Regression under Cox's Model for Recall-based Time-to-Event Data in Observational Studies

Technical Report No. ASU/2014/10
Dated 19 December, 2014

Sedigheh Mirzaei Salehabadi
and
Debasis Sengupta

Indian Statistical Institute

Applied Statistics Unit
Kolkata 700 108

Regression under Cox's Model for Recall-based Time-to-Event Data in Observational Studies

Sedigheh Mirzaei Salehabadi
Applied Statistical Unit, Indian Statistical Institute
and
Debasis Sengupta
Applied Statistical Unit, Indian Statistical Institute

Abstract

In some retrospective observational studies, the subject is asked to recall the age at a particular landmark event. The resulting data may be partially incomplete because of the inability of the subject to recall. Even though this type of incompleteness may be regarded as interval censoring, the censoring is likely to be informative. For such data, methods of parametric and non-parametric estimation of the survival function of the time-to-event have been provided by the authors. In this paper, we consider the problem of regression through Cox's proportional hazards model. While a partial likelihood is not available, a method of semi-parametric inference of the regression parameters as well as the baseline distribution is proposed. Monte Carlo simulations show reasonable performance of the regression parameters, compared to Cox estimators of the same parameters computed from the complete version of the data. The proposed method is illustrated through the analysis of data on age at menarche from a recent Anthropometric study of adolescent and young adult females in Kolkata, India.

Keywords: Informative censoring, Interval censoring, Proportional hazards model, Retrospective study, Turnbull estimator.

1 Introduction

Retrospective studies on a landmark event can produce dichotomous data on the current status of an individual (whether or not the event has occurred till the day of observation). From the perspective of the time to event, these data can be regarded as left or right censored. In some retrospective studies, the subject is asked to recall the time of the landmark event, in case it has already taken place. Such retrospective data can be incomplete because of the possibility that the time is forgotten. Sometimes the subject may be able to specify only a range for the time-to-event. For some other subjects, the event may be found not to have happened till the time of visit. Thus, data arising from this kind of retrospective studies are interval-censored. However, the chance of recall may depend on the time span between the occurrence of the event and the time of interview. For instance, between two females interviewed at the same age, the one having experienced menarche more recently may have a higher chance of recalling the date. Thus, the censoring mechanism in this set-up is likely to be informative. Mirzaei et al. (2015) and Mirzaei and Sengupta (2013) have proposed parametric and nonparametric methods of likelihood based inference, when the data are subjected to informative interval censoring of this type. They have shown theoretically as well as through simulation that estimators of survival function that ignore the informative nature of censoring can have large bias even when the sample size is large (Mirzaei et al., 2015). On the other hand, the problem of regression with the above type of censored data has not been addressed yet.

The proportional hazards regression model is widely used in the analysis of event time data with covariates. The method of analysis proposed by Cox (1972) can accommodate right-censored data which are usual in survival problems, and left-truncated data which arise when there are delayed entries in a cohort (Breslow et al., 1983). Other models which are used for more complex observation schemes include the accelerated failure time (AFT) model (Wei, 1992), the additive hazard regression model (Klein and Moeschberger, 2003), proportional odds ratio model (Dabrowska and Doksum, 1988) and so on (Vonta, 1996). There has also been some work on more general regression models for survival data, such as single index regression models (Chaudhuri, 2007) and models with random effect/frailty (Wienke, 2010).

Discrete-time regression models for right-truncated data have been developed and applied in the analysis of AIDS incidence and induction time distributions (Kalbfleisch and Lawless, 1991; Gross and Huber-Carol, 1992). Finkelstein (1986) and Finkelstein et al. (1993) discussed methods for fitting a discrete proportional hazards model for the case where the data are either interval-censored or right-truncated. In both cases, a score test was developed for testing the hypothesis of a zero regression coefficient. Tu et al. (1993) discussed a discrete proportional hazards model and an associated EM algorithm for data that are censored as well as truncated. Alioum and Commenges (1996) discussed a method for fitting a proportional hazards model for arbitrarily interval censored data. Their method assumes the censoring to be non-informative.

DeMasi et al. (1997) and Tanaka and Rao (2005) considered the regression problem for informatively censored data. Their model treats informative censoring as a type of risk in a competing risks setup, where the subject experiences two types of mutually exclusive events. This set-up is not meant to model the informative censoring naturally occurring in recall data.

In this paper, we consider regression under Cox's model for a special type of informatively censored data arising from uncertainly recalled event time in a retrospective study. In Section 2, we develop a semiparametric maximum likelihood estimator of the regression coefficients under the model. In Section 3, we report results of a simulation study of the performance of the proposed maximum likelihood estimator. Section 4 illustrates this method with data on menarcheal age of adolescent and young adult females, collected during the course of a project undertaken by the Indian Statistical Institute, Kolkata. Proofs of all the results are given in the Appendix.

2 Model and Inference

2.1 Model

Consider a subject having time of occurrence of the landmark event T_i, which is a single sample from a distribution F_i with density f_i and support $[t_{min}, t_{max}]$, for $i = 1, \ldots, n$. Let these subjects be interviewed at times $S_1, \ldots, S_n \in [t_{min}, t_{max}]$, respectively. Suppose U_i

is the unobservable time that the ith subject would take to forget the epoch of his/her landmark event. In order to minimize errors in recall, which has been recognized as a problematic issue with recall data (Rabe-Hesketh et al., 2001; Wen and Chen, 2014), we regarded a date as not recalled at all even when a range of possible dates was recalled. There are observable indicators δ_i and ε_i of the events $T_i \leq S_i$ and $U_i > S_i - T_i \geqslant 0$, respectively. We assume that $U_1, \ldots, U_n$ are samples from a distribution with distribution function π, and that these are independent of both T_i and S_i. It follows that, given S_i and T_i, the non-recall probability depends on the time elapsed since the landmark event as

$$P(\varepsilon_i = 0 | T_i = t, S_i = s) = \pi(s - t), \qquad s > t. \tag{1}$$

According to this model, the likelihood, conditional on the ages at interview, is

$$\prod_{i=1}^{n} [\bar{F}_i(S_i)]^{1-\delta_i} \left[\{f_i(T_i)(1 - \pi(S_i - T_i))\}^{\varepsilon_i} \left(\int_0^{S_i} f_i(u)\pi(S_i - u)du \right)^{1-\varepsilon_i} \right]^{\delta_i}. \tag{2}$$

Here the informativeness of the censoring mechanism is captured through the function π. If π is a constant, then the likelihood (2) becomes a multiple of the likelihood for non-informatively left- or right censored data with possibility of no censoring. As a further special case, if $\pi = 1$, then the likelihood (2) simplifies to the likelihood for dichotomous data. If $\pi = 0$, i.e., there is perfect recall with probability 1, then $\varepsilon_i = 1$ for all i such that $\delta_i = 1$, and the likelihood reduces to that for right-censored data.

Let Z_i be the r-dimensional vector of covariates, assumed to be independent of both S_i and U_i. Note that the distribution of T_i would depend on Z_i. Under Cox's relative risk regression model, the probability of the individual i, with covariate vector Z_i, having the event after time t is

$$\bar{F}_i(t) = [\bar{F}_0(t)]^{\exp(\beta^T Z_i)}, \tag{3}$$

where $\bar{F}_0$ is the baseline survival function, assumed to have a density f.

2.2 Identifiability

Before embarking on developing a method of estimation, we need to check the identifiability of β, F_0 and π. By substituting (3) in the likelihood (2), after dropping the subscript i for simplicity and following Theorem 1 of Mirzaei et al. (2015), one can show that a typical factor in the product likelihood is equal to the conditional density of the observable vector (V, δ), given S and Z, where $V = (S-T)\varepsilon$. The conditional density is written alternatively as

$$h(v,\delta|s,z;\beta) = \begin{cases} \bar{F}_0(s)^{\exp(\beta^T z)} & \text{if } v=0 \text{ and } \delta=0, \\ \int_0^s -\frac{d}{du}\left(\bar{F}_0(u)^{\exp(\beta^T z)}\right)\pi(s-u)du & \text{if } v=0 \text{ and } \delta=1, \\ -\frac{d}{dv}\left(\bar{F}_0(s-v)^{\exp(\beta^T z)}\right)(1-\pi(v)) & \text{if } v>0 \text{ and } \delta=1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Among the unknown parameters β, F_0 and π, the interest lies mainly in β, and possibly in F_0. We address the question as to whether β, F_0 and π are identifiable from h, in the next theorem.

Theorem 1 *Suppose, for any number τ in the support of F_0, the model* (4) *holds for some $s > \tau$, and for $z = 0$ as well as for any r linearly independent values of the vector z, where r is the dimension of z. Then the parameters β, F_0 and π are identifiable from h under this model.*

We now proceed with the estimation problem, after assuming that the distribution of the covariate vector ensures the identifiability of all the unknown parameters.

2.3 Piecewise Constant Non-recall Probability

The integral contained in the likelihood (2) makes it difficult to maximize. For the sake of mathematical tractability, we now assume that π is a piecewise constant function of the form

$$\pi(x) = b_1 I(x_1 < x \leq x_2) + b_2 I(x_2 < x \leq x_3) + \ldots + b_k I(x_k < x < \infty), \tag{5}$$

where $0 = x_1 < x_2 < \cdots < x_k$; $0 < b_1, b_2, \ldots, b_k \leq 1$. Note that it is possible to arrange the parameters $b_1, b_2, \ldots, b_k$ to be in increasing order, so that π is a non-decreasing function.

In view of (5), the likelihood (2) reduces to

$$L = \prod_{i=1}^{n} [\bar{F}_i(S_i)]^{1-\delta_i} \left[\left\{ f_i(T_i) \left(1 - \sum_{l=1}^{k} b_l I\big(W_{l+1}(S_i) < T_i \le W_l(S_i)\big) \right) \right\}^{\varepsilon_i} \right.$$
$$\left. \cdot \left\{ \sum_{l=1}^{k} b_l \big(F_i(W_l(S_i)) - F_i(W_{l+1}(S_i))\big) \right\}^{1-\varepsilon_i} \right]^{\delta_i}, \tag{6}$$

where $W_l(S_i) = (S_i - x_l) \vee t_{min}$ for $l = 1, \ldots, k$ and $W_{k+1}(S_i) = t_{min}, i = 1, 2, \ldots, n$. Note that

$$t_{min} = W_{k+1}(S_i) \le W_k(S_i) \le W_{k-1}(S_i) \le \cdots \le W_1(S_i). \tag{7}$$

Depending on the value of S_i, some of the above inequalities may in fact be equalities. Specifically, if l is an index such that $S_i - x_{l+1} \le t_{min} < S_i - x_l$ then $t_{min} = W_{k+1}(S_i) = \cdots = W_{l+1}(S_i)$. The remaining inequalities would be strict.

