE estimation procedure is most visible for the estimation of β and θ since, as can be seen from Figures 2 and 3, there is a significant gain in terms of bias, whereas the variance only slightly increases.

Since our tables to not directly report on the mean squared error of the estimators, we plot the square root of the mean squared errors (rmse) as a function of the sample size for the three estimation procedures. These plots are based on the calculated mean squared errors based on 1000 simulations for sample sizes $n = 15, 25, 50, 75, 100, 150, 200, 300, 500$ and 800. The parameter set-ups here were the same as those for Figures 1—3. The results are depicted in Figures 4—6 for τ, β and θ respectively. It is observed from these plots that the maximum likelihood method performs better than the other two estimators for estimating τ for most of the sample sizes and for large samples, the rmse of the LSE decreases more rapidly than the other two. For the estimation of β, the CCH and the MLE perform almost similarly and better than the LSE, and the difference being maximum for smaller sample sizes. Finally, it is observed that the LSE performs notably better than the others for estimating θ, uniformly for all sample sizes.

3.2 Real Data Examples

In this section we illustrate the use of the least squares procedure for two real data sets. We also apply the alternative methods for comparison purposes.

Lymphoma Data: The available data is the one on non-Hodgkin's lymphoma described and analyzed in Matthews, Farewell and Pyke (1985), which consists of the survival times (in months), defined as the time from diagnosis to death, of 31 individuals with advanced non-Hodgkin's lymphoma with clinical symptoms. Of these 11 are censored, including the largest observation which is treated as an uncensored data in the analysis. For completeness, the data is reproduced below, where an (*) indicates a censored observation.

2.5, 4.1, 4.6, 6.4, 6.7, 7.4, 7.6, 7.7, 7.8, 8.8, 13.3, 13.4, 18.3, 19.7, 21.9, 24.7, 27.5, 29.7, 30.1*, 32.9. 33.5, 35.4*, 37.7*, 40.9*, 42.6*, 45.4*, 48.5*, 48.9*, 60.4*, 64.4*, 66.4*.

In their study, Matthews, Farewell and Pyke (1985) find a very small significance level for the null hypothesis of constant hazard rate for the above data using $\tau_0 = 10$ and $\tau_1 = 35$. We also used the same $[\tau_0, \tau_1]$ interval for the estimation of τ for all the three methods that we compare. The function $Y_n(t)$, estimating $Y(t)$, is shown in Figure 7. The estimated values for the (τ, β, θ) triplet by the LS method was (22.09, 0.03, -0.011). In the LS method, since the censoring proportion is quite high in the right tail of the data, we set $\alpha_1 = 0$ and $\alpha_2 = 0.15$ in our analysis. The MLE of the parameters were (18.065, 0.028, -0.004) and the CCH estimator yielded (18.3, 0.029, -0.009). For the CCH estimator, $T = 50$ is used. All three estimation methods indicate a change point in the second half of the second year, and the MLE estimates a relatively smaller jump size than the other two.

Heart Transplant Data: The second data set we used is the well known Stanford heart transplant data, which is analyzed by several authors, including Miller and Halperin (1982), and Cox and Oakes (1984). Between October 1967 and February 1980, 184 out of 249 patients received heart transplant. The data we use consists of the survival times of these 184 patients after the transplant, among which 113 were censored. Loader (1991) also considers this data set for the application of the proposed change point inference methods. The survival times start from 0.0 and the largest observed survival time is 3695, which is censored, whereas the largest uncensored survival time is 2878. The time is measured in days which leaded to many ties in the sample. The empirical cumulative hazard function is depicted in Figure 8 and the kernel smoothed nonparametric estimator of the hazard rate function is given in Figure 9 with different choices of smoothing parameter. Once the estimator for τ is obtained, all three methods estimate β and θ by the same method and the estimation of τ is sensitive to the search interval $[\tau_0, \tau_1]$. We therefore report the results for a set of chosen intervals as given in Table 7. The lower end of the search interval is taken to be $\tau_0 = 20$ and the cases $\tau_1 = 120, 200, 500, 100$ are considered for the upper limit. In all cases 170 equally spaced grid points are considered. We observe that for all three methods, the estimators of β and θ are close to each other but they differ in the estimates of τ. They all capture the decreasing property of the hazard rate function by estimating a jump size of -0.003 to -0.004. The MLE and LSE methods for τ result in closer values as the search interval gets wider and in general LSE results in larger estimates for τ. The function $Y(y)$ in the LSE method is depicted in Figure 10 over the search interval. For the LSE estimator, the truncation proportion at the tails were taken to be $\alpha_1 = 0.03$ and $\alpha_2 = 0.05$, and other reasonable choices of these quantities did not make a difference in the estimation. We also note that for this particular data set, the CCH method sets the estimator of τ to the smallest grid point larger than the lower end point of the search interval, which was 22 in our choice. Therefore we report a single result for the CCH method.

