Local likelihood and local partial likelihood in hazard regression

Jianqing Fan[1]
Department of Statistics
Hong Kong Chinese University
Shatin, Hong Kong

Irène Gijbels[2]
Institut de Statistique
Université Catholique de Louvain
B-1348 Louvain-la-Neuve, Belgium

Martin King[3]
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27599-3260

Abstract

In survival analysis a convenient way to model the relationship between a survival time and a covariate is the proportional hazards regression model. This model usually assumes that the covariate has a log-linear effect on the hazard function. In this paper we consider the proportional hazards regression model with nonparametric risk effect. We discuss estimation of the risk function and its derivatives in two cases: when the baseline hazard function is parametrized and when it is not parametrized. In the case of a parametric baseline hazard function inference is based on a local version of the likelihood function, while in the case of a nonparametric baseline hazard we use a local version of the partial likelihood. This results in respectively the maximum local likelihood estimators and the maximum local partial likelihood estimators. We establish the asymptotic normality of the estimators. Surprisingly, both methods have the same asymptotic bias and variance for a commonly-used situation, even though the local likelihood method uses the information about the baseline hazard function.

1 The proportional hazards regression model

In survival analysis one is interested in finding the possible relationships between a survival time T and a covariate X. It is often convenient to work with the *conditional hazard function* of T given

[1] On leave from the University of North Carolina at Chapel Hill and supported by NSF DMS9504414 and an NSA grant.
[2] Supported by 'Projet d'Actions de Recherche Concertées' (No. 93/98–164) and research grant (No. 15.001.95F) of the National Science Foundation (FNRS), Belgium.

Abbreviated title: Local hazard regression.
AMS 1991 subject classification. Primary 62G05; secondary 60Gxx, 60G44.
Key words and phrases: censored data, local likelihood, local partial likelihood, proportional hazards.

$X = x$, defined as

$$\lambda(t|x) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\{t \leq T < t + \Delta t | T \geq t, X = x\}, \tag{1.1}$$

which is the instantaneous rate of failure at time t, given a particular value x for the covariate X. A popular approach to studying the covariate effect on the survival time is the *proportional hazards model*

$$\lambda(t|x) = \lambda_0(t)\Psi(x). \tag{1.2}$$

When $\Psi(0) = 1$, the function $\lambda_0(t)$ is the conditional hazard function of T given $X = 0$, and is called the *baseline hazard function.* Under model (1.2) the conditional failure rates associated with any two values of the covariate X are proportional. The most common form of the proportional hazards regression model is obtained by taking its reparametrization form

$$\Psi(x) = \exp\{\psi(x)\}, \tag{1.3}$$

as is done in this paper. Model (1.3) was introduced by Cox (1972). See Fleming and Harrington (1991), Andersen, Borgan, Gill and Keiding (1993), and references therein for the literature concerning this model.

In many applications, the survival times of studied subjects are not always fully observed, but subject to censoring. Consider the bivariate data $\{(X_i, T_i), i = 1, \cdots, n\}$ which form an i.i.d. sample from the population (X, T). Due to for example termination of the study or early withdrawal from the study, the survival times $T_1, \cdots, T_n$ are typically not fully observable, i.e. some of the observations are right-censored. The observed data can then be formulated as follows: Under an *independent censoring* scheme in which we have the i.i.d. censoring times $C_1, \cdots, C_n$ that are independent of the survival times given the covariates, we observe for the i^{th} participant an *event-time* $Z_i = \min(T_i, C_i)$, a censoring indicator $\delta_i = I\{T_i \leq C_i\}$ as well as the associated covariate X_i. The observed data are

$$\{(X_i, Z_i, \delta_i), i = 1, \cdots, n\}$$

which is an i.i.d. sample from the population

$$(X, \min(T, C), I\{T \leq C\}).$$

For convenience we assume throughout this paper that the random variables T and C are positive and continuous. The covariate X is assumed to remain constant over time.

It can be shown that under the proportional hazards model (1.2)

$$\Psi(x) = \frac{E\{\delta|X=x\}}{E\{\Lambda_0(Z)|X=x\}}, \tag{1.4}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ denotes the *cumulative baseline hazard function.* This implies that $\Psi(x)$ can easily be estimated if the baseline hazard function is known.

If a parametric form is assumed for both the baseline hazard function $\lambda_0(\cdot)$ and the risk factor $\psi(\cdot)$ inference is often based on the likelihood function. See for example Aitkin and Clayton (1980). In the case the function $\psi(\cdot)$ is not parametrized we work with a local version of this likelihood function. See Gentleman and Crowley (1991). In the first part of this paper we discuss estimation of the parametric baseline hazard function, of the risk regression function $\psi(x)$ and its derivatives based on a local likelihood. The asymptotic distribution of the resulting local likelihood estimators is established.

The situation is more complicated when the baseline hazard function is completely unknown. In the case that the function $\psi(x)$ is linear, Cox (1972) proposed to base the inference on the *partial likelihood.* In the case that $\psi(x)$ is not parametrized, a local version of this partial likelihood function will be used. This local partial likelihood has been introduced in Tibshirani and Hastie (1987). We will derive the local partial likelihood function from a different prospect and establish the asymptotic distribution of the resulting local partial likelihood estimators. The connections between the local likelihood and the local partial likelihood will also be elucidated.

When the baseline hazard function is parametrized, both the local likelihood and the local partial likelihood estimator can be employed. The former uses the knowledge of the baseline hazard function while the latter ignores this knowledge. To gain some insight into the efficiency of the local partial likelihood estimators, we compare the asymptotic biases and variances of both estimators. Surprisingly, we find that both methods have the same asymptotic bias and variance for the most-commonly used situation where the derivative curve is estimated by using a local quadratic fit. In other words, asymptotically the knowledge on the form of the baseline hazard function does not provide any extra information about the derivative curve and hence the local partial likelihood method is preferable.

There is an extensive literature on nonparametric regression techniques and it is impossible to exhaust them. See the references in Hastie and Tibshirani (1990a) and Fan and Gijbels (1996). For closely-related work on censored data we mention the following. Marron and Padgett (1987), Müller and Wang (1990, 94), Stute and Wang (1993) and Hjort (1995), among others have studied the estimation of hazard rates via kernel methods. Nonparametric estimation of the conditional hazard and distribution function using local linear fit was investigated in Li and Doss (1995). The nonparametric proportional hazards model has been studied by Tibshirani and Hastie (1987), Gentleman and Crowley (1991) via a local modeling principle, and by O'Sullivan (1988), Hastie and Tibshirani (1990b) and Kooperberg, Stone and Truong (1995a, 95b), among others, via spline models. These papers give convincing examples that demonstrate the usefulness of the nonparametric techniques and inspire our current work.