The likelihood (6) can be rewritten as

$$L = \prod_{i=1}^{n} [\bar{F}_0(S_i)^{\exp(\beta^T Z_i)}]^{1-\delta_i} \left[\left\{ \left(\bar{F}_0(T_i-)^{\exp(\beta^T Z_i)} - \bar{F}_0(T_i)^{\exp(\beta^T Z_i)} \right) \right. \right.$$
$$\left. \cdot \left(1 - \sum_{l=1}^{k} b_l I\big(W_{l+1}(S_i) < T_i \le W_l(S_i)\big) \right) \right\}^{\varepsilon_i}$$
$$\left. \cdot \left\{ \sum_{l=1}^{k} b_l \big(\bar{F}_0(W_{l+1}(S_i))^{\exp(\beta^T Z_i)} - \bar{F}_0(W_l(S_i))^{\exp(\beta^T Z_i)} \big) \right\}^{1-\varepsilon_i} \right]^{\delta_i}, \tag{8}$$

which is mathematically tractable.

2.4 Maximum Likelihood Estimation

Following the Turnbull (1976) (non-information interval censoring) and Mirzaei and Sengupta (2013) (informative interval censoring), to maximize the likelihood (8), we need to look for the related intervals and maximize the likelihood with respect to the probability masses assigned to those intervals. Assume that the parameters k and $x_1, x_2, \ldots, x_k$ of the function π are known. The likelihood (8) involves probabilities assigned to intervals of the type $[t, t_{max}]$ and $(t, t_{max}]$, as per the baseline probability distribution. Since these intervals have overlap, we express them as unions of some disjoint intervals. Let $\mathcal{I}_1$, $\mathcal{I}_2$ and

$\mathcal{I}_3$ be sets of indices i (between 1 and n) that satisfy the conditions $\delta_i = 0$, $\delta_i\varepsilon_i = 1$ and $\delta_i(1-\varepsilon_i) = 1$, respectively. Consider the intervals

$$\begin{aligned}
A_i &= (S_i, t_{max}] && \text{for } i \in \mathcal{I}_1; \\
A_i &= [T_i, t_{max}] && \text{for } i \in \mathcal{I}_2; \\
A_i' &= (T_i, t_{max}] && \text{for } i \in \mathcal{I}_2; \\
A_{il} &= \begin{cases} (W_l(S_i), t_{max}], & l = 1, \ldots, k, \\ [W_l(S_i), t_{max}], & l = k+1, \end{cases} && \text{for } i \in \mathcal{I}_2 \cup \mathcal{I}_3.
\end{aligned} \tag{9}$$

and the sets

$$\begin{aligned}
\mathcal{A}_1 &= \{A_i : \ \ i \in \mathcal{I}_1\}; \\
\mathcal{A}_2 &= \{A_i \setminus A_i' : \ \ i \in \mathcal{I}_2\}; \\
\mathcal{A}_3 &= \{A_i' : \ \ i \in \mathcal{I}_2\}; \\
\mathcal{A}_4 &= \{A_{i(l+1)} \setminus A_{il} : \ \ 1 \le l \le k \text{ and } i \in \mathcal{I}_3\}.
\end{aligned} \tag{10}$$

As the baseline distribution is absolutely continuous, the elements of $\mathcal{A}_2$ and $\mathcal{A}_3$ are all distinct with probability 1. Let n_2 be the number of elements of $\mathcal{I}_2$. We arrange the singleton elements of $\mathcal{A}_2$ in increasing order, and denote them $B_1, B_2, \ldots, B_{n_2}$. We also arrange the elements of $\mathcal{A}_3$ in the corresponding order and denote them as $B_{n_2+1}, B_{n_2+2}, \ldots, B_{2n_2}$. We then collect the unique elements of $\mathcal{A}_1 \cup \mathcal{A}_4$ that are distinct from $B_1, B_2, \ldots, B_{2n_2}$, and denote them as $B_{2n_2+1}, B_{2n_2+2}, \ldots, B_M$. Observe that the collection $B_1, B_2, \ldots, B_M$ consist of the distinct elements of $\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$, arranged in a particular order. Denote the non-empty subsets of the index set $\{1, 2, \ldots, M\}$ by $s_1, s_2, \ldots, s_{2^M-1}$. Define

$$I_r = \left\{ \bigcap_{i \in s_r} B_i \right\} \bigcap \left\{ \bigcap_{i \notin s_r} B_i^c \right\} \qquad \text{for } r = 1, 2, \ldots, 2^M - 1. \tag{11}$$

Some of the I_r's may be empty sets, denoted here by ϕ. Let

$$\mathcal{C} = \{s_r : \ I_r \ne \phi, \ 1 \le r \le 2^M - 1\}. \tag{12}$$

It can be verified that all the non-empty I_r's are distinct and disjoint. Let $\mathcal{A}$ be the class of all sets I_r such that $s_r \in \mathcal{C}$.

Note that each of the intervals $B_1, \ldots, B_M$ is a union of disjoint sets that are members of $\mathcal{A}$. For any Borel set A, suppose $P_0(A)$ is the probability assigned to A as per the baseline probability distribution corresponding to the survival function $\bar{F}_0$. Let $p_r = P_0(I_r)$, for $I_r \in \mathcal{A}$. Then the likelihood (8) reduces to

$$
\begin{aligned}
L = \prod_{i \in \mathcal{I}_1} \left(\sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}}} p_r \right)^{\exp(\beta^T Z_i)} & \times \prod_{i \in \mathcal{I}_2} \left(1 - \sum_{l=1}^{k} b_l I\big(T_i \in A_{i(l+1)} \backslash A_{il}\big) \right) \\
& \cdot \left[\left(\sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}}} p_r \right)^{\exp(\beta^T Z_i)} - \left(\sum_{\substack{r: I_r \subseteq A_i' \\ s_r \in \mathcal{C}}} p_r \right)^{\exp(\beta^T Z_i)} \right] \\
\times \prod_{i \in \mathcal{I}_3} \left[\sum_{l=1}^{k} b_l \left\{ \left(\sum_{\substack{r: I_r \subseteq A_{i(l+1)} \\ s_r \in \mathcal{C}}} p_r \right)^{\exp(\beta^T Z_i)} - \left(\sum_{\substack{r: I_r \subseteq A_{il} \\ s_r \in \mathcal{C}}} p_r \right)^{\exp(\beta^T Z_i)} \right\} \right]. \qquad (13)
\end{aligned}
$$

Thus, maximizing the likelihood (8) is equivalent to maximizing the likelihood (13) with respect to β and the set of p_r's with $s_r \in \mathcal{C}$. The latter parameters are nuisance parameters when the main objective is to estimate β. The number of nuisance parameters can be very high. This problem can be simplified if it can be shown algebraically that some of the estimated p_r's are zero. With this goal, we consider the following subsets of $\mathcal{C}$.

$$
\begin{aligned}
\mathcal{C}_1 &= \{s : s \in \mathcal{C};\ \text{there is another element } s' \in \mathcal{C}, \text{ such that } s \subset s'\}, \\
\mathcal{C}_2 &= \{s : s \in \mathcal{C};\ \text{there is another element } s' \in \mathcal{C}, \text{ such that} \\
&\qquad s' \backslash (s \cap s') \text{ consists of a singleton } j \text{ and } s \backslash (s \cap s') = \{j + n_2\}\}, \\
\mathcal{C}_0 &= \mathcal{C} \backslash (\mathcal{C}_1 \cup \mathcal{C}_2). \qquad (14)
\end{aligned}
$$

The next result shows that the maximization of the likelihood can be restricted to $\mathcal{C}_0$.

Theorem 2 *For a fixed value of β, maximizing the likelihood* (13) *with respect to p_r for*

$s_r \in \mathcal{C}$ is almost surely equivalent to maximizing it with respect to p_r for $s_r \in \mathcal{C}_0$, i.e.,

$$\max_{p_r : p_r \in [0,1], \sum_{s_r \in \mathcal{C}} p_r = 1} L \quad = \quad \max_{p_r : p_r \in [0,1], \sum_{s_r \in \mathcal{C}_0} p_r = 1} L.$$

Let us relabel the intervals I_j, $s_j \in \mathcal{C}_0$, by $J_1, J_2, \ldots, J_v$. Further, let $q_j = P(J_j)$ for $j = 1, 2, \ldots, v$ and $\eta = (b_1, b_2, \cdots, b_k)$. Theorem 2 implies that maximizing the likelihood (13) is almost surely equivalent to maximizing

$$\begin{aligned} L(q_1, \ldots, q_v, \eta, \beta) = \prod_{i \in \mathcal{I}_1} \left(\sum_{j : J_j \subseteq A_i} q_j \right)^{\exp(\beta^T Z_i)} & \times \prod_{i \in \mathcal{I}_2} \left(1 - \sum_{l=1}^{k} b_l I\big(T_i \in A_{i(l+1)} \backslash A_{il}\big) \right) \\ & \cdot \left[\left(\sum_{j : J_j \subseteq A_i} q_j \right)^{\exp(\beta^T Z_i)} - \left(\sum_{j : J_j \subseteq A_i'} q_j \right)^{\exp(\beta^T Z_i)} \right] \\ \times \prod_{i \in \mathcal{I}_3} \left[\sum_{l=1}^{k} b_l \left\{ \left(\sum_{j : J_j \subseteq A_{i(l+1)}} q_j \right)^{\exp(\beta^T Z_i)} \right. \right. & \left. \left. - \left(\sum_{j : J_j \subseteq A_{il}} q_j \right)^{\exp(\beta^T Z_i)} \right\} \right]. \end{aligned} \tag{15}$$

with respect to $q_1, q_2, \ldots, q_v$, η and β, subject to the restriction $\sum_{j=1}^{v} q_j = 1$.

In order to maximize the likelihood (15), we need to identify the sets J_j, $j = 1, \ldots, v$, that is, the intervals I_j, $s_j \in \mathcal{C}_0$, defined through (11) and (14). This identification involves elaborate combinatorial calculations.

The following section shows that one can restrict the domain, to maximize the likelihood with respect to probability masses assigned to exact recall point without losing much information.