References

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.*, **6**, 701–726.

Chang, I.-S., Chen, C.-H. and Hsiung, C.A. (1994). Estimation in change-point hazard rate models with random censorship. In *Change-point problems*, IMS Lecture Notes – Monograph Series, Volume 23, 78–92.

Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data.* London, Chapman and Hall.

Henderson, R. (1990). A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika*, **77**, 835–843.

Loader, C.R. (1991). Inference for a hazard rate change point. *Biometrika*, **78**, 749–757.

Matthews, D.E. and Farewell, V.T. (1982). On testing for a constant against a change-point alternative. *Biometrics*, **38**, 463–468.

Matthews, D.E., Farewell, V.T. and Pyke, R. (1985). Asymptotic score-statistic processes and tests for constant hazard rate against a change-point alternative. *Ann. Statist.*, **13**, 583–591.

Miller, R. and Halpern, J. (1982). Regression with censored data. *Biometrika*, **69**, 521–531.

Müller, H.G. and Wang, J.-L. (1990). Nonparametric analysis of changes in hazard rates for censored survival data: an alternative to change-point problems. *Biometrika*, **77**, 305–314.

Müller, H.-G. and Wang, J.-L. (1994). Change-point models for hazard functions. In *Change-point problems*, IMS Lecture Notes – Monograph Series, Volume 23, 224–241.

Nguyen, H.T., Rogers, G.S. and Walker, E.A. (1984). Estimation in change-point hazard rate models. *Biometrika*, **71**, 299–304.

Yao, Y.-C. (1986). Maximum likelihood estimation in hazard rate models with a change-point. *Comm. Statist., Series A*, **15**, 2455–2466.

Table 1: *Performances of the maximum likelihood estimators for the true values of the parameters* $\tau = 1$, $\beta = 1$ *and* $\theta = 1$.

		estimation of τ			estimation of β			estimation of θ		
			estimated			estimated			estimated	
$[\tau_0, \tau_1]$	n	$\widehat{\tau}$	bias	stdev.	$\widehat{\beta}$	bias	stdev.	$\widehat{\theta}$	bias	stdev.
$[0.25, 1.75]$	50	1.089	0.089	0.360	0.980	-0.020	0.241	14.431	13.431	77.664
	100	1.067	0.067	0.293	0.986	-0.014	0.157	1.637	0.637	1.705
	400	1.025	0.025	0.113	0.998	-0.002	0.068	1.104	0.104	0.255
$[0.5, 1.5]$	50	1.062	0.062	0.245	0.978	-0.022	0.209	9.258	8.258	11.763
	100	1.060	0.060	0.203	0.989	-0.011	0.144	1.479	0.479	0.835
	400	1.024	0.024	0.096	0.998	-0.002	0.068	1.101	0.101	0.244
$[0.75, 1.15]$	50	1.004	0.004	0.105	0.976	-0.024	0.196	1.519	0.519	1.041
	100	1.009	0.009	0.094	0.984	-0.016	0.134	1.287	0.287	0.569
	400	1.014	0.014	0.057	0.996	-0.004	0.066	1.085	0.085	0.233

Table 2: *Performances of the CCH estimators for the true values of the parameters* $\tau = 1$, $\beta = 1$ *and* $\theta = 1$.