The paper is organized as follows. Section 2 deals with the situation that the baseline hazard function is parametrized, and discusses inference based on the local likelihood. The case of a nonparametric baseline function is studied in Section 3. In that section we introduce the local partial likelihood estimators and study their asymptotic behavior. Section 4 briefly discusses the relative efficiency between the local likelihood method and the local partial likelihood method. The proofs of the presented results are given in Section 5.

2 Parametric baseline hazard function

2.1 Likelihood function

Assume now that the baseline hazard function $\lambda_0(\cdot)$ has been parametrized as $\lambda_0(t) = \lambda_0(t;\theta)$ and that $\psi(x)$ has been parametrized as $\psi(x) = \psi(x;\beta)$. Let $f(t|x)$ denote the conditional density function of T given $X = x$, and let $S(t|x) = P\{T > t|X = x\}$ be its conditional survivor function. The conditional distribution function of C given $X = x$ is denoted by $G(t|x)$. Under independent and *noninformative censoring*, i.e. $G(t|x)$ does not involve the parameters θ and β, it can be shown that the conditional likelihood function is given by

$$L = \prod_{\mathrm{u}} f(Z_i|X_i) \prod_{\mathrm{c}} S(Z_i|X_i), \tag{2.1}$$

where $\prod_{\mathrm{u}}$ and $\prod_{\mathrm{c}}$ denote respectively the product involving the uncensored and the censored observations. This kind of likelihood appears often in the literature (c.f. Aitkin and Clayton, 1980). See also Section 5.3.4 of Fan and Gijbels (1996) for a simple derivation of this likelihood function. Under the proportional hazards model (1.2) we have

$$\log L = \sum_{i=1}^{n} [\delta_i\{\log \lambda_0(Z_i;\theta) + \psi(X_i;\beta)\} - \Lambda_0(Z_i;\theta)\exp\{\psi(X_i;\beta)\}]. \tag{2.2}$$

Maximization of (2.2) leads to the maximum likelihood estimators of θ and β.

2.2 Local likelihood

Suppose now that the form of $\psi(x)$ is not specified, and that the p^{th} order derivative of $\psi(x)$ at the point x exists. Then, by Taylor's expansion,

$$\psi(X) \approx \psi(x) + \psi'(x)(X-x) + \cdots + \frac{\psi^{(p)}(x)}{p!}(X-x)^p, \tag{2.3}$$

for X in a neighborhood of x. Let h be the bandwidth parameter that controls the size of the local neighborhood and let K be a kernel function, which smoothly weighs down the contribution of a remote datum point. Set

$$\underset{\sim}{X} = \{1, X-x, \cdots, (X-x)^p\}^T \qquad \text{and} \qquad \underset{\sim}{X}_i = \{1, X_i - x, \cdots, (X_i - x)^p\}^T,$$

where T denotes the transpose of a vector. Then, locally around x, $\psi(X)$ can be modeled as

$$\psi(X) \approx \underset{\sim}{X}^T \beta, \tag{2.4}$$

where $\beta = (\beta_0, \cdots, \beta_p)^T = \{\psi(x), \cdots, \psi^{(p)}(x)/p!\}^T$. Using the local model (2.4), and incorporating the localizing weights, we obtain the local (log) likelihood

$$\ell_n(\beta, \theta) = n^{-1} \sum_{i=1}^{n} \left[\delta_i \{\log \lambda_0(Z_i; \theta) + \underset{\sim}{X}_i^T \beta\} - \Lambda_0(Z_i; \theta) \exp(\underset{\sim}{X}_i^T \beta)\right] K_h(X_i - x), \tag{2.5}$$

where $K_h(t) = h^{-1} K(t/h)$. Compare with (2.2). This local likelihood was used by Gentleman and Crowley (1991). It can be easily seen that $\ell_n(\beta, \theta)$ is concave with respect to β. Hence, this local log-likelihood has a unique maximizer with respect to β. However, $\ell_n(\beta, \theta)$ is not always concave with respect to β and θ. See Section 2.4 for further discussion on this issue.

Let $\widehat{\beta}$ and $\widehat{\theta}$ be the maximizers of (2.5). Then, according to our parametrization a natural estimator of $\psi^{(\nu)}(x)$, with $\nu = 0, \cdots, p$, is

$$\widehat{\psi}_\nu(x) = \nu! \widehat{\beta}_\nu. \tag{2.6}$$

As x varies in a range of a set, we obtain an estimated (derivative) curve on that set.

2.3 Sampling properties

Throughout this section we assume that the censoring scheme is independent and noninformative. We will always use the superscripts $'$ and $''$ to indicate respectively the gradient vector and the Hessian matrix of a certain function. For example, $\lambda_0'(t; \theta)$ denotes the gradient vector of $\lambda_0(t; \theta)$

with respect to the parameter vector θ. For identifiability purpose we assume that $\lambda_0(1;\theta) = 1$. The local maximum likelihood estimators $\widehat{\beta}$ and $\widehat{\theta}$ solve the local likelihood equation $\ell_n'(\beta,\theta) = 0$, i.e.,

$$\begin{cases} \dfrac{\partial \ell_n(\beta,\theta)}{\partial \beta} = n^{-1} \displaystyle\sum_{i=1}^n \left[\delta_i - \Lambda_0(Z_i;\theta)\exp(\underset{\sim}{X}_i{}^T\beta)\right] \underset{\sim}{X}_i K_h(X_i - x) = 0 \\ \\ \dfrac{\partial \ell_n(\beta,\theta)}{\partial \theta} = n^{-1} \displaystyle\sum_{i=1}^n \left\{\delta_i \bar{\lambda}_0'(Z_i;\theta) - \Lambda_0'(Z_i;\theta)\exp(\underset{\sim}{X}_i{}^T\beta)\right\} K_h(X_i - x) = 0, \end{cases} \tag{2.7}$$

where we introduced the notation $\bar{\lambda}_0(t;\theta) = \log \lambda_0(t;\theta)$. Let θ_0 be the true parameter of $\lambda_0(t;\theta)$, and similarly let

$$\beta^0 = \{\psi(x), \cdots, \psi^{(p)}(x)/p!\}^T, \tag{2.8}$$

be the true parameter in the local model (2.4). In order to have a consistency property for $\widehat{\beta}$ and $\widehat{\theta}$ it is necessary that the true parameters β^0 and θ_0 solve the asymptotic counterpart of (2.7). That this indeed holds true for the first local likelihood equation can be seen easily via invoking (1.4). For the second local likelihood equation, this is justified by the following proposition.