2.5 Approximate MLE

Let $\mathcal{A}_0 = \{J_1, J_2, \ldots, J_v\}$, and $\mathcal{A}_2 = \{\{T_i\},\ i \in \mathcal{I}_2\}$ as already defined in (10). Further, let n_i be the cardinality of $\mathcal{I}_i$, $i = 1, 2, 3$. The task of maximizing the likelihood (15) can be simpler for large n_2, as the following result shows.

Theorem 3 *The set $\mathcal{A}_2$ is contained in the set $\mathcal{A}_0$ almost surely. Further, if the inspection times take values from a finite set and the range of values of $\beta^T Z_i$ in (15) is bounded, then* *the probability of $\mathcal{A}_0$ being equal to $\mathcal{A}_2$ goes to one as $n_2 \to \infty$.*

One can form a computationally simpler estimator on the basis of Theorem 3. According to this theorem, the maximum likelihood estimator has mass only at points of exact recall of the event, when n_2 is large. In such a case, the likelihood (15) involves J_j's that are singletons only. Therefore, irrespective of the value of n_2, one can maximize (15) with respect to point masses corresponding to the the times of exact recall.

Formally, let $t_1, \ldots, t_{n_2}$ be the ordered set of distinct ages at event that have been exactly recalled, and $q_1^*, \ldots, q_{n_2}^*$ be the probability masses allocated to them. Maximizing the likelihood (15), subject to the constraint that $q_j = 0$ whenever $J_j \notin \mathcal{A}_2$, is equivalent to maximizing the following approximate likelihood subject to $\sum_{j=1}^{n_2} q_j^* = 1$ and $q_j^* \geq 0$:

$$L_a(q_1^*, \ldots, q_{n_2}^*, \eta, \beta) = \prod_{i \in \mathcal{I}_1} \left(\sum_{j:t_j \geqslant t_{m_i}} q_j^* \right)^{\exp(\beta^T Z_i)} \times$$
$$\prod_{i \in \mathcal{I}_2} \left(1 - \sum_{l=1}^{k} b_l I\big(T_i \in A_{i(l+1)} \backslash A_{il})\big) \right) \cdot \left[\left(\sum_{j:t_j \geqslant t_{m_i}} q_j^* \right)^{\exp(\beta^T Z_i)} - \left(\sum_{j:t_j > t_{m_i}} q_j^* \right)^{\exp(\beta^T Z_i)} \right]$$
$$\times \prod_{i \in \mathcal{I}_3} \left[\sum_{l=1}^{k} b_l \left\{ \left(\sum_{j:t_j \geqslant t_{m_{i(l+1)}}} q_j^* \right)^{\exp(\beta^T Z_i)} - \left(\sum_{j:t_j \geqslant t_{m_{il}}} q_j^* \right)^{\exp(\beta^T Z_i)} \right\} \right]. \qquad (16)$$

where $m_i = \inf\{j : t_j \in A_i\}$ for $i \in \mathcal{I}_1 \cup \mathcal{I}_2$, and, $m_{il} = \inf\{j : t_j \in A_{il}\}$, $l = 1, 2, \ldots, k$ for $i \in \mathcal{I}_3$.

In order to remove the range restriction on the parameters $q_1^*, \ldots, q_{n_2}^*$, we use the reparametrization $\gamma_d = \log(-\log(\sum_{j:t_j \geq t_d} q_j^*))$, $d = 1, 2, \ldots, n_2$. Thus, the approximate likelihood (16) can be expressed as

$$L_a(\gamma, \eta, \beta) = \prod_{i \in \mathcal{I}_1} \left(e^{-e^{Z_i\beta + \gamma_{m_i}}} \right) \times \prod_{i \in \mathcal{I}_2} \left(1 - \sum_{l=1}^{k} b_l I\big(T_i \in A_{i(l+1)} \backslash A_{il})\big) \right)$$
$$\cdot \left[\big(e^{-e^{Z_i\beta + \gamma_{m_i}}}\big) - \big(e^{-e^{Z_i\beta + \gamma_{m_i+1}}}\big) \right]$$
$$\times \prod_{i \in \mathcal{I}_3} \left[\sum_{l=1}^{k} b_l \left\{ \big(e^{-e^{Z_i\beta + \gamma_{m_{i(l+1)}}}}\big) - \big(e^{-e^{Z_i\beta + \gamma_{m_{il}}}}\big) \right\} \right]. \qquad (17)$$

where $\gamma = (\gamma_1, \ldots, \gamma_{n_2+1})$ and $\gamma_{n_2+1} = \infty$. The log of the above expression simplifies to

$$\ell_a(\gamma, \eta, \beta) = \sum_{i=1}^{n} \log \left[\sum_{j=1}^{n_2} \alpha_{ij} g_{ij} \right], \tag{16}$$

where

$$\alpha_{ij} = \begin{cases} I(J_j \subseteq A_i) & \text{if } i \in \mathcal{I}_1, \\ \left(1 - \sum_{l=1}^{k} b_l \cdot I(T_i \in A_{i(l+1)} \backslash A_{il})\right) \cdot I(J_j \subseteq A_i \backslash A_i') & \text{if } i \in \mathcal{I}_2, \\ \sum_{l=1}^{k} b_l \cdot I(J_j \subseteq A_{i(l+1)} \backslash A_{il}) & \text{if } i \in \mathcal{I}_3, \end{cases} \tag{17}$$

and

$$g_{ij} = e^{-e^{Z_i^T \beta + \gamma_j}} - e^{-e^{Z_i^T \beta + \gamma_{j+1}}}.$$

We obtain the approximate maximum likelihood estimate (AMLE) of the parameters γ, η and β by maximizing the above approximate log-likelihood. Note that α_{ij} depends only on η and g_{ij} depends only on β and γ. Further,

$$\begin{aligned}
\frac{\partial \alpha_{ij}}{\partial b_l} &= -I(T_i \in A_{i(l+1)} \backslash A_{il}) \cdot I(J_j \subseteq A_i \backslash A_i') \cdot I(i \in \ell_2) \\
&\quad + I(J_j \subseteq A_{i(l+1)} \backslash A_{il}) \cdot I(i \in \ell_3) \\
&= \xi_{ijl} \text{ (say)}, \\
\frac{\partial g_{it}}{\partial \gamma_j} &= -\delta_{tj} e^{Z_i^T \beta + \gamma_j} e^{-e^{Z_i^T \beta + \gamma_j}} + \delta_{(t+1)j} e^{Z_i^T \beta + \gamma_j} e^{-e^{Z_i^T \beta + \gamma_j}} \\
&= (\delta_{tj} - \delta_{(t+1)j}) h_{ij}, \\
\text{where } h_{ij} &= -e^{Z_i^T \beta + \gamma_j} e^{-e^{Z_i^T \beta + \gamma_j}} = -e^{Z_i^T \beta + \gamma_j - e^{Z_i^T \beta + \gamma_j}}, \\
\frac{\partial g_{it}}{\partial \beta} &= -Z_i e^{Z_i^T \beta + \gamma_t} e^{-e^{Z_i^T \beta + \gamma_t}} + Z_i e^{Z_i^T \beta + \gamma_{t+1}} e^{-e^{Z_i^T \beta + \gamma_{t+1}}} \\
&= Z_i (h_{it} - h_{i(t+1)}).
\end{aligned}$$

It follows that

$$
\begin{aligned}
\frac{\partial \ell_a}{\partial b_l} &= \sum_{i=1}^{n} \frac{\sum_{t=1}^{n_2} \xi_{itl} g_{it}}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}}, \\
\frac{\partial \ell_a}{\partial \gamma_j} &= \sum_{i=1}^{n} \frac{\sum_{t=1}^{n_2} \alpha_{it}(\delta_{tj} - \delta_{(t+1)j}) h_{ij}}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} \quad = \quad \sum_{i=1}^{n} \frac{(\alpha_{ij} - \alpha_{i(j-1)}) h_{ij}}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} \quad = \quad \sum_{i=1}^{n} \mu_{ij} h_{ij},
\end{aligned}
$$

where $\alpha_{i0} = 0$, $\mu_{ij} = (\alpha_{ij} - \alpha_{i(j-1)}) / \sum_{t=1}^{n_2} \alpha_{it} g_{it}$. Also,

$$
\frac{\partial \ell_a}{\partial \beta} = \sum_{i=1}^{n} Z_i \frac{\sum_{t=1}^{n_2} \alpha_{it}(h_{it} - h_{i(t+1)})}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} = \sum_{i=1}^{n} Z_i \frac{\sum_{t=1}^{n_2} (\alpha_{it} - \alpha_{i(t-1)}) h_{it}}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} = \sum_{i=1}^{n} Z_i \sum_{t=1}^{n_2} \mu_{it} h_{it}.
$$

We now compute the second derivatives.