			estimation of τ			estimation of β			estimation of θ		
				estimated			estimated			estimated	
$[\tau_0, \tau_1]$	T	n	$\widehat{\tau}$	bias	stdev.	$\widehat{\beta}$	bias	stdev.	$\widehat{\theta}$	bias	stdev.
$[0.25, 1.75]$	2	50	1.167	0.167	0.317	1.015	0.015	0.193	2.371	1.371	7.751
		100	1.150	0.150	0.262	1.017	0.017	0.140	1.575	0.575	1.047
		400	1.045	0.045	0.119	1.006	0.006	0.070	1.102	0.102	0.270
$[0.25, 1.75]$	3	50	0.995	-0.005	0.358	0.973	-0.027	0.191	1.432	0.432	0.920
		100	1.079	0.079	0.290	1.000	0.000	0.135	1.326	0.326	0.623
		400	1.133	0.133	0.231	1.023	0.023	0.078	1.156	0.156	0.316
$[0.25, 1.75]$	4	50	0.675	-0.325	0.345	0.927	-0.073	0.221	0.937	-0.063	0.665
		100	0.846	-0.154	0.289	0.961	-0.039	0.153	0.990	-0.010	0.495
		400	0.966	-0.034	0.150	0.992	-0.008	0.067	1.007	0.007	0.253

Table 3: *Performances of the least squares estimators for the true values of the parameters $\tau = 1$, $\beta = 1$ and $\theta = 1$.*

			estimation of τ			estimation of β			estimation of θ		
				estimated			estimated			estimated	
$[\tau_0, \tau_1]$	(α_1, α_2)	n	$\widehat{\tau}$	bias	stdev.	$\widehat{\beta}$	bias	stdev.	$\widehat{\theta}$	bias	stdev.
$[0.25, 1.75]$	(0,0)	50	1.020	0.020	0.405	0.997	-0.000	0.223	2.455	1.455	25.290
		100	0.995	-0.0053	0.351	1.003	0.003	0.221	1.229	0.229	1.260
		400	0.969	-0.031	0.155	1.004	0.004	0.073	0.961	-0.039	0.280
$[0.25, 1.75]$	(0,3)	50	0.990	0.010	0.400	0.996	-0.004	0.221	1.530	0.530	1.783
		100	0.990	-0.010	0.350	1.002	0.002	0.156	1.195	0.195	0.896
		400	0.969	-0.031	0.156	1.004	0.004	0.073	0.961	-0.039	0.280
$[0.25, 1.75]$	(5,5)	50	0.978	-0.022	0.395	0.996	-0.004	0.221	1.491	0.491	1.715
		100	0.984	-0.016	0.348	1.002	0.002	0.155	1.190	0.190	0.894
		400	0.969	-0.031	0.156	1.004	0.004	0.073	0.961	-0.039	0.280
$[0.25, 1.75]$	(3,7)	50	0.951	-0.049	0.382	0.994	-0.006	0.221	1.377	0.377	1.248
		100	0.958	-0.042	0.338	1.000	0.000	0.154	1.149	0.149	0.767
		400	0.969	-0.032	0.154	1.004	0.004	0.073	0.960	-0.040	0.279
$[0.5, 1.5]$	(0,0)	50	1.002	0.002	0.276	1.008	0.008	0.210	1.529	0.529	1.886
		100	0.999	-0.001	0.249	1.005	0.005	0.146	1.185	0.185	0.712
		400	0.975	-0.025	0.115	1.003	0.003	0.069	0.981	-0.019	0.266
$[0.5, 1.5]$	(0,3)	50	1.016	0.016	0.274	1.007	0.007	0.209	1.494	0.494	1.498
		100	0.999	-0.001	0.249	1.005	0.005	0.146	1.185	0.185	0.712
		400	0.975	-0.025	0.115	1.003	0.003	0.069	0.981	-0.019	0.266
$[0.5, 1.5]$	(5,5)	50	1.012	0.012	0.273	1.006	0.006	0.208	1.475	0.475	1.413
		100	0.999	-0.001	0.249	1.005	0.005	0.146	1.185	0.185	0.712
		400	0.975	-0.025	0.115	1.003	0.003	0.069	0.981	-0.019	0.266
$[0.5, 1.5]$	(3,7)	50	0.100	-0.000	0.270	1.005	0.005	0.208	1.429	0.429	1.248
		100	0.994	-0.006	0.247	1.005	0.005	0.144	1.177	0.177	0.692
		400	0.975	-0.025	0.115	1.003	0.003	0.069	0.981	-0.019	0.266
$[0.75, 1.15]$	(0,0)	50	0.987	-0.013	0.118	1.009	0.009	0.198	1.251	0.251	0.928
		100	0.986	-0.014	0.111	1.004	0.004	0.138	1.115	0.115	0.544
		400	0.989	-0.011	0.071	1.001	0.001	0.067	1.008	0.008	0.234
$[0.75, 1.15]$	(0,3)	50	0.987	-0.013	0.118	1.009	0.009	0.198	1.251	0.251	0.928
		100	0.986	-0.014	0.111	1.004	0.004	0.138	1.115	0.115	0.544
		400	0.989	-0.011	0.071	1.001	0.001	0.067	1.008	0.008	0.234
$[0.75, 1.15]$	(5,5)	50	0.987	-0.013	0.118	1.009	0.009	0.198	1.251	0.251	0.928
		100	0.986	-0.014	0.111	1.004	0.004	0.138	1.115	0.115	0.544
		400	0.989	-0.011	0.071	1.001	0.001	0.067	1.008	0.008	0.234
$[0.75, 1.15]$	(3,7)	50	0.987	-0.013	0.118	1.009	0.009	0.198	1.251	0.251	0.928
		100	0.986	-0.014	0.111	1.004	0.004	0.138	1.115	0.115	0.544
		400	0.989	-0.011	0.071	1.001	0.001	0.067	1.008	0.008	0.234