Proposition 1. If θ_0 is an interior point of the parameter space, and

$$E\left\{\|\delta\bar{\lambda}_0'(Z,\theta_0)\|\,|X = x\right\} < \infty \qquad \text{and} \qquad E\left\{\|\Lambda_0'(Z,\theta_0)\|\,|X = x\right\} < \infty,$$

then

$$E\left\{\delta\bar{\lambda}_0'(Z;\theta_0) - \Lambda_0'(Z;\theta_0)\Psi(x)|X = x\right\} = 0, \tag{2.9}$$

where $\Psi(x) = \exp\{\psi(x)\}$.

Proof. Let $N(t) = I\{Z \le t, \delta = 1\}$, $Y(t) = I\{Z \ge t\}$ and let

$$\mathcal{F}_t = \sigma\{X, N(u), Y(u), 0 \le u \le t\}$$

be the history up to time t. Set

$$M(t) = N(t) - \int_0^t Y(u)\Psi(X)\lambda_0(u;\theta_0)du.$$

Then, $M(t)$ is an $\mathcal{F}_t$-martingale. Note that

$$\delta\bar{\lambda}_0'(Z;\theta_0) - \Lambda_0'(Z;\theta_0)\Psi(X) = \int_0^\infty \bar{\lambda}_0'(t;\theta_0)dM(t).$$

Since $\bar{\lambda}_0'(t;\theta_0)$ is $\mathcal{F}_t$-measurable, (2.9) follows by taking the conditional expectation of the above equality with respect to Z given $X = x$. This completes the proof. □

Equation (2.9) gives the first order Bartlett identity of the local likelihood. We also need the following second order Bartlett identity.

Proposition 2. If θ_0 is an interior point of the parameter space, and

$$E\left[\|\delta\{\bar{\lambda}_0'(Z;\theta_0)\}^{\otimes 2}\|\,|X=x\right]<\infty \qquad \text{and} \qquad E\left\{\|\Lambda_0''(Z;\theta_0)\|\,|X=x\right\}<\infty,$$

then

$$E\left\{\delta\bar{\lambda}_0''(Z;\theta_0)-\Lambda_0''(Z;\theta_0)\Psi(x)|X=x\right\}=-E\left[\delta\{\bar{\lambda}_0'(Z;\theta_0)\}^{\otimes 2}|X=x\right], \tag{2.10}$$

where $A^{\otimes 2}$ denotes AA^T for a vector or matrix A.

Proof. Using the notation of the proof of Proposition 1, we have

$$\delta\frac{\lambda_0''(Z;\theta_0)}{\lambda_0(Z;\theta_0)}-\Lambda_0''(Z;\theta_0)\Psi(X)=\int_0^\infty\frac{\lambda_0''(u;\theta_0)}{\lambda_0(u;\theta_0)}dM(u).$$

Hence,

$$E\left\{\delta\frac{\lambda_0''(Z;\theta_0)}{\lambda_0(Z;\theta_0)}-\Lambda_0''(Z;\theta_0)\Psi(x)|X=x\right\}=0.$$

Using this and

$$\bar{\lambda}_0''(Z;\theta_0)=\frac{\lambda_0''(Z;\theta_0)}{\lambda_0(Z;\theta_0)}-\left\{\frac{\lambda_0'(Z;\theta_0)}{\lambda_0(Z;\theta_0)}\right\}^{\otimes 2}$$

we obtain (2.10). □

Before we state the main results of this section, we first introduce some notations. Let

$$H=\operatorname{diag}\{1,h,\cdots,h^p\}^T, \qquad \underset{\sim}{u}=\{1,u,\cdots,u^p\}^T,$$

and put

$$S_0(x;\theta_0) = \int_{-\infty}^{+\infty}E\{\delta\begin{pmatrix}\underset{\sim}{u}\\ \bar{\lambda}_0'(Z;\theta_0)\end{pmatrix}^{\otimes 2}|X=x\}K(u)du \tag{2.11}$$

$$S_1(x;\theta_0) = \int_{-\infty}^{+\infty}E\{\delta\begin{pmatrix}\underset{\sim}{u}\\ \bar{\lambda}_0'(Z;\theta_0)\end{pmatrix}^{\otimes 2}|X=x\}K^2(u)du, \tag{2.12}$$

and

$$b_n(x)=\frac{\psi^{(p+1)}(x)}{(p+1)!}h^{p+1}\left\{\int\underset{\sim}{u}\,\underset{\sim}{u}^TK(u)du\right\}^{-1}\int u^{p+1}\underset{\sim}{u}K(u)du. \tag{2.13}$$

We now impose some convenient conditions.

Condition A:

(i). θ_0 is an interior point of the parameter space;

(ii). There exists an $\eta > 0$ such that

$$E\left\{\Lambda_0(Z;\theta_0)^{2+\eta}|X\right\}, \qquad E\left\{|\bar\lambda_0'(Z;\theta_0)|^{2+\eta}|X\right\} \qquad \text{and} \qquad E\left\{|\Lambda_0'(Z;\theta_0)|^{2+\eta}|X\right\}$$

are finite and continuous at the point $X = x$;

(iii). The functions $E(\delta|X)$, $E\{\Lambda_0(Z;\theta_0)|X\}$, $E\{\Lambda_0'(Z;\theta_0)|X\}$, $E\{\Lambda_0''(Z;\theta_0)|X\}$, $E\{\delta\bar\lambda_0'(Z;\theta_0)|X\}$ and $E\{\delta\bar\lambda_0''(Z;\theta_0)|X\}$ are continuous at the point $X = x$;

(iv). There exists a function $M(z)$, with $EM(Z) < \infty$, such that

$$\left|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_\ell}\bar\lambda_0(z;\theta)\right| < M(z), \qquad \left|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_\ell}\Lambda_0(z;\theta)\right| < M(z),$$

for all z, and for all θ in a neighborhood of θ_0;

(v). The kernel function $K \geq 0$ is a bounded density with a compact support;

(vi). The function $\psi(\cdot)$ has a continuous $(p+1)^{th}$ derivative around the point x;

(vii). The density $f(\cdot)$ of X is continuous at the point x and $f(x) > 0$;

(viii). $nh \to \infty$ and nh^{2p+3} is bounded.