$$
\begin{aligned}
\frac{\partial^2 \ell_a}{\partial b_l \partial b_{l'}} &= -\sum_{i=1}^{n} \frac{\left(\sum_{t=1}^{n_2} \xi_{itl} g_{it}\right) \left(\sum_{t=1}^{n_2} \frac{\partial \alpha_{ij}}{\partial b_{l'}} g_{it}\right)}{\left(\sum_{t=1}^{n_2} \alpha_{it} g_{it}\right)^2} \quad = \quad -\sum_{i=1}^{n} \frac{\left(\sum_{t=1}^{n_2} \xi_{itl} g_{it}\right) \left(\sum_{t=1}^{n_2} \xi_{itl'} g_{it}\right)}{\left(\sum_{t=1}^{n_2} \alpha_{it} g_{it}\right)^2}, \\
\frac{\partial^2 \ell_a}{\partial \gamma_j \partial \gamma_{j'}} &= \sum_{i=1}^{n} \left(\frac{\partial \mu_{ij}}{\partial \gamma_{j'}} h_{ij} + \mu_{ij} \frac{\partial h_{ij}}{\partial \gamma_{j'}} \right) \\
&= \sum_{i=1}^{n} \frac{\left\{(\alpha_{i(j-1)} - \alpha_{ij})\right\} \left\{(\alpha_{ij'} - \alpha_{i(j'-1)})\right\} h_{ij'}}{\left(\sum_{t=1}^{n_2} \alpha_{it} g_{it}\right)^2} h_{ij} + \sum_{i=1}^{n} \mu_{ij} \delta_{jj'} (1 - e^{Z_i^T \beta + \gamma_j}) h_{ij} \\
&= \sum_{i=1}^{n} \mu_{ij} [d_{ij} \delta_{jj'} - \mu_{ij'} h_{ij} h_{ij'}], \\
\text{where } d_{ij} &= (1 - e^{Z_i^T \beta + \gamma_j}) h_{ij}, \\
\frac{\partial^2 \ell_a}{\partial \beta \partial \beta^T} &= \sum_{i=1}^{n} Z_i \sum_{t=1}^{n_2} \left(\frac{\partial \mu_{it}}{\partial \beta^T} h_{it} + \mu_{it} \frac{\partial h_{it}}{\partial \beta^T} \right) \\
&= \sum_{i=1}^{n} Z_i \sum_{t=1}^{n_2} \left[-\frac{(\alpha_{it} - \alpha_{i(t-1)}) \sum_{t'=1}^{n_2} \alpha_{it'} Z_i^T (h_{it'} - h_{i(t'+1)})}{\left(\sum_{t'=1}^{n_2} \alpha_{it'} g_{it'}\right)^2} h_{it} + \mu_{it} Z_i^T (1 - e^{Z_i^T \beta + \gamma_t}) h_{it} \right] \\
&= \sum_{i=1}^{n} Z_i Z_i^T \sum_{t=1}^{n_2} \left[-\mu_{it} h_{it} \sum_{t'=1}^{n_2} \mu_{it'} h_{it'} + \mu_{it} d_{it} \right] \\
&= \sum_{i=1}^{n} Z_i Z_i^T \left[\sum_{t=1}^{n_2} \mu_{it} d_{it} - \left(\sum_{t=1}^{n_2} \mu_{it} h_{it} \right)^2 \right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell_a}{\partial b_l \partial \gamma_j} &= \sum_{i=1}^{n} \left[\frac{\sum_{t=1}^{n_2} \xi_{itl} \frac{\partial g_{it}}{\partial \gamma_j}}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} - \frac{\left(\sum_{t=1}^{n_2} \xi_{itl} g_{it}\right) \left(\sum_{t=1}^{n_2} \alpha_{it} \frac{\partial g_{it}}{\partial \gamma_j}\right)}{\left(\sum_{t=1}^{n_2} \alpha_{it} g_{it}\right)^2} \right] \\
&= \sum_{i=1}^{n} \left[\frac{\sum_{t=1}^{n_2} \xi_{itl} (\delta_{tj} - \delta_{(t+1)j}) h_{ij}}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} - \frac{\left(\sum_{t=1}^{n_2} \xi_{itl} g_{it}\right) \left(\sum_{t=1}^{n_2} \alpha_{it} (\delta_{tj} - \delta_{(t+1)j}) h_{ij}\right)}{\left(\sum_{t=1}^{n_2} \alpha_{it} g_{it}\right)^2} \right] \\
&= \sum_{i=1}^{n} \left[\frac{(\xi_{ijl} - \xi_{i(j-1)l}) h_{ij}}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} - \frac{\left(\sum_{t=1}^{n_2} \xi_{itl} g_{it}\right) (\alpha_{ij} - \alpha_{i(j-1)}) h_{ij}}{\left(\sum_{t=1}^{n_2} \alpha_{it} g_{it}\right)^2} \right] \\
&= \sum_{i=1}^{n} \left[\frac{(\xi_{ijl} - \xi_{i(j-1)l}) h_{ij}}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} - \frac{\left(\sum_{t=1}^{n_2} \xi_{itl} g_{it}\right)}{\left(\sum_{t=1}^{n_2} \alpha_{it} g_{it}\right)} \times \mu_{ij} h_{ij} \right], \\
\frac{\partial^2 \ell_a}{\partial b_l \partial \beta^T} &= \sum_{i=1}^{n} \left[\frac{\sum_{t=1}^{n_2} \xi_{itl} \frac{\partial g_{it}}{\partial \beta^T}}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} - \frac{\left(\sum_{t=1}^{n_2} \xi_{itl} g_{it}\right) \left(\sum_{t=1}^{n_2} \alpha_{it} \frac{\partial g_{it}}{\partial \beta^T}\right)}{\left(\sum_{t=1}^{n_2} \alpha_{it} g_{it}\right)^2} \right] \\
&= \sum_{i=1}^{n} \left[\frac{\sum_{t=1}^{n_2} \xi_{itl} Z_i (h_{it} - h_{i(t+1)})}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} - \frac{\left(\sum_{t=1}^{n_2} \xi_{itl} g_{it}\right) \left(\sum_{t=1}^{n_2} \alpha_{it} Z_i (h_{it} - h_{i(t+1)})\right)}{\left(\sum_{t=1}^{n_2} \alpha_{it} g_{it}\right)^2} \right] \\
&= \sum_{i=1}^{n} Z_i \left[\frac{\sum_{t=1}^{n_2} (\xi_{itl} - \xi_{i(t-1)l}) h_{it}}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} - \frac{\sum_{t=1}^{n_2} \xi_{itl} g_{it}}{\sum_{t=1}^{n_2} \alpha_{it} g_{it}} \left(\sum_{t=1}^{n_2} \mu_{it} h_{it} \right) \right], \\
\frac{\partial^2 \ell_a}{\partial \beta \partial \gamma_j} &= \sum_{i=1}^{n} Z_i \sum_{t=1}^{n_2} \left[\frac{\partial \mu_{it}}{\partial \gamma_j} h_{it} + \mu_{it} \frac{\partial h_{it}}{\partial \gamma_j} \right] \\
&= \sum_{i=1}^{n} Z_i \sum_{t=1}^{n_2} \left[\frac{\left\{(\alpha_{i(t-1)} - \alpha_{it})\right\} \left\{(\alpha_{ij} - \alpha_{i(j-1)})\right\} h_{ij}}{\left(\sum_{t=1}^{n_2} \alpha_{it} g_{it}\right)^2} h_{it} + \mu_{it} \delta_{tj} (1 - e^{Z_i^T \beta + \gamma_j}) h_{ij} \right] \\
&= \sum_{i=1}^{n} Z_i \left[-\mu_{ij} h_{ij} \sum_{t=1}^{n_2} \mu_{it} h_{it} + \mu_{ij} d_{ij} \right].
\end{aligned}
$$

A Newton-Raphson iteration can be used to compute the AMLEs $\hat{\gamma}, \hat{\beta}$. The corresponding AMLE of the baseline distribution function $\hat{\bar{F}}_0$ is

$$
\hat{\bar{F}}_0(t) = \sum_{j: t_j \geq t} \hat{q}_j^*, \tag{18}
$$

where $\hat{q}_j * = e^{-e^{\hat{\gamma}_j *}} - e^{-e^{\hat{\gamma}_{j*+1}}}$ and $(\hat{\gamma}, \hat{\eta}, \hat{\beta})$ is the maximizer of (16).

3 Simulation Study of Small Sample Performance

For the purpose of simulation, we generate samples of time-to-event from a proportional hazards model with survival function $\bar{F}_i(t) = [\bar{F}_0(t)]^{\exp(\beta^T Z_i)}$, where the baseline distribution function $F_0(t)$ is Weibull with shape and scale parameters $\alpha = 11$ and $\beta = 13$, respectively, and discard the samples lying outside the interval [8,16]. This truncated distribution has median 11.57. The vector of covariates, $Z = (Z_1, Z_2)$, consists of a binary variable, taking values 1 and 0 with probabilities 0.25 and 0.75, and a continuous variable having the uniform distribution over the interval [0,5]. We choose the vector of regression coefficients as $\beta = (\beta_1, \beta_2) = (1.5, 1.5)$. The 'time of interview' is generated from the discrete uniform distribution over the set of integers $\{7, 8, \ldots, 21\}$. These choices are in line with the data analytic example of the next section, where the time to landmark event is the age at menarche in years. As for the forgetting probability π_η, we use (5) with $k = 7$, $x_1 = 0$, $x_2 = 1.7$, $x_3 = 3.4$, $x_4 = 5.1$, $x_5 = 6.8$, $x_6 = 8.5$ and $x_7 = 10.2$ and the vector parameter $\eta = (b_1, b_2, \ldots, b_7) = (0.01, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15)$.

Note that the approximate log-likelihood (16) is maximized alternately with respect to η and (γ, β). For the present simulations, we use an isotonic version of the estimator of π_η in the following way. After each step of maximization with respect to η (with (γ, β) held fixed), we use isotonic regression, through the usual algorithm of pooling adjacent violators, on the estimated π_η to obtain a monotonically non-decreasing estimate of it. Maximization with respect to (γ, β) is then performed after holding π_η fixed. These steps are repeated till convergence.

As a benchmark of performance, one can consider the hypothetical situation when all the event times are perfectly recalled, that is, the data are right censored. In this case, one can use the MLE obtained by maximizing Cox's partial likelihood. We refer to this estimator based on 'complete recall' data as the 'complete recall MLE'. On the other hand, if one uses only the 'current status' information, namely whether the event of interest has happened till the time of interview, then the corresponding likelihood is

$$\prod_{i=1}^{n} \left[\bar{F}_0^{\exp(\beta^T Z_i)}(S_i)\right]^{1-\delta_i} \cdot \left[1 - \bar{F}_0^{\exp(\beta^T Z_i)}(S_i)\right]^{\delta_i},$$

which can be maximized with respect to β and the values of $\bar{F}_0$ at the possible times of inspection (namely, the integers 7 to 21). We refer to this estimator as the 'current status MLE'. Another option is to use the recalled data whenever available, but to disregard the informativeness of the censoring while handling the non-recall data. A penalized version of

Table 1: Bias, Stdev and MSE of estimated regression coefficients

Estimator	Property	$n = 50$		$n = 200$		$n = 1000$	
		β_1	β_2	β_1	β_2	β_1	β_2
Complete recall MLE	Bias	0.2981	–0.0734	0.0089	0.0037	–0.0026	0.0008
	Stdev	0.8321	0.8499	0.1293	0.5048	0.0848	0.2271
	MSE	0.7812	0.7277	0.0168	0.2548	0.0072	0.0515
Proposed AMLE	Bias	0.2593	–0.0547	0.0105	–0.0047	0.0083	–0.0011
	Stdev	1.3145	1.2913	0.1739	0.5057	0.0885	0.2272
	MSE	1.7904	1.6658	0.0303	0.2553	0.0079	0.0516
Current status MLE	Bias	–0.392	–0.3316	–0.026	–0.0367	–0.0032	0.0010
	Stdev	1.4048	1.3405	0.3225	0.9847	0.2353	0.6113
	MSE	2.1271	1.9069	0.1047	0.9709	0.0553	0.3737
Smooth-Hazard MLE	Bias	–0.170	3.551	–0.310	2.841	0.191	1.782
	Stdev	0.740	1.739	0.322	0.850	0.123	0.219
	MSE	0.569	15.648	0.198	8.780	0.0505	3.250

the corresponding likelihood is maximizied in the function `shr` of the `SmoothHazard` package of R, which fits the Cox model by using an approximation of the hazard function by a linear combination of M-splines. We refer to this estimator as the 'SmoothHazard MLE'.