Table 4: *Performances of the maximum likelihood estimators for the true values of the parameters $\tau = 1$, $\beta = 1$ and $\theta = 0.7$.*

$[\tau_0, \tau_1]$	n	estimation of τ $\widehat{\tau}$	estimated bias	estimated stdev.	estimation of β $\widehat{\beta}$	estimated bias	estimated stdev.	estimation of θ $\widehat{\theta}$	estimated bias	estimated stdev.
$[0.25, 1.75]$	50	1.031	0.031	0.366	0.937	-0.063	0.255	2.978	2.278	26.167
	100	1.003	0.003	0.284	0.928	-0.072	0.157	1.240	0.540	0.920
	400	0.874	-0.126	0.103	0.910	-0.090	0.073	0.864	0.164	0.173
$[0.50, 1.50]$	50	1.012	0.012	0.242	0.927	-0.073	0.213	1.565	0.865	2.276
	100	0.980	-0.020	0.199	0.929	-0.071	0.146	1.118	0.418	0.590
	400	0.871	-0.129	0.090	0.914	-0.086	0.072	0.846	0.146	0.171

Table 5: *Performances of the CCH estimators for the true values of the parameters $\tau = 1$, $\beta = 1$ and $\theta = 0.7$. The upper limit is $T = 4$.*

$[\tau_0, \tau_1]$	n	estimation of τ $\widehat{\tau}$	estimated bias	estimated stdev.	estimation of β $\widehat{\beta}$	estimated bias	estimated stdev.	estimation of θ $\widehat{\theta}$	estimated bias	estimated stdev.
$[0.25, 1.75]$	50	0.749	-0.251	0.340	0.904	-0.096	0.209	0.905	0.205	0.606
	100	0.825	-0.175	0.270	0.914	-0.086	0.140	0.882	0.182	0.421
	400	0.886	-0.114	0.156	0.916	-0.084	0.074	0.850	0.150	0.185
$[0.50, 1.50]$	50	0.856	-0.144	0.203	0.924	-0.076	0.193	0.958	0.258	0.588
	100	0.876	-0.124	0.152	0.921	-0.079	0.135	0.904	0.204	0.388
	400	0.876	-0.124	0.115	0.915	-0.085	0.072	0.844	0.144	0.174

Table 6: *Performances of the least squares estimators for the true values of the parameters* $\tau = 1$, $\beta = 1$ *and* $\theta = 0.7$.

$[\tau_0, \tau_1]$	n	estimation of τ $\widehat{\tau}$	estimated bias	estimated stdev.	estimation of β $\widehat{\beta}$	estimated bias	estimated stdev.	estimation of θ $\widehat{\theta}$	estimated bias	estimated stdev.
$[0.25, 1.75]$	50	1.010	0.010	0.418	0.978	-0.022	0.216	1.243	0.543	2.012
	100	1.028	0.028	0.384	0.984	-0.016	0.154	0.965	0.265	0.830
	400	0.945	-0.055	0.187	0.974	-0.026	0.074	0.729	0.029	0.243
$[0.50, 1.50]$	50	1.019	0.019	0.293	0.976	-0.024	0.208	1.131	0.431	1.113
	100	0.984	-0.016	0.252	0.972	-0.028	0.145	0.900	0.200	0.536
	400	0.928	-0.072	0.136	0.965	-0.035	0.072	0.744	0.044	0.227