Theorem 1. Under Condition A, there exists a solution $\widehat\beta$ and $\widehat\theta$ to the local likelihood equation (2.7) such that

$$H(\widehat\beta - \beta^0) \xrightarrow{P} 0 \qquad \text{and} \qquad \widehat\theta - \theta_0 \xrightarrow{P} 0.$$

Theorem 2. Under Condition A, the solution given in Theorem 1 is asymptotically normal:

$$\sqrt{nh}\begin{pmatrix} H(\widehat\beta - \beta^0) - b_n(x) \\ \widehat\theta - \theta_0 \end{pmatrix} \xrightarrow{D} \mathrm{N}\{0, f^{-1}(x)S_0(x;\theta_0)^{-1}S_1(x;\theta_0)S_0(x;\theta_0)^{-1}\}.$$

Remark 1. The bias of $\widehat\beta$ admits the same expression as that of the least-squares nonparametric regression. See (3.8) of Fan and Gijbels (1996). The explanation for this is that the bias comes from the approximation error and hence is independent of the model.

Remark 2. For the parametric linear model $\psi(X;\beta) = \underset{\sim}{X}^T\beta$, one would directly maximize the log-likelihood (2.2). Our proofs of Theorems 1 and 2 show that

$$\sqrt{n}\begin{pmatrix} \widehat{\beta} - \beta^0 \\ \widehat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{D} \mathrm{N}(0, \Sigma^{-1}),$$

where $\Sigma = E\{\delta \begin{pmatrix} \underset{\sim}{X} \\ \bar{\lambda}_0'(Z;\theta_0) \end{pmatrix}^{\otimes 2}\}$.

Remark 3. For an efficient estimation of θ one should rely on all data points. A possible approach is to maximize (2.5) with θ fixed, to obtain an estimate of $\widehat{\psi}(x,\theta)$ for each θ and then to maximize (2.2) with $\psi(X_i;\beta)$ replaced by $\widehat{\psi}(X_i,\theta)$. This is basically the profile likelihood method.

2.4 Concavity of the local likelihood

Like in most of the parametric likelihood theory [see for example Sections 6.3 and 6.4 of Lehmann (1983)], we only know that there exists a consistent solution to the local likelihood equation. But, we don't know which solution is consistent if there are multiroots. However, if $\ell_n(\beta,\theta)$ is strictly concave, then the solution to (2.7) is unique and must be consistent.

The Hessian matrix of $\ell_n(\beta,\theta)$ is given by

$$\begin{aligned}
&\ell_n''(\beta,\theta) \\
&= n^{-1}\sum_{i=1}^n K_h(X_i - x) \\
&\quad\times \begin{pmatrix} -\Lambda_0(Z_i;\theta)\exp(\underset{\sim}{X}_i^T\beta)\underset{\sim}{X}_i\underset{\sim}{X}_i^T & -\exp(\underset{\sim}{X}_i^T\beta)\underset{\sim}{X}_i\Lambda_0'(Z_i;\theta)^T \\ -\exp(\underset{\sim}{X}_i^T\beta)\Lambda_0'(Z_i;\theta)\underset{\sim}{X}_i^T & -\exp(\underset{\sim}{X}_i^T\beta)\Lambda_0''(Z_i;\theta) + \delta_i\bar{\lambda}_0''(Z_i;\theta) \end{pmatrix} \\
&= -n^{-1}\sum_{i=1}^n K_h(X_i - x)\Lambda_0(Z_i;\theta)\exp(\underset{\sim}{X}_i^T\beta)\begin{pmatrix} X_i \\ \Lambda_0'(Z_i;\theta)/\Lambda_0(Z_i;\theta) \end{pmatrix}^{\otimes 2} \\
&\quad -n^{-1}\sum_{i=1}^n K_h(X_i - x)\begin{pmatrix} 0 & 0 \\ 0 & \exp(\underset{\sim}{X}_i^T\beta)\Lambda_0(Z_i;\theta)\{\log\Lambda_0(Z_i;\theta)\}'' - \delta_i\bar{\lambda}_0''(Z_i;\theta) \end{pmatrix}.
\end{aligned} \tag{2.14}$$

Clearly the first term on the left-hand side of (2.14) is negatively definite and if

$$\{\log\Lambda_0(Z_i;\theta)\}'' \geq 0 \qquad \text{and} \qquad \bar{\lambda}_0''(Z_i;\theta) \leq 0, \tag{2.15}$$

then $\ell_n(\beta,\theta)$ is strictly concave. We summarize the result as follows.

Theorem 3. A necessary condition for the strict concavity of $\ell_n(\beta, \theta)$ is condition (2.15). For this special case, the maximum local likelihood estimator is unique and possesses the property stated in Theorem 2.

Example (Weibull distribution). For the Weibull baseline distribution with scale parameter ρ and index κ, the hazard and cumulative hazard function are respectively given by

$$\lambda_0(t; \rho, \kappa) = \kappa\rho(\rho t)^{\kappa-1} \quad \text{and} \quad \Lambda_0(t; \rho, \kappa) = (\rho t)^{\kappa}. \tag{2.16}$$

Model (1.2) can be written as

$$\lambda(t|x) = t^{\kappa-1} \exp\{\psi_0(t)\}, \quad \psi_0(x) = \psi(x) + \log \kappa + \kappa \log \rho$$

and our estimation equation (2.7) estimates κ and $\psi_0(\cdot)$ (ρ is not identifiable). This can also be seen from the identifiability condition $\lambda_0(1; \theta) = 1$, which implies that $\kappa\rho^{\kappa} = 1$ and hence $\lambda_0(t; \rho, \kappa) = t^{\kappa-1}$. Clearly, $\lambda_0(t; \rho, \kappa)$ satisfies Condition (2.15).