We now compare the performance of the proposed AMLE of the regression coefficients with the three estimators described above. Table 1 shows the bias, the standard deviation (Stdev) and the mean squared error (MSE) of the estimated regression coefficients. The results reported here are based on 500 simulation runs for sample sizes $n = 50$, 200 and 1000. It is clear that the standard deviation of the proposed AMLE, as well as its mean square error, is larger than those of the (hypothetical) 'complete recall MLE', but smaller than the 'current status MLE'. The gap between the performances of the first two estimators becomes small as the sample size increases, though the gap between the AMLE and the 'current status MLE' does not reduce as much. The 'SmoothHazard MLE' has a persistent bias even when n is large. This outcome is expected, as the estimator is based on the assumption that the censoring is non-informative. Thus, neither the 'current status MLE' nor the 'SmoothHazard MLE' is able to successfully utilize the information contained in the recalled time-to-event data, while the proposed AMLE is able to do so.

Figure 1 shows the plot of the average of estimated baseline survival functions, for $n = 50$. Figure 2 and Figure 3 show the same plot for $n = 200$ and $n = 1000$, respectively.

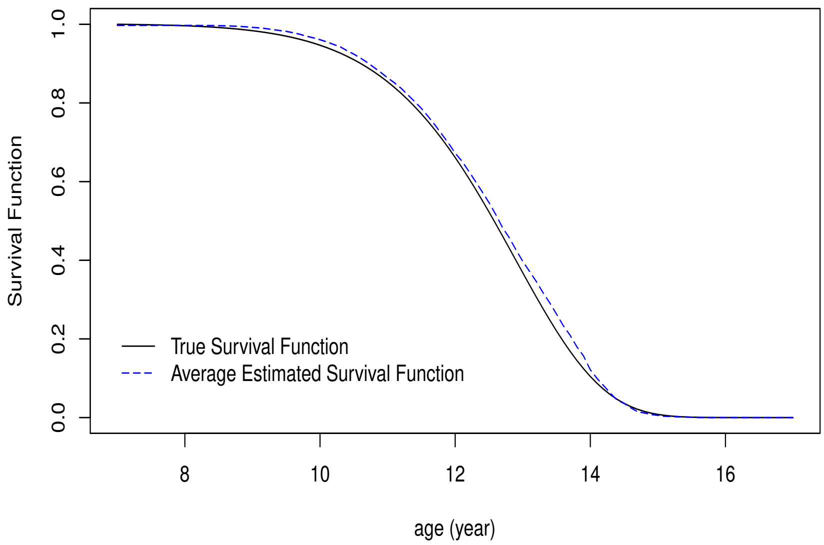

Figure 1: Average estimated baseline survival function with $n = 50$

In all the plots, it is seen that the estimated baseline survival function is close to the true baseline survival function, and more so when n is larger.

We now turn to the problem of testing for the significance of the estimators of the regression coefficients. The standard theory of parametric estimation generally does not hold for an infinite dimensional nuisance parameter. However, in the case of the Cox regression model for randomly right censored data, it has been shown that an asymptotic theory based on partial likelihood works in an analogous manner to that based on the asymptotic theory of parametric likelihood (Andersen and Gill, 1982), and that the partial likelihood may be viewed as the full likelihood maximized with respect to the baseline hazard subject to a piecewise linear constraint (Johansen, 1983). We now run some simulations to check whether the likelihood (16) with the nuisance parameters $\bar{F}_0$ replaced by the estimator (18) can be used similarly to obtain an approximate test of significance of the regression coefficients, even though there is no asymptotic theory as yet to justify such an approximation.

The 'score vector' (borrowing terminology of parametric likelihood theory) based on $\frac{\partial \ell_a(\gamma,\beta)}{\partial \beta}$ evaluated at $\beta = 0$ and $\gamma = \hat{\gamma}_0$, the restricted AMLE of γ subject to $\beta = 0$, can be

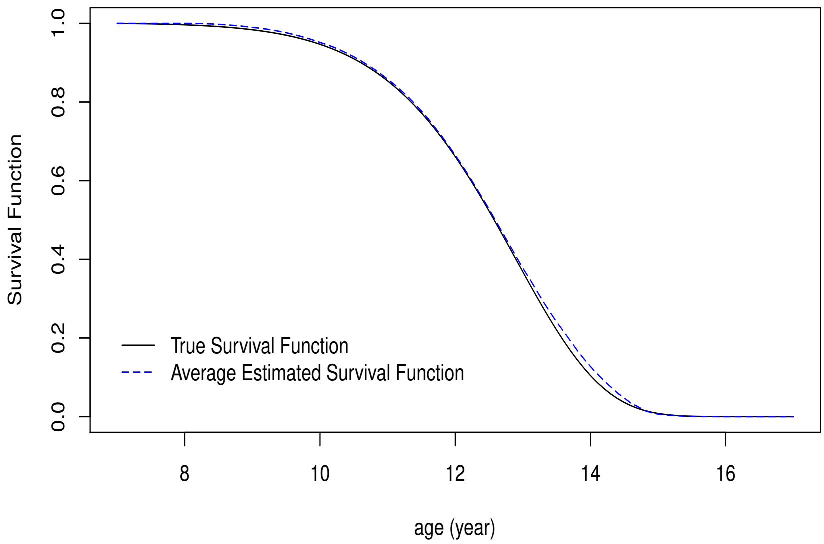

Figure 2: Average estimated baseline survival function with $n = 200$

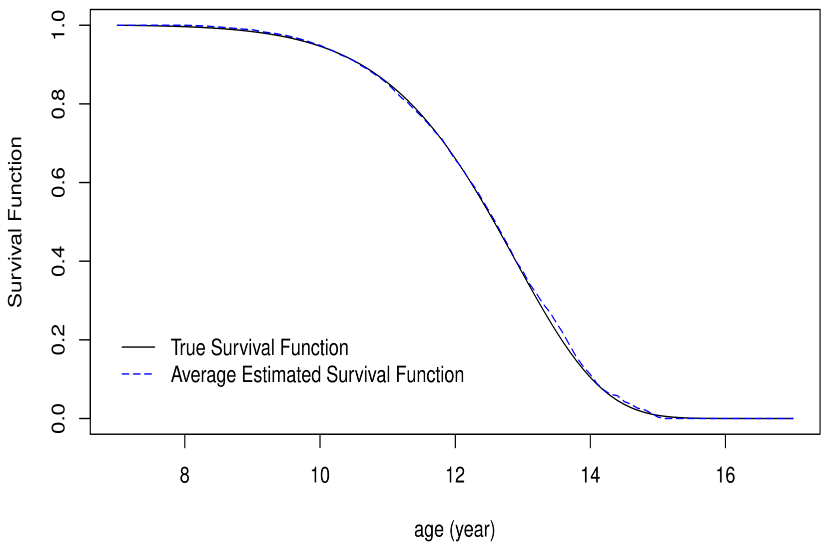

Figure 3: Average estimated baseline survival function with $n = 1000$

Table 2: The empirical type I error probability of test $H_0 : \beta = 0$

Asymtotic	$n = 50$	$n = 200$	$n = 1000$
Type I error	0.041	0.038	0.022

written as

$$U = \sum_{i=1}^{n} \sum_{j=1}^{n_2} \frac{\alpha_{ij} \left(\hat{\hat{F}}(t_j) \log(\hat{\hat{F}}(t_j)) - \hat{\hat{F}}(t_{j+1}) \log(\hat{\hat{F}}(t_{j+1})) \right) Z_i}{\sum_l \alpha_{il} g_{il}}. \tag{19}$$

Let $V = A_{22} - A_{21} A_{11}^{-1} A_{12}$, where

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

and $A_{11} = -\partial^2 \ell_a / \partial \gamma^T \partial \gamma, A_{12} = A_{21}^T = -\partial^2 \ell_a / \partial \gamma^T \partial \beta, A_{22} = -\partial^2 \ell_a / \partial \beta^T \partial \beta$, the quantities being estimated at $\beta = 0$ and $\gamma = \hat{\gamma}_0$, the AMLE at $\beta = 0$. The hypothesis $\beta = 0$ may be tested by taking $U^T V^{-1} U$ as an approximate χ^2 statistic on 2 degrees of freedom. To check the behavior of this statistic, we generate data of sizes $n = 50$, 200 and 1000 for 1000 runs, under the null hypothesis when the baseline distribution function $F_0(t)$ is the Weibull distribution with shape and scale parameters $\alpha = 11$ and $\beta = 13$, respectively, truncated to the interval [8,16], The vector of covariates, $Z = (Z_1, Z_2)$, consists of a binary variable, taking values 1 and 0 with probabilities 0.25 and 0.75, and a continuous variable having the uniform distribution over the interval [0,5]. Table 2 shows the value of empirical type I error probability of this test for different sizes of data. It can be seen that the error probability is less than 0.05. This indicates that the 'score test' is somewhat conservative.

4 Data Analysis

In this section, we illustrate the use of the proposed method with data collected from a recent anthropometric study conducted by the Biological Anthropology Unit of the Indian Statistical Institute in and around the city of Kolkata, India from 2005 to 2011 (ISI, 2012, p.108)). In this retrospective data set, individuals aged between 7 and 21 years were surveyed through stratified sampling among students of educational institutions sampled at the first stage. The subjects were interviewed on or around their birthdays. The data set contains age, menarcheal status, age at menarche (if recalled), and some other socioeconomic information. For this data set, the landmark event is the onset of menarche. Whenever the subject reported having had menarche but could not recall the date exactly, we regarded it as a case of non-recall.