Table 7: *Results of MLE, LSE and CCH estimators for Stanford Heart Transplant data.*

METHOD	$[\tau_0, \tau_1]$	τ	β	θ
MLE	$[20, 120]$	68.2352	0.0044	-0.0039
	$[20, 200]$	68.7058	0.0044	-0.0039
	$[20, 500]$	70.8235	0.0043	-0.0038
	$[20, 1000]$	71.8824	0.0042	-0.0037
LSE	$[20, 120]$	80.58813	0.0038	-0.0034
	$[20, 200]$	81.41171	0.0038	-0.0034
	$[20, 500]$	79.29412	0.0039	-0.0037
	$[20, 1000]$	71.88235	0.0042	-0.0038
CCH	$[20, 120]$	22.00000	0.0034	-0.0027
	$[20, 200]$	22.00000	0.0034	-0.0027

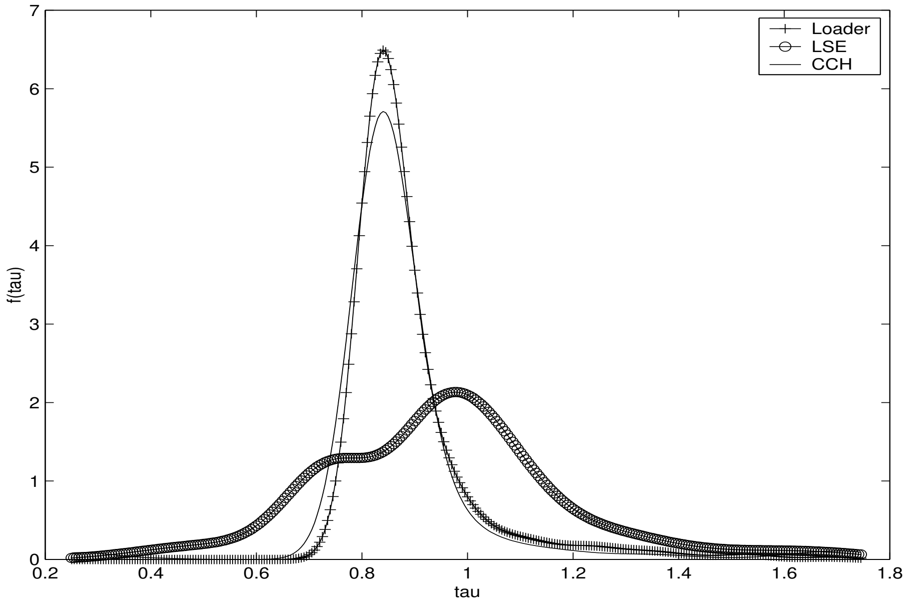

Figure 1: *Density estimates for the distribution of* $\widehat{\tau}$ *for the three estimation procedures, based on 1000 samples of size* $n = 300$.

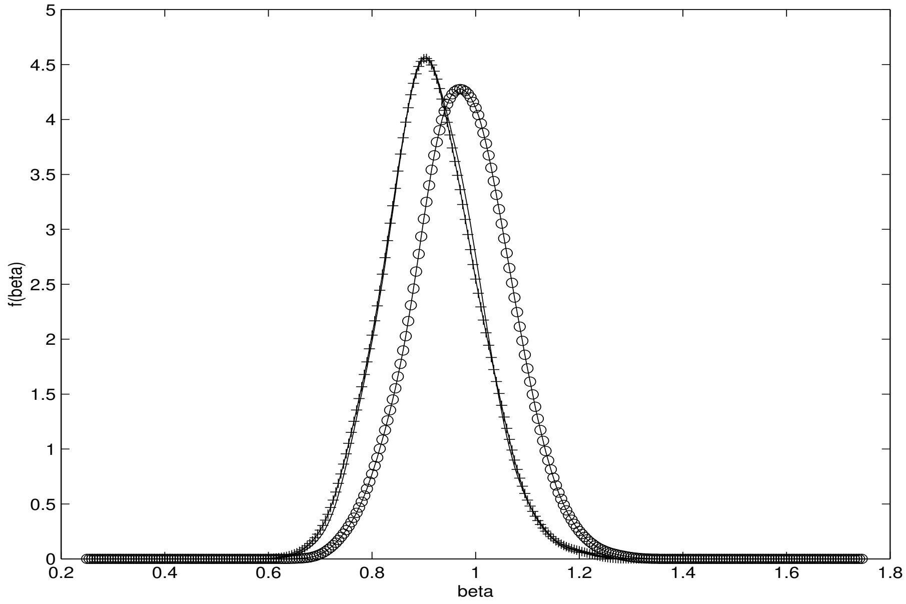

Figure 2: *Density estimates for the distribution of* $\widehat{\beta}$ *for the three estimation procedures, based on 1000 samples of size* $n = 300$.