3 Nonparametric baseline hazard function

3.1 From likelihood to partial likelihood

Let $t_1 < \cdots < t_N$ denote the ordered failure times and let (j) denote the label of the item failing at t_j. Denote by $\mathcal{R}_j$ the risk set at time t_j-, i.e. $\mathcal{R}_j = \{i : Z_i \geq t_j\}$. Consider the "least informative" nonparametric modeling for $\Lambda_0(\cdot)$, i.e. $\Lambda_0(t)$ has a possible jump θ_j only at t_j. More precisely,

$$\Lambda_0(t; \theta) = \sum_{j=1}^{N} \theta_j I\{t_j \leq t\}. \tag{3.1}$$

Then

$$\Lambda_0(Z_i; \theta) = \sum_{j=1}^{N} \theta_j I\{i \in \mathcal{R}_j\}.$$

Substituting these two expressions into (2.2), we obtain that

$$\log L = \sum_{j=1}^{N} \left[\log \theta_j + \psi\{X_{(j)}; \beta\}\right] - \sum_{i=1}^{n} \left[\sum_{j=1}^{N} \theta_j I\{i \in \mathcal{R}_j\} \exp\{\psi(X_i; \beta)\}\right]. \tag{3.2}$$

The maximum of $\log L$ with respect to θ_j $(j = 1, \cdots, N)$ is obtained at

$$\widehat{\theta}_j = \left[\sum_{i \in \mathcal{R}_j} \exp\{\psi(X_i; \beta)\}\right]^{-1}. \tag{3.3}$$

This is indeed the Breslow estimator of the baseline hazard function [see Breslow (1972, 74)]. Substituting (3.3) into (3.2), we obtain

$$\max_{\lambda_0} \log L = \sum_{j=1}^{N} \left(\psi\{X_{(j)};\beta\} - \log[\sum_{i \in \mathcal{R}_j} \exp\{\psi(X_i;\beta)\}] \right) - N. \tag{3.4}$$

Therefore, the maximum likelihood estimate of β under the nonparametric model (3.1) is to find β that maximizes (3.4). The objective function in (3.4) is the same as the partial likelihood function in Cox (1975) and is a profile likelihood. This kind of derivation is due to Breslow (1972). Theory on partial likelihood can be found in Wong (1986).

In summary, when $\lambda_0(\cdot)$ are not specified, one would use the maximum partial likelihood estimator. This method is equivalent to the maximum likelihood estimator with the "least informative" baseline hazard function, namely with the cumulative hazard function parametrized as in (3.1) with a large number of parameters.

3.2 Local partial likelihood

When the form of $\psi(x;\beta)$ and $\Lambda_0(t;\theta)$ are not specified, one would use the local modeling (2.4) along with a local version of the partial likelihood in (3.4). Namely, finding β to maximize the local partial likelihood

$$\sum_{j=1}^{N} K_h\{X_{(j)} - x\} \left[\underset{\sim}{X}_{(j)}^{T}\beta - \log\{ \sum_{i \in \mathcal{R}_j} \exp(\underset{\sim}{X}_i^{T}\beta) K_h(X_i - x)\} \right]. \tag{3.5}$$

Clearly (3.5) is just a localized version of (3.4). It can also be derived from the local log-likelihood (2.5) using the "least-informative" nonparametric modeling (3.1). In other words, the maximum local partial likelihood estimator is the maximum local likelihood estimator when $\Lambda_0(\cdot;\theta)$ is parametrized with a large number of parameters as in (3.1).

Let $\widehat{\beta}$ maximize (3.5) with respect to $\beta = \{\beta_0, \cdots, \beta_p\}^T$. Then an obvious estimator of $\psi^{(\nu)}(x)$ is as in (2.6), namely,

$$\widehat{\psi}_\nu(x) = \nu!\widehat{\beta}_\nu, \tag{3.6}$$

where $\widehat{\beta}_\nu$ estimates β_ν. Note that the function value $\psi(x)$ is not directly estimable: (3.5) does not involve the intercept $\beta_0 = \psi(x)$ since it cancels out. This is not surprising since from model (1.2) it is already clear that $\psi(x)$ is only identifiable within a constant factor. The identifiability of $\psi(x)$ is ensured by imposing the condition $\psi(0) = 0$. Then the function $\psi(x) = \int_0^x \psi'(t)dt$ can be estimated by

$$\widehat{\psi}(x) = \int_0^x \widehat{\psi}_1(t)dt. \tag{3.7}$$

For practical implementation, Tibshirani and Hastie (1987) suggested to approximate the integration by trapezoidal rule.

At this point, one may wonder why not estimating $\Psi(x) = \exp\{\psi(x)\}$ directly. There are a few reasons which support the estimation of $\psi(x)$ rather than $\Psi(x)$. With $\psi(x)$ the estimation problem is range-preserving, i.e. both $\psi(\cdot)$ and its local approximant $\underset{\sim}{X}^T\beta$ have range $(-\infty, +\infty)$. Further, with $\psi(x)$ the local log-likelihood is concave in β, which ensures a unique maximizer. Finally, when h is large, the local partial likelihood reduces to the parametric partial likelihood and hence provides a parsimony to the parametric model.

3.3 Asymptotic property of the maximum local partial likelihood estimator

Since the local partial likelihood (3.5) does not involve $\beta_0 = \psi(x)$, we write

$$\beta^* = (\beta_1, \cdots, \beta_p)^T, \qquad \widehat{\beta}^* = (\widehat{\beta}_1, \cdots, \widehat{\beta}_p)^T \qquad \text{and} \qquad \underset{\sim}{X_i}^* = \{X_i - x, \cdots, (X_i - x)^p\}^T.$$

Correspondingly, let

$$\beta^{0*} = \{\psi'(x), \cdots, \psi^{(p)}(x)/p!\}^T, \qquad H^* = \operatorname{diag}(h, \cdots, h^p)^T, \qquad \underset{\sim}{u}^* = \{u, \cdots, u^p\}^T.$$

Set $\nu_1 = \int \underset{\sim}{u}^* K(u)du$,

$$A = \int \underset{\sim}{u}^* \underset{\sim}{u}^{*T} K(u)du - \nu_1\nu_1^T, \quad \text{and} \quad B = \int K^2(u)(\underset{\sim}{u}^* - \nu_1)^{\otimes 2}du. \tag{3.8}$$

Put

$$P(u|x) = P\{Z \geq u | X = x\} \qquad \text{and} \qquad \Lambda(t, x) = \int_0^t P(u|x)\lambda_0(u)du. \tag{3.9}$$

Writing the conditional probability as the conditional expectation of the indicator, it can easily be seen via exchanging the integration with the conditional expectation that

$$\Lambda(t, x) = E[\Lambda_0\{\min(Z, t)\}|X = x],$$

where $\Lambda_0(\cdot)$ is the cumulative baseline hazard function. Finally, let $\widehat{\beta}^*$ be the maximizer of the local partial likelihood

$$\sum_{j=1}^N K_h\{X_{(j)} - x\} \left[\underset{\sim}{X}_{(j)}^{*T}\beta^* - \log\{\sum_{i \in \mathcal{R}_j} \exp(\underset{\sim}{X_i}^{*T}\beta^*)K_h(X_i - x)\} \right] \tag{3.10}$$

with respect to β^*.