There are many studies concerning the effects of socioeconomic factors on the measures of body shape (anthropometric indices or ratios) and physical maturation (e.g. biological parameters of the adolescent growth spurt) of children. Some of the important factors which effect age at menarche (maturation in girls) are home diet and physical activity which directly related to parents education and montly expenditure (Khan et al., 1996; Padez, 2003; Aryeetey et al., 2011)

We considered a subset of the original data, consisting of 673 respondents who came from a nuclear family and were the only child of their respective parents. Among 673 samples, 241 individuals did not have menarche, 147 individuals had menarche and recall the date of its onset and 285 individuals had menarche but could not recall the date. There were 492 individuals with higher education of father and 420 individuals whose mother had education higher than school. The median of montly expenditure is 7808.031 Indian Rupees. As for the forgetting probability π, we modeled it over the interval 0 to 13 years (maximum possible separation between menarcheal age and age at observation in the sample). We used a piecewise constant model, with $k = 8$ and equal length of the intervals over which the probability is constant. There are two binary covariates indicating whether the parents had passed high school, and a continuous covariate representing monthly expenditure of the family.

Table 3 shows the estimated regression coefficients and the corresponding p-values. It is found that all the coefficients are significant at the 1% level. The vector of three regression coefficients has p-value 0.00093.

Figure 4 shows a plot of the estimated baseline survival function as well as estimated survival functions in three different cases as follow.

CASE (a) Parents didn't pass the high school and monthly family expenditure is around the median. We represent this case as $Z = (0, 0, 7808)$.

Table 3: Estimated regression coefficients and their p–values

Covariates	Estimated value	p–value
Father passed high school	0.091	0.0036
Mother passed high school	0.249	0.0061
Monthly family expenditure	0.0002	0.0047

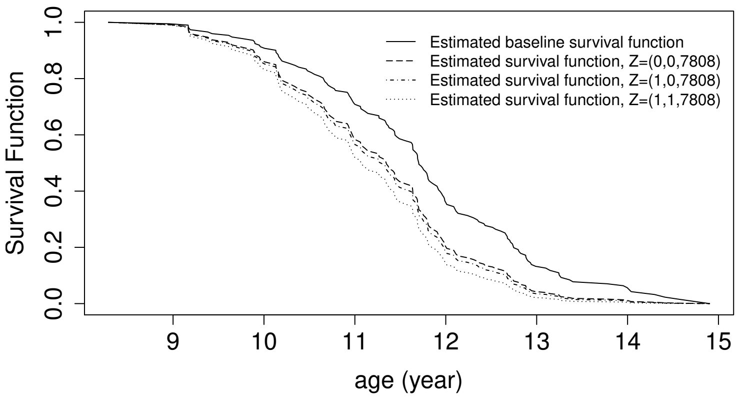

Figure 4: Estimated survival function in different cases

CASE (b) Father passed the high school, but mother didn't and monthly family expenditure is around the median. We represent this case as $Z = (1, 0, 7808)$.

CASE (c) Parents passed the high school and monthly family expenditure is around the median. We represent this case as $Z = (1, 1, 7808)$.

The chosen value of k was obtained after considering a coarser and a finer partition for the piecewise constant model of π. Specifically, the range 0 to 13 years was split experimentally into k equal intervals, with $k = 4$, 8 and 16, and the resulting estimated baseline survival functions were compared. Figure 5 shows plots of the estimated baseline survival function for different values of k. It is seen that by increasing k from 4 to 8, one observes a substantial change in the estimated baseline survival function, though the change is much less when k is increased from 8 to 16. The integrated mean square difference between baseline survival functions (scaled by the integral of the square of the function for the lower value of k) is 0.92 when one compares $k = 4$ with $k = 8$. The same criterion produces the value 0.021 when the comparison is between the curves for $k = 8$ and $k = 16$.

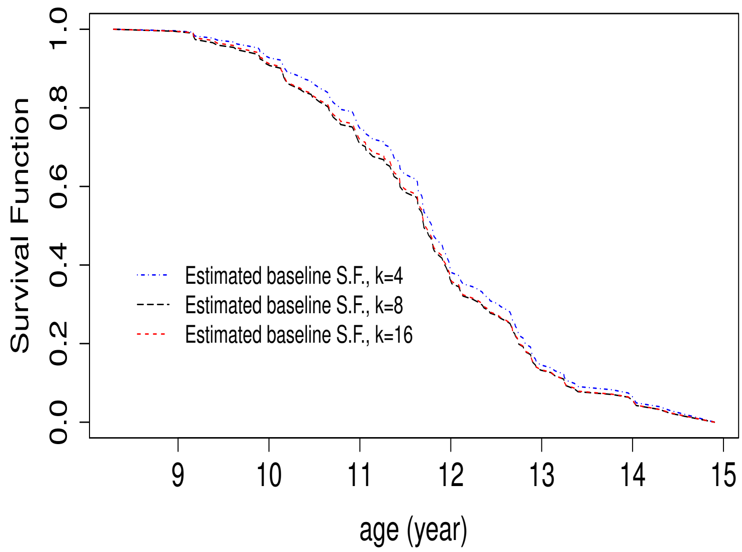

Figure 5: Estimated baseline survival function with different k

We have chosen $k = 8$, as the alternative choice $k = 16$ does not produce a substantially different estimate of the baseline survival function. Figure 6 shows the estimated function π for different values of k. Once again, the estimates of π for $k = 8$ and $k = 16$ differ much less than those for $k = 4$ and $k = 8$.

5 Concluding Remarks

In this paper, we have presented a method for fitting the Cox regression model to recall-based time-to-event data with covariates, where there is informative censoring. Simulation results indicate that the estimators of the regression coefficients are reasonable, though there is no proof of consistency of these estimators as of now. It may be recalled that there is no known proof of consistency of MLEs of the Cox regression parameters even in the case of non-informatively interval-censored data. Some results are available in the special case of status data with fixed and multiple inspection times (and in particular, for the further special case of current status data) (Huang, 1996; Yu et al., 2006; Liu and Shen, 2009). The problem of establishing consistency may be considered in future.

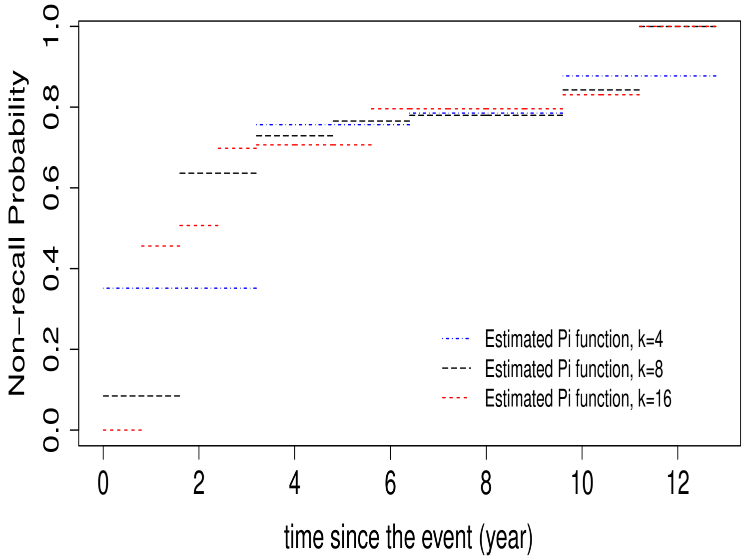

Figure 6: Estimated π function with different k

Fitting of a semi-parametric regression model is generally a more complex inferential problem than that of estimating only a distribution. The complexity in the present case is even greater because the informative interval censoring model leads to a large number of nuisance parameters, which are the probability masses allocated, as per the baseline distribution of the Cox model, to intersections of different intervals. The tasks of formation of these intervals and tracking of their probability masses are greatly simplified by the approximation inspired by Theorem 3. The Cox regression model appears to be suited to the formulation of the approximate likelihood through masses at the times of exactly recalled events. It is this matching of the models that makes the AMLE computationally tractable. A different approach may be needed for other regression models.

The proposed approach can be adapted to handle left truncated data. Assuming that there is a time of left truncation associated with each observation, each term in the likelihood would have to be divided by the upper tail probability at the point of truncation. It can be shown that the simplification given through Theorem 2 will continue to hold, since the shift of mass envisaged in the proof of that theorem does not alter the factors in the denominator. The objective funtion (16) would then be replaced by

$$\ell_a(\gamma, \beta) = \sum_{i=1}^{n} \log \left[\sum_{j=1}^{n_2} \alpha_{ij} \left[e^{(-e^{Z_i\beta+\gamma_j})} - e^{(-e^{Z_i\beta+\gamma_{j+1}})} \right] \right] - \log \left[\sum_{j=1}^{n_2} \psi_{ij} e^{(-e^{Z_i\beta+\gamma_j})} \right], \quad (20)$$

where ψ_{ij}'s are known constants like α_{ij}'s. The optimazation problem is therefore similar.

The isotonic version of the estimator of π in Sections 3 and 4 has been used in keeping with the general perception that memory fades with time. A wider set of simulations (not reported here) indicated that the variability of the estimator is reduced when the constraint of monotonicity is incorporated.

The data set analysed in Section 4 also contains 'partial' recall data relating to the week/month/year of menarche. More sophisticated modeling would be required for handling data of such complex nature. The work presented in this paper can be used as a point of departure for analysis under such models. Another direction of future research could be extension of this model to include frailty. Methodology for other forms of regression models, such as the accelerated failure time model, may be developed also.

Acknowledgements

This research is partially sponsored by the project "Physical growth, body composition and nutritional status of the Bengal school aged children, adolescent, and young adults of Calcutta, India: Effects of socioeconomic factors on secular trends", funded by the Neys Van Hoogstraten Foundation of the Netherlands, and the project "Optimization and Reliability Modeling" funded by the Indian Statistical Institute, Kolkata. The authors thank Professor Parasmani Dasgupta, leader of the first project, for making the data available for this research and Professor Biswabrata Pradhan, leader of the second project, for facilitating useful discussions. The investigation into the issue of identifiability was enriched by a discussion with Professor Probal Chaudhuri.