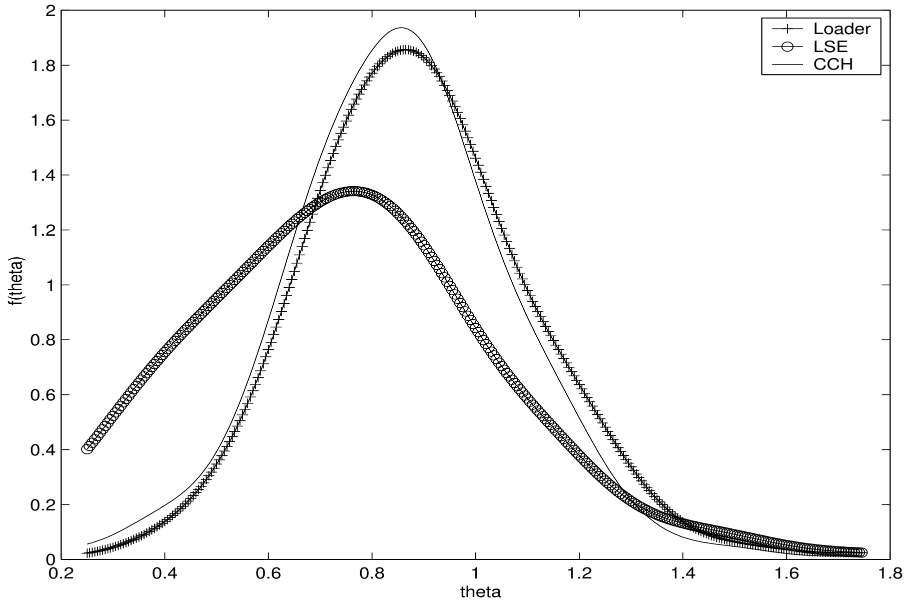

Figure 3: *Density estimates for the distribution of $\widehat{\theta}$ for the three estimation procedures, based on 1000 samples of size $n = 300$.*

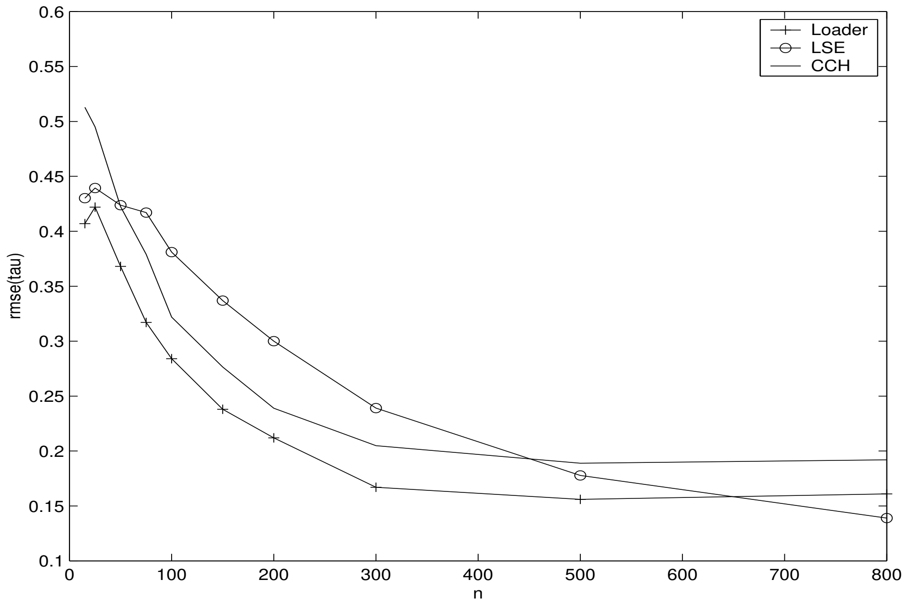

Figure 4: *Root mean squared errors of the three estimators for τ as a function of sample size, based on 1000 simulations.*