We now impose some convenient technical conditions for asymptotic normality.

Condition B:

(i). The kernel function $K \geq 0$ is a bounded density function with a compact support;

(ii). The function $\psi(\cdot)$ has a continuous $(p+1)^{th}$ derivative around the point x;

(iii). The density $f(\cdot)$ of X is continuous at the point x and $f(x) > 0$;

(iv). The conditional probability $P(u|\cdot)$ is equi-continuous at x.

We now state the main result of this section.

Theorem 4. Under Condition B, if $nh \to \infty$ and nh^{2p+3} is bounded, then

$$\sqrt{nh}\left\{H^*(\widehat{\beta}^* - \beta^{0*}) - \frac{\psi^{(p+1)}(x)}{(p+1)!}A^{-1}bh^{p+1}\right\} \xrightarrow{D} \mathrm{N}\{0, \frac{\sigma^2(x)}{f(x)}A^{-1}BA^{-1}\}, \tag{3.11}$$

where $\sigma^2(x) = \exp\{-\psi(x)\}E\{\Lambda_0(Z)|X=x\}^{-1} = E\{\delta|X=x\}^{-1}$ and $b = \int u^{p+1}(\underset{\sim}{u}^* - \nu_1)K(u)du$.

Note that in general the bias vector and variance matrix in (3.11) depend on K in a different way from that in nonparametric regression (compare with for example (3.18) and (3.19) of Fan and Gijbels, 1996). Theorem 4 gives the joint asymptotic normality of the derivative estimators. In particular, the bias and variance of $\widehat{\psi}_\nu(x)$ can be obtained by taking the ν^{th} marginal component of (3.11). For example, if $p=2$ and $\nu=1$, we have the following result.

Corollary 1. Under the conditions of Theorem 4 with $p=2$, if K is symmetric,

$$\sqrt{nh^3}\left\{\widehat{\psi}_1(x) - \psi'(x) - \frac{1}{6}\int t^3K_1^*(t)dt\ \psi^{(3)}(x)h^2\right\} \xrightarrow{D} \mathrm{N}\{0, \frac{\sigma^2(x)}{f(x)}\int K_1^*(t)^2dt\}, \tag{3.12}$$

where $K_1^*(t) = tK(t)/\int t^2K(t)dt$.

As a consequence of (3.12), the theoretical optimal bandwidth, which minimizes the asymptotic weighted mean integrated squared error

$$\int\Big[\{\frac{1}{6}\int t^3K_1^*(t)dt\ \psi^{(3)}(x)h^2\}^2 + \frac{\sigma^2(x)}{nh^3f(x)}\int K_1^*(t)^2dt\Big]w(x)dx, \tag{3.13}$$

is given by

$$h_{\text{opt}} = C(K)\left[\frac{\int\sigma^2(x)w(x)/f(x)dx}{\int\{\psi^{(3)}(x)\}^2w(x)dx}\right]^{1/7}n^{-1/7}, \tag{3.14}$$

with

$$C(K) = \left[\frac{27\int K_1^*(t)^2dt}{\{\int t^3K_1^*(t)dt\}^2}\right]^{1/7} = \left[\frac{27\int t^2K^2(t)dt}{\{\int t^4K(t)dt\}^2}\right]^{1/7}.$$

4 Relative efficiency

In this section, we study the relative efficiency between the local likelihood and the local partial likelihood estimators, when the baseline hazard function is correctly parametrized. We are interested in knowing how much efficiency gets lost by using the local partial likelihood method, which ignores the form of the baseline hazard function.

For convenience of the discussion, we assume that we use a local quadratic fit ($p = 2$) to estimate the derivative function $\psi'(x)$, using a symmetric kernel. For this special case, the local likelihood and the local partial likelihood estimator have the same asymptotic bias (see Theorem 2 and Corollary 1). Thus, we need to concentrate only on the variance expressions. Let $\sigma_c^2(x)$ be the asymptotic variance of the local likelihood estimator, which is the third diagonal element of the asymptotic variance provided in Theorem 2. Correspondingly, let $\sigma_p^2(x)$ denote the asymptotic variance of the local partial likelihood estimator, given by the second diagonal element of the asymptotic variance in Theorem 4 and expressed explicitly in (3.12). We compare the two estimators via

$$e(x) = \frac{\sigma_c^2(x)}{\sigma_p^2(x)}.$$

After some straightforward calculations and using (1.4), we find $e(x) = 1$. The result is somewhat surprising. Even though the information of the baseline hazard function was used in the local likelihood, it does not improve the sampling property of the local likelihood estimator. The reason for this asymptotic result is that we assume that $h \to 0$ and the kernel K is symmetric. Hence, in (2.5), the local parameter β_1 is orthogonal to the other parameters β_0, β_2 and θ. Our theoretical result is consistent with the simulated results in Tables 1 and 2 of Gentleman and Crowley (1991), though the latter paper considers estimation of $\psi(\cdot)$ using a local linear fit.

The above result is asymptotic in nature. We are not sure how small h should be in order to have the above orthogonality and hence it is possible at finite sample that the local likelihood estimator performs better than the local partial likelihood estimator.

We choose to compare derivative estimators instead of estimators for ψ itself because the local partial likelihood does not directly involve the local parameter for $\psi(x)$. We don't intend to compare the two methods for other cases such as the local cubic and quartic fit since such a large order of local fit is rarely used in applications.