Appendix

A.1 Proof of Theorem 1

Without loss of generality, we assume that a possible value of z is the 0 vector (this corresponds to a shift of origin, the effect which can be absorbed through the baseline survival function). For the sake of contradiction, let us assume there are two values of the triplet $(\beta, \bar{F}_0, \pi)$, say $(\beta_1, \bar{F}_{01}, \pi_1)$ and $(\beta_2, \bar{F}_{02}, \pi_2)$, such that their substitutions in the right hand side of (4) produce the same function. Then we have, for all z and s and all positive $v < s$,

$$\begin{aligned} -\frac{d}{dv}\left(\bar{F}_{01}(s-v)^{\exp(\beta_1^T z)}\right)(1-\pi_1(v)) &= h(v, 1|s, z; \beta) \\ &= -\frac{d}{dv}\left(\bar{F}_{02}(s-v)^{\exp(\beta_2^T z)}\right)(1-\pi_2(v)). \end{aligned}$$

Hence,

$$\frac{\frac{d}{dv}\left(\bar{F}_{01}(s-v)^{\exp(\beta_1^T z)}\right)}{\frac{d}{dv}\left(\bar{F}_{02}(s-v)^{\exp(\beta_2^T z)}\right)} = \frac{1-\pi_2(v)}{1-\pi_1(v)} \qquad \forall\ z,\ s,\ v < s, \tag{A.1}$$

i.e.,

$$\frac{\exp(\beta_1^T z)\bar{F}_{01}(s-v)^{\exp(\beta_1^T z)-1} f_{01}(s-v)}{\exp(\beta_2^T z)\bar{F}_{02}(s-v)^{\exp(\beta_2^T z)-1} f_{02}(s-v)} = \frac{1-\pi_2(v)}{1-\pi_1(v)} \qquad \forall\ z,\ s,\ v < s. \tag{A.2}$$

In particular, the above identity holds for $z = 0$, i.e.,

$$\frac{f_{01}(s-v)}{f_{02}(s-v)} = \frac{1-\pi_2(v)}{1-\pi_1(v)} \qquad \forall\ s,\ v < s, \tag{A.3}$$

After combining the above equation with (A.2), we obtain

$$\frac{\bar{F}_{01}(s-v)^{\exp(\beta_1^T z)-1}}{\bar{F}_{02}(s-v)^{\exp(\beta_2^T z)-1}} = \exp((\beta_2 - \beta_1)^T z) \qquad \forall\ z,\ s,\ v < s. \tag{A.4}$$

By taking the limit of the left hand side as v goes to s, we obtain $\exp((\beta_2 - \beta_1)^T z) = 1$, i.e.,

$$\beta_2^T z = \beta_1^T z \qquad \forall\ z. \tag{A.5}$$

Since the above equation holds for r linearly independent values of the vector z (as assumed in the statement of the theorem), we have

$$\beta_1 = \beta_2. \tag{A.6}$$

It follows from equations (A.4) and (A.5) that

$$\left[\frac{\bar{F}_{01}(s-v)}{\bar{F}_{02}(s-v)}\right]^{\exp(\beta_1^T z)-1} = 1 \qquad \forall\ z,\ s,\ v < s. \tag{A.7}$$

Therefore,

$$\bar{F}_{01} = \bar{F}_{02}. \tag{A.8}$$

From equations (A.1), (A.6) and (A.8), we have

$$\pi_1(v) = \pi_2(v) \qquad \forall\ v, \tag{A.9}$$

i.e.,

$$(\beta_1, \bar{F}_{01}, \pi_1) = (\beta_2, \bar{F}_{02}, \pi_2),$$

which is a contradiction.

A.2 Proof of Theorem 2

Since $\mathcal{C}$ is the union of disjoint sets $\mathcal{C}_0$ and $\mathcal{C}_1 \cup \mathcal{C}_2$, we can rewrite the likelihood (13) as

$$\begin{aligned}
L = \prod_{i \in \mathcal{I}_1} & \left(\sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}_1 \cup \mathcal{C}_2}} p_r \right)^{\exp(\beta^T Z_i)} \times \prod_{i \in \mathcal{I}_2} \left(1 - \sum_{l=1}^{k} b_l I\big(T_i \in A_{il}\big) \right) \\
& \cdot \left[\left(\sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}_1 \cup \mathcal{C}_2}} p_r \right)^{\exp(\beta^T Z_i)} - \left(\sum_{\substack{r: I_r \subseteq A_i' \\ s_r \in \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_i' \\ s_r \in \mathcal{C}_1 \cup \mathcal{C}_2}} p_r \right)^{\exp(\beta^T Z_i)} \right] \times \\
\prod_{i \in \mathcal{I}_3} & \left[\sum_{l=1}^{k} b_l \left\{ \left(\sum_{\substack{r: I_r \subseteq A_{i(l+1)} \\ s_r \in \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_{i(l+1)} \\ s_r \in \mathcal{C}_1 \cup \mathcal{C}_2}} p_r \right)^{\exp(\beta^T Z_i)} - \left(\sum_{\substack{r: I_r \subseteq A_{il} \\ s_r \in \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_{il} \\ s_r \in \mathcal{C}_1 \cup \mathcal{C}_2}} p_r \right)^{\exp(\beta^T Z_i)} \right\} \right].
\end{aligned} \tag{A.10}$$

For every $s_r \in \mathcal{C}_2$, there exists a unique $s_{r^*} \in \mathcal{C}_0$ such that $s_{r^*} \setminus s_{r^*} \cap s_r = \{j_r\}$ and $s_r \setminus s_{r^*} \cap s_r = \{n_2 + j_r\}$ for some integer j_r in between 1 and n_2, where n_2 is as defined after (10). If any probability mass is shifted from I_r to I_{r^*}, the likelihood (A.10) can possibly be affected only through terms that involve the sets B_{j_r} and $B_{n_2+j_r}$, defined in (11). Given the fact that the baseline distribution is absolutely continuous, there is almost surely a unique $i_r \in \mathcal{I}_2$ such that $B_{j_r} = A_{i_r} \setminus A'_{i_r} = \{T_{i_r}\}$ and $B_{n_2+j_r} = A'_{i_r}$. The individual indexed by i_r is the only one whose contribution to the likelihood is affected by the change. For this individual, $I_{r^*} \subset B_{j_r} \subseteq A_{i_r}$, but $I_{r^*} \nsubseteq A'_{i_r}$. On the other hand, $I_r \subseteq B_{n_2+j_r} = A'_{i_r} \subseteq A_{i_r}$. Therefore, the first exponentiated term in the second line of (A.10) remains the same after the shift of mass, while there is a reduction in the subtracted term in that line. The likelihood increases as a result.

We now turn to shifting of probability mass out of I_r, where $s_r \in \mathcal{C}_1$. For any such s_r, define the non-empty set $\mathcal{C}_{s_r} = \{s' : s' \in \mathcal{C}_0, \quad s_r \subset s'\}$. If $\mathcal{C}_{s_r}$ is a singleton, we denote the only member by s_{r^*}. If $\mathcal{C}_{s_r}$ is not a singleton, we denote by s_{r^*} that member which satisfies the condition: 'for all $\beta \in \cup_{j:\, s_j \in \mathcal{C}_{s_r}; s_j \neq s_{r^*}} I_j$, there is a real number $\alpha \in I_{r^*}$ such that $\alpha < \beta$'. Thus, for every $s_r \in \mathcal{C}_1$, we have a uniquely defined $s_{r^*} \in \mathcal{C}_0$.

If p_r is increased at the expense of p_{r^*}, the likelihood (A.10) can possibly change only through terms that involve sets B_j such that $j \in s_{r^*} \setminus s_r$. We shall show that for an individual i, whose contribution to the likelihood involves such sets, that contribution generally increases due to the said shift of probability mass. In a particular case (Case (iii) below), where this shift cannot be proved to increase the likelihood, there is another way of shifting mass out of p_r that would definitely increase the likelihood.

CASE (i). Let $j \in s_{r^*} \setminus s_r$ and $B_j = A_{i_j}$ for some $i_j \in \mathcal{I}_1$. Any shift of probability mass from I_r to I_{r^*} would increase the contribution of the i_jth individual to the likelihood, since $I_{r^*} \subseteq A_{i_j}$ but $I_r \nsubseteq A_{i_j}$.

CASE (ii). Let $j \in s_{r^*} \setminus s_r$ and $B_j = A_{i_j} \setminus A'_{i_j}$ for some $i_j \in \mathcal{I}_2$. In this case, $I_{r^*} \subseteq A_{i_j}$ but $I_{r^*} \nsubseteq A'_{i_j}$. By construction, $B_{n_2+j} = A'_{i_j}$, which is disjoint with B_j. In order that I_{r^*} is not a null set, we must have $n_2 + j \notin s_r$. It follows that I_r is not contained in B_j or B_{n_2+j}. Thus, $I_r \nsubseteq A_{i_j}$ and $I_r \nsubseteq A'_{i_j}$. Clearly, a transfer of probability mass from I_r to I_{r^*} would increase the contribution of the i_jth individual to the likelihood.

CASE (iii). Let $j \in s_{r^*} \setminus s_r$ and $B_j = A'_{i_j}$ for some $i_j \in \mathcal{I}_2$. Since $j \notin s_r$, we have $I_r \subseteq B_j^c = [t_{min}, T_{i_j})$. Therefore, for each of the intervals B_l with $l \in s_r$, $B_l \cap [t_{min}, T_{i_j}) \neq \phi$. On the other hand, since $I_{r^*} \neq \phi$, we have $B_l \cap (T_{i_j}, t_{max}] \neq \phi$ for $l \in s_r$. It follows that each of the intervals B_l, $l \in s_r$, contains a left- and a right-neighborhood of the point T_{i_j}. Consequently, T_{i_j} is contained in these intervals. Hence, the set $s_{r^\dagger} = \{l : T_{i_j} \in B_l\}$ is a superset of s_r contained in $\mathcal{C}_0$, with $I_{r^\dagger} = \{T_{i_j}\} \neq \phi$. As argued in Case (ii), a transfer of probability mass from I_r to $I_{r^\dagger}$ would increase the contribution of the i_jth individual to the likelihood.

CASE (iv). Let $j \in s_{r^*} \setminus s_r$ and $B_j = A_{i_j(l+1)} \setminus A_{i_j l}$ for some $l \in \{1, \ldots, k\}$ and some $i_j \in \mathcal{I}_3$. A transfer of probability mass from I_r to I_{r^*} would increase the contribution of the i_jth individual to the likelihood. This is because of the fact that $I_{r^*} \subseteq A_{i_j(l+1)}$ and $I_{r^*} \nsubseteq A_{i_j l}$, whereas I_r is not contained in either of these sets.

It transpires that maximization of L can be achieved even in the presence of the constraint $p_r = 0$ for $s_r \in \mathcal{C}_1 \cup \mathcal{C}_2$. Thus, L can be fully maximized over the restricted set $\{p_r : s_r \in \mathcal{C}_0\}$.