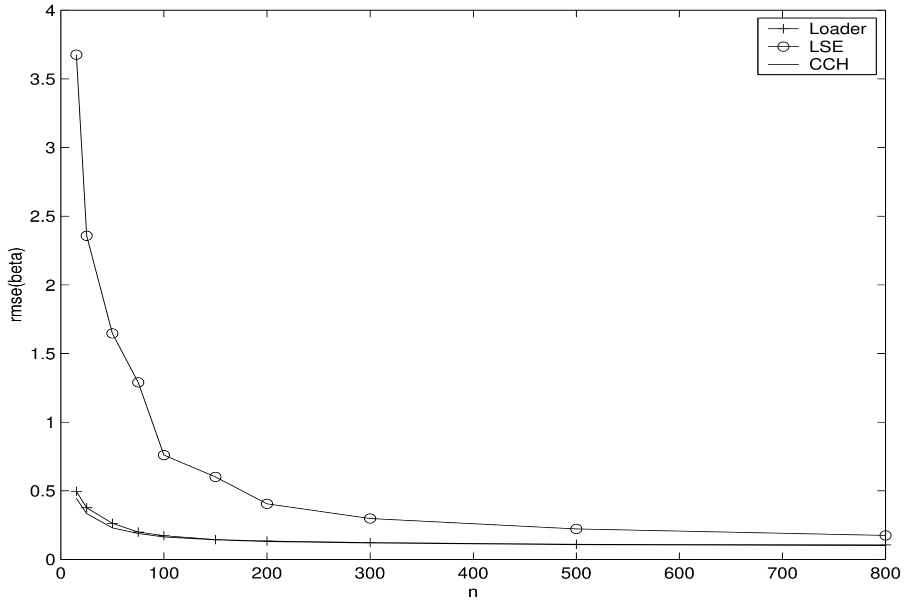

Figure 5: *Root mean squared errors of the three estimators for β as a function of sample size, based on 1000 simulations.*

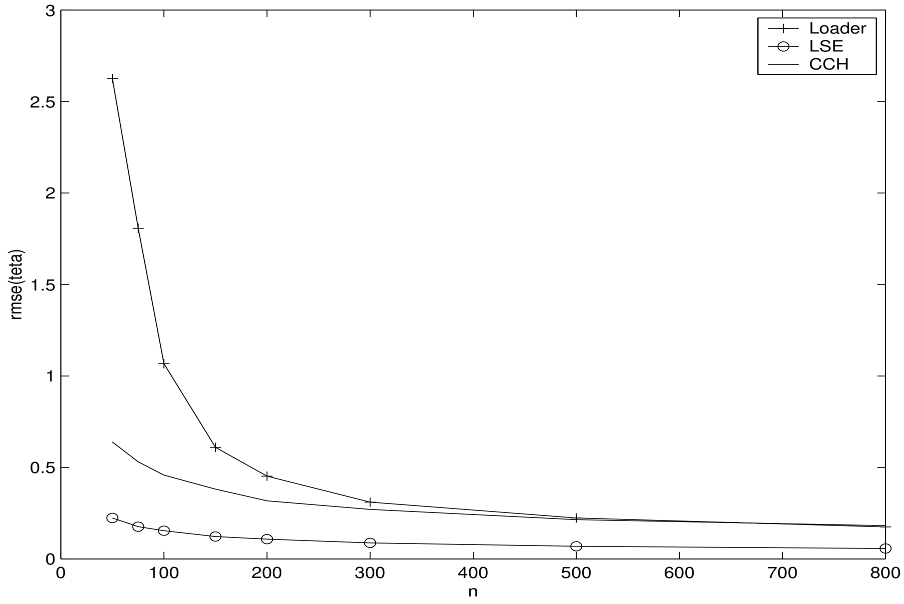

Figure 6: *Root mean squared errors of the three estimators for θ as a function of sample size, based on 1000 simulations.*