5 Proofs

5.1 Proof of Theorem 1

Let $\alpha = H(\beta - \beta^0)$, $\widehat{\alpha} = H(\widehat{\beta} - \beta^0)$ and $U_i = H^{-1}\underset{\sim}{X}_i$. Put

$$\ell_n(\alpha,\theta) = n^{-1}\sum_{i=1}^{n}\left[\delta_i\{\bar{\lambda}_0(Z_i;\theta) + \underset{\sim}{X}_i^T\beta^0 + U_i^T\alpha\} - \Lambda_0(Z_i;\theta)\exp(\underset{\sim}{X}_i^T\beta^0 + U_i^T\alpha)\right]K_h(X_i - x).$$

Then, the problem is equivalent to showing that there exists a solution $\widehat{\alpha}$ and $\widehat{\theta}$ to the likelihood equation

$$\left\{\begin{array}{l} \dfrac{\partial \ell_n(\alpha,\theta)}{\partial\alpha} = n^{-1}\displaystyle\sum_{i=1}^{n}\left\{\delta_i - \Lambda_0(Z_i;\theta)\exp(\underset{\sim}{X}_i^T\beta^0 + U_i^T\alpha)\right\}U_iK_h(X_i - x) = 0 \\ \\ \dfrac{\partial \ell_n(\alpha,\theta)}{\partial\theta} = n^{-1}\displaystyle\sum_{i=1}^{n}\left\{\delta_i\bar{\lambda}_0'(Z_i;\theta) - \Lambda_0'(Z_i;\theta)\exp(\underset{\sim}{X}_i^T\beta^0 + U_i^T\alpha)\right\}K_h(X_i - x) = 0, \end{array}\right. \tag{5.1}$$

such that

$$\widehat{\alpha} \xrightarrow{P} 0, \qquad \widehat{\theta} - \theta_0 \xrightarrow{P} 0.$$

Let $\gamma = (\alpha^T, \theta^T)^T$ and $\gamma_0 = (0^T, \theta_0^T)^T$. Denote by S_ε the sphere centered at γ_0 with radius ε. We will show that for any sufficiently small ε, the probability that

$$\sup_{\gamma\in S_\varepsilon}\ell_n(\gamma) < \ell(\gamma_0) = \ell_n(0,\theta_0) \tag{5.2}$$

tends to one. Hence $\ell_n(\gamma)$ has a local maximum in the interior of S_ε. Since at a local maximum the likelihood equation (5.1) must be satisfied, it follows that for any $\varepsilon > 0$, with probability tending to one, the likelihood equation has a solution $\{\widehat{\alpha}(\varepsilon), \widehat{\theta}(\varepsilon)\}$ within S_ε. Let $(\widehat{\alpha}, \widehat{\theta})$ be the closest root to γ_0. Then

$$P\{\|\widehat{\alpha}\|^2 + \|\widehat{\theta} - \theta_0\|^2 \le \varepsilon\} \to 1.$$

This in turn implies that

$$H(\widehat{\beta} - \beta^0) \xrightarrow{P} 0 \qquad \text{and} \qquad \widehat{\theta} - \theta_0 \xrightarrow{P} 0.$$

We now establish (5.2). Denote by γ_j and γ_{0j} the j^{th} element of γ and γ_0 respectively. By Taylor's expansion around the point γ_0,

$$\ell_n(\gamma) - \ell_n(\gamma_0) = \ell_n'(\gamma_0)^T(\gamma - \gamma_0) + \frac{1}{2}(\gamma - \gamma_0)^T\ell_n''(\gamma_0)(\gamma - \gamma_0) + R_n(\gamma^*), \tag{5.3}$$

with γ^* lying between γ_0 and γ, and where

$$R_n(\gamma) = \frac{1}{6}\sum_{j,k,\ell}(\gamma_j - \gamma_{0j})(\gamma_k - \gamma_{0k})(\gamma_\ell - \gamma_{0\ell})\frac{\partial^3}{\partial\gamma_j\partial\gamma_k\partial\gamma_\ell}\ell_n(\gamma).$$

First of all, by recalling that $\beta_0 = \psi(x)$, we have

$$\begin{aligned}
\ell_n'(\gamma_0) &= n^{-1}\sum_{i=1}^n K_h(X_i - x)\begin{pmatrix} \{\delta_i - \Lambda_0(Z_i;\theta_0)\exp(\underset{\sim}{X}_i^T\beta^0)\}U_i \\ \delta_i\bar\lambda_0'(Z_i;\theta_0) - \Lambda_0'(Z_i;\theta_0)\exp(\underset{\sim}{X}_i^T\beta^0)\end{pmatrix} \\
&\xrightarrow{P} f(x)\begin{pmatrix} E\{\delta - \Lambda_0(Z;\theta_0)\Psi(x)|X=x\}\int \underset{\sim}{u}K(u)du \\ E\left[\delta\bar\lambda_0'(Z;\theta_0) - \Lambda_0'(Z;\theta_0)\Psi(x)|X=x\right]\end{pmatrix}.
\end{aligned}$$

By (1.4) and (2.9) we conclude that

$$\ell_n'(\gamma_0) \xrightarrow{P} 0.$$

Thus, with probability tending to one,

$$|\ell_n'(\gamma_0)^T(\gamma - \gamma_0)| \le \varepsilon^3. \tag{5.4}$$

Analogously,

$$\begin{aligned}
&\ell_n''(\gamma_0) \\
&= -n^{-1}\sum_{i=1}^n K_h(X_i - x)\begin{pmatrix} \exp(\underset{\sim}{X}_i^T\beta^0)\Lambda_0(Z_i;\theta_0)U_iU_i^T & \exp(\underset{\sim}{X}_i^T\beta^0)U_i\Lambda_0'(Z_i;\theta_0)^T \\ \exp(\underset{\sim}{X}_i^T\beta^0)\Lambda_0'(Z_i;\theta_0)U_i^T & -\delta_i\bar\lambda_0''(Z_i;\theta_0) + \Lambda_0''(Z_i;\theta_0)\exp(\underset{\sim}{X}_i^T\beta^0)\end{pmatrix} \\
&\xrightarrow{P} -f(x)\times \\
&\begin{pmatrix} \Psi(x)E\{\Lambda_0(Z;\theta_0)|X=x\}\int\underset{\sim}{u}\underset{\sim}{u}^TK(u)du & \Psi(x)\int\underset{\sim}{u}K(u)du\,E\{\Lambda_0'(Z;\theta_0)|X=x\}^T \\ \Psi(x)E\{\Lambda_0'(Z;\theta_0)|X=x\}\int\underset{\sim}{u}^TK(u)du & E\{-\delta\bar\lambda_0''(Z;\theta_0) + \Psi(x)\Lambda_0''(Z;\theta_0)|X=x\}\end{pmatrix}.
\end{aligned}$$