A.3 Proof of Theorem 3

Let $i \in \mathcal{I}_2$ and the index j_i be such that $s_{j_i} = \{j : T_i \in B_j\}$. Since each time-to-event has an absolutely continuous distribution, the recalled times T_i, $i \in \mathcal{I}_2$ are distinct with probability 1. Therefore, $\{T_i\} \in \{B_1, B_2, \ldots, B_{n_2}\}$ almost surely. It follows that $T_i \in I_{j_i} \subseteq \{T_i\}$, i.e., $I_{j_i} = \{T_i\}$ with probability 1. It is also easy to see that s_{j_i} does not belong to $\mathcal{C}_1$ or $\mathcal{C}_2$, with probability 1. Therefore, $s_{j_i} \in \mathcal{C}_0$ and hence $\mathcal{A}_2 \subseteq \mathcal{A}_0$ almost surely.

Let $J_j \in \mathcal{A}_0 \setminus \mathcal{A}_2$. There is an index r such that $I_r = J_j \neq \phi$ and $s_r \in \mathcal{C}_0$, even though $I_r \neq \{T_i\}$ for any $i \in \mathcal{I}_2$. We shall show that the existence of I_r implies an event with probability going to zero.

It is easy to see that $i \notin s_r$ for $i = 1, 2, \ldots, n_2$. Thus, I_r can be written as

$$I_r = \left\{ \bigcap_{i \in s_r} B_i \right\} \bigcap \left\{ \bigcap_{i \notin s_r} B_i^c \right\} = I'_r \setminus \{T_i,\ i \in \mathcal{I}_2\},$$

where

$$I'_r = L_r \bigcap R_r, \quad L_r = \left\{ \bigcap_{i \in s_r} B_i \right\}, \quad R_r = \left\{ \bigcap_{i \notin s_r, i > n_2} B_i^c \right\}.$$

If there is an $i \in \mathcal{I}_2$ such that $T_i \in I'_r$, then the index set $s_{r^*} = s_r \cup \{i\}$ corresponds to the non-null set $I_{r^*} = \{T_i\}$. It follows that $s_r \in \mathcal{C}_1$, which leads to the contradictory conclusion $s_r \notin \mathcal{C}_0$. Therefore $T_i \notin I'_r$ for any $i \in \mathcal{I}_2$.

We now show that an upper bound of the probability of the above event goes to zero as $n_2 \to \infty$. Since the set L_r is obtained as an intersection of sets of the form $(S_i, t_{max}]$, $(T_i, t_{max}]$, $(W_l(S_i), t_{max}]$ or $[t_{min}, t_{max}]$, the intersection itself must be an interval of the form $(l_r, t_{max}]$. On the other hand, since the set R_r is obtained as an intersection of sets that are complements of sets of the above type, the intersection itself must be an interval of the form $[t_{min}, m_r]$. Thus, the set I'_r is the interval $(l_r, m_r]$. By the argument given in the preceding paragraph, neither l_r nor m_r is equal to T_i for any $i \in \mathcal{I}_2$ (otherwise s_r would not be in $\mathcal{C}_0$). Therefore, both l_r and m_r are of the form S_i for some $i \in \mathcal{I}_1$ or of the form $S_i - x_l$ for some $i \in \mathcal{I}_3$ and some $l \in \{1, \ldots, k\}$.

Let $w_1 < w_2 < \cdots < w_K$ be the feasible values of S_i and $S_i - x_l$ (where $1 \leq l \leq k$) that are strictly between t_{min} and t_{max}. Since the baseline distribution is absolutely continuous, we have $1 > \bar{F}(w_1) > \bar{F}(w_2) > \cdots > \bar{F}(w_K) > 0$. The values of l_r and m_r are taken from the set $w_1, w_2, \ldots, w_K$.

The probability of the event "$T_i \notin I'_r$ for any $i \in \mathcal{I}_2$" is

$$\begin{aligned}&\prod_{i\in\mathcal{I}_2}\left[1-\left\{\bar{F}^{\exp(\beta^T Z_i)}(l_r)-\bar{F}^{\exp(\beta^T Z_i)}(m_r)\right\}\right]\\&\quad=\prod_{i\in\mathcal{I}_2}\left[1-\left\{\bar{F}^{B}(l_r)\right\}^{\exp(\beta^T Z_i)/B}+\left\{\bar{F}^{B}(m_r)\right\}^{\exp(\beta^T Z_i)/B}\right],\end{aligned}$$

where B is an upper bound on $\exp(\beta^T Z_i)$. Since $u^{\exp(\beta^T Z_i)/B}$ is a strictly concave function of u, we have

$$1-(1-u_2+u_1)^{\exp(\beta^T Z_i)/B} < u_2^{\exp(\beta^T Z_i)/B} - u_1^{\exp(\beta^T Z_i)/B}$$

for $0 < u_1 < u_2 < 1$. Using this inequality for $u_1 = \bar{F}^B(m_r)$ and $u_2 = \bar{F}^B(l_r)$, we have

$$\begin{aligned}&\prod_{i\in\mathcal{I}_2}\left[1-\left\{\bar{F}^{\exp(\beta^T Z_i)}(l_r)-\bar{F}^{\exp(\beta^T Z_i)}(m_r)\right\}\right]\\&\quad<\prod_{i\in\mathcal{I}_2}\left[1-\bar{F}^{B}(l_r)+\bar{F}^{B}(m_r)\right]^{\exp(\beta^T Z_i)/B}\\&\quad<\left[1-\bar{F}^{B}(l_r)+\bar{F}^{B}(m_r)\right]^{n_2 L/B},\end{aligned}$$

where L is a lower bound on $\exp(\beta^T Z_i)$. Since $[1-\bar{F}^B(w_{j_1})+\bar{F}^B(w_{j_2})] \in (0,1)$ for any j_1 and j_2 with $1 \le j_1 < j_2 \le K$, we have $[1-\bar{F}^B(l_r)+\bar{F}^B(m_r)] \in (0,1)$. Therefore, the last expression goes to zero as $n_2 \to \infty$. This completes the proof.

References

Alioum, A. and Commenges, D. (1996). A proportional hazards model for arbitrarily censored and truncated data. *Biometrics* **52,** 512–524.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10,** 1100–1120.

Aryeetey, R., Ashinyo, A., and Adjuik, M. (2011). Age at menarche among basic level school girls in medina, accra. *African Journal of Reproductive Health* **103,** 103–110.

Breslow, N. E., Lubin, J. H., Marek, P., and Langholz, B. (1983). Multiplicative models and cohort analysis. *J. Amer. Statist. Assoc.* **78,** 1–12.

Chaudhuri, P. (2007). On single index regression models for multivariate survival time data. *In Nair, V., editor, Advances in Statistical Modeling and Inference:Essays in Honor of Kjell A. Doksum,* pages 223–232.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34,** 187–220. With discussion by F Downton, Richard Peto, D J Bartholomew, D V Lindley, P W Glassborow, D E Barton, Susannah Howard, B Benjamin, John J Gart, L D Meshalkin, A R Kagan, M Zelen, R E Barlow, Jack Kalbfleisch, R L Prentice and Norman Breslow, and a reply by D R Cox.

Dabrowska, D. M. and Doksum, K. A. (1988). Estimation and testing in a two-sample generalized odds-rate model. *J. Amer. Statist. Assoc.* **83,** 744–749.

DeMasi, R. A., Qaqish, B., and Sen, P. K. (1997). Statistical models and asymptotic results for multivariate failure time data with generalized competing risks. *Sankhyā Ser. A* **59,** 408–434.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42,** 845–854.

Finkelstein, D. M., Moore, D. F., and Schoenfeld, D. A. (1993). A proportional hazards model for truncated AIDS data. *Biometrics* **49,** 731–740.

Gross, S. T. and Huber-Carol, C. (1992). Regression models for truncated survival data. *Scand. J. Statist.* **19,** 193–213.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24,** 540–568.

ISI (2012). *Annual Report of the Indian Statistical Institute 2011-12.* Indian Statistical Institute. Available at URL http://library.isical.ac.in/jspui/handle/10263/5345?mode=full.

Johansen, S. (1983). An extension of Cox's regression model. *Internat. Statist. Rev.* **51,** 165–174.

Kalbfleisch, J. D. and Lawless, J. F. (1991). Regression models for right-truncated data with applications to aids incubation times and reporting lags. *Statist. Sinica* **81,** 19–32.

Khan, A. D., Schroeder, D. G., Martorell, R., Haas, J. D., and Rivera, J. (1996). Early childhood determinants of age at menarche in rural guatemala. *American Journal of Human Biology* **8,** 717–723.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data.* Springer-Verlag, New York.

Liu, H. and Shen, Y. (2009). A semiparametric regression cure model for interval-censored data. *J. Amer. Statist. Assoc.* **104,** 1168–1178.

Mirzaei, S. S., Das, R., and Sengupta, D. (2015). Parametric estimation of menarcheal age distribution based on recall data. *Scand. J. Statist.* **DOI: 10.1111/sjos.12107, to be published,**.

Mirzaei, S. S. and Sengupta, D. (2013). Nonparametric estimation of time-to-event distribution based on recall data in observational studies. *Thecnical Report No.ASD/2013/7, Applied Statistical Unit, Indian Statistical Institue* **7,**. Available at http://www.isical.ac.in/ asu/TR/TechRepASU201307.pdf.

Padez, C. (2003). Age at menarche of schoolgirls in maputo, mozambique. *Ann. Hum. Biol.* **30,,** 487–495.

Rabe-Hesketh, S., Yang, S., and Pickles, A. (2001). Multilevel models for censored and latent responses. *Stat Methods Med Res.* **10,,** 409–427.

Tanaka, Y. and Rao, P. V. (2005). A proportional hazards model for informatively censored survival times. *J. Statist. Plann. Inference* **129,** 253–262.

Tu, X. M., Meng, X. L., and Pagano, M. (1993). The aids epidemic: Estimating survival after aids diagnosis from surveillance data. *J. Amer. Statist. Assoc.* **88,** 26–36.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser. B* **38,** 290–295.

Vonta, F. (1996). Efficient estimation in a non-proportional hazards model in survival analysis. *Scand. J. Statist.* **23,** 49–61.

Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the cox regression model in survival analysis (with discussion). *Statist. Medicine* **11,** 1871–1879.

Wen, C.-C. and Chen, Y.-H. (2014). Functional inference for interval-censored data in proportional odds model with covariate measurement error. *Statistica Sinica* **24,**.

Wienke, A. (2010). *Frailty Models in Survival Analysis.* Chapman and Hall /CRC.

Yu, Q., Wong, G. Y. C., and Kong, F. (2006). Consistency of the semi-parametric MLE in linear regression models with interval-censored data. *Scand. J. Statist.* **33,** 367–378.