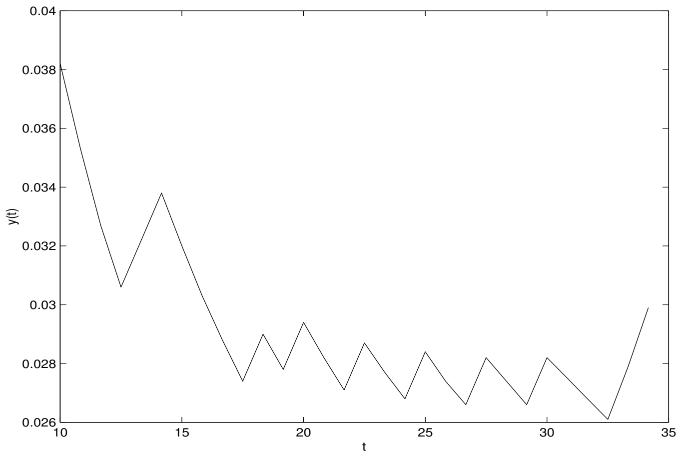

Figure 7: *Estimated $Y(t)$ for the LS estimator.*

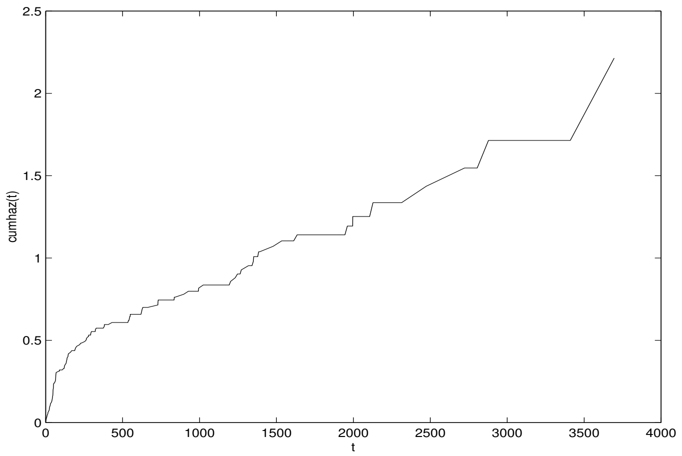

Figure 8: *Estimated cumulative hazard for the Stanford heart transplant data.*

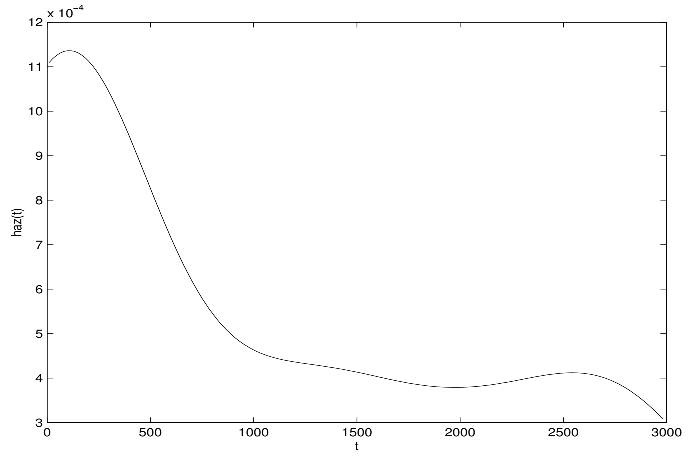

Figure 9: *Kernel estimator of the hazard rate function for the Stanford heart transplant data.*

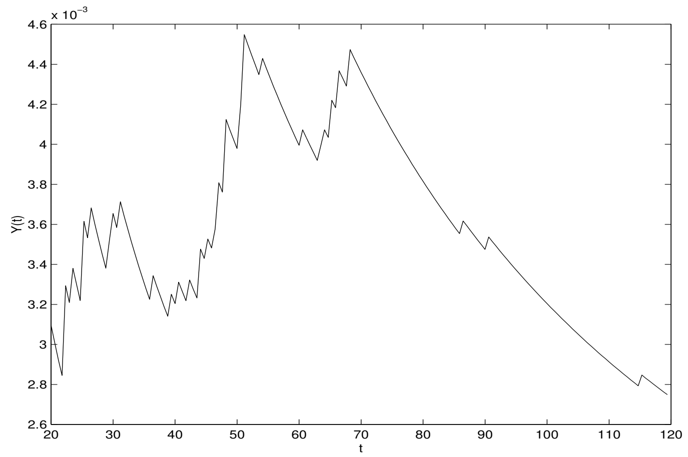

Figure 10: *Estimated* $Y(t)$ *of the Stanford heart transplant data for the LS estimator.*