By (1.4), (2.9) and (2.10)

$$\begin{aligned}
&\ell_n''(\gamma_0) \\
&= -f(x)\begin{pmatrix} E\{\delta|X=x\}\int\underset{\sim}{u}\underset{\sim}{u}^TK(u)du & \int\underset{\sim}{u}K(u)du\,E\{\delta\bar\lambda_0'(Z;\theta_0)|X=x\}^T \\ E\{\delta\bar\lambda_0'(Z;\theta_0)|X=x\}\int\underset{\sim}{u}^TK(u)du & E[\delta\{\bar\lambda_0'(Z;\theta_0)\}^{\otimes 2}|X=x]\end{pmatrix} \\
&\qquad + o_P(1) \\
&= -f(x)S_0(x;\theta_0) + o_P(1).
\end{aligned} \tag{5.5}$$

Thus, with probability tending to one

$$(\gamma - \gamma_0)^T\ell_n''(\gamma_0)(\gamma - \gamma_0) < -af(x)\varepsilon^2, \qquad \text{for all } \gamma \in S_\varepsilon, \tag{5.6}$$

where a is the smallest eigenvalue of $S_0(x;\theta_0)$.

By Condition A (iv),

$$|R_n(\gamma)| \le C\varepsilon^3 n^{-1} \sum_{i=1}^{n} M(Z_i) = C\varepsilon^3\{EM(Z) + o_P(1)\}, \tag{5.7}$$

for some constant $C > 0$.

Substituting (5.4), (5.6) and (5.7) into (5.3), we conclude with probability tending to one that when ε is small enough,

$$\ell_n(\gamma) - \ell_n(\gamma_0) \le 0 \qquad \text{for all } \gamma \in S_\varepsilon.$$

This completes the proof of Theorem 1. □

5.2 Proof of Theorem 2

We continue to use the notations introduced in the proof of Theorem 1. By Taylor's expansion and Condition A (iv), we have

$$0 = \ell_n'(\widehat{\gamma}) = \ell_n'(\gamma_0) + \ell_n''(\gamma_0)(\widehat{\gamma} - \gamma_0) + O_P(\|\widehat{\gamma} - \gamma_0\|^2).$$

Hence, by (5.5)

$$\begin{aligned} (\widehat{\gamma} - \gamma_0) &= \{\ell_n''(\gamma_0) + o_P(1)\}^{-1}\, \ell_n'(\gamma_0) \\ &= \{-f(x)S_0(x;\theta_0) + o_P(1)\}^{-1}\, \ell_n'(\gamma_0). \end{aligned} \tag{5.8}$$

Thus, we only need to establish the asymptotic normality of $\ell_n'(\gamma_0)$. We first compute the mean and the variance of $\ell_n'(\gamma_0)$.

First of all, by Taylor's expansion,

$$\exp\{\psi(X_i)\} - \exp(\underset{\sim}{X}_i^T \beta^0) = \Psi(x) \frac{\psi^{(p+1)}(x)}{(p+1)!}(X_i - x)^{p+1}\{1 + o_P(1)\}. \tag{5.9}$$

By (5.1), (1.4) and (2.9), we get

$$\begin{aligned} E\ell_n'(\gamma_0) &= EK_h(X_i - x) \begin{pmatrix} \{\delta_i - \Lambda_0(Z_i;\theta_0)\exp(\underset{\sim}{X}_i^T\beta^0)\}U_i \\ \delta_i \bar{\lambda}_0'(Z_i;\theta_0) - \Lambda_0'(Z_i;\theta_0)\exp(\underset{\sim}{X}_i^T\beta^0) \end{pmatrix} \\ &= EK_h(X_i - x) \begin{pmatrix} \Lambda_0(Z_i;\theta_0)[\exp\{\psi(X_i)\} - \exp(\underset{\sim}{X}_i^T\beta^0)]U_i \\ \Lambda_0'(Z_i;\theta_0)[\exp\{\psi(X_i)\} - \exp(\underset{\sim}{X}_i^T\beta^0)] \end{pmatrix}. \end{aligned}$$

By (5.9) it can be calculated that

$$\begin{aligned}
E\ell_n'(\gamma_0) &= f(x)\Psi(x)\frac{\psi^{(p+1)}(x)}{(p+1)!}h^{p+1}\begin{pmatrix} E\{\Lambda_0(Z;\theta_0)|X=x\}\int \underset{\sim}{u}\, u^{p+1}K(u)du \\ E\{\Lambda_0'(Z;\theta_0)|X=x\}\int u^{p+1}K(u)du \end{pmatrix} + o(1) \\
&= f(x)\frac{\psi^{(p+1)}(x)}{(p+1)!}h^{p+1}\begin{pmatrix} E\{\delta|X=x\}\int \underset{\sim}{u}\, u^{p+1}K(u)du \\ E\{\delta\bar{\lambda}_0'(Z;\theta_0)|X=x\}\int u^{p+1}K(u)du \end{pmatrix} + o(1) \\
&\equiv f(x)\bar{b}_n(x) + o(1). \qquad (5.10)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathrm{Var}\{\ell_n'(\gamma_0)\} \\
&= n^{-1}EK_h^2(X-x)\begin{pmatrix} \{\delta - \Lambda_0(Z;\theta_0)\exp(\underset{\sim}{X}^T\beta^0)\}U \\ \delta\bar{\lambda}_0'(Z;\theta_0) - \Lambda_0'(Z;\theta_0)\exp(\underset{\sim}{X}^T\beta^0) \end{pmatrix}^{\otimes 2} + O(n^{-1}h^{2p+2}) \\
&= n^{-1}EK_h^2(X-x)\begin{pmatrix} \{\delta - \Lambda_0(Z;\theta_0)\Psi(X)\}U \\ \delta\bar{\lambda}_0'(Z;\theta_0) - \Lambda_0'(Z;\theta_0)\Psi(X) \end{pmatrix}^{\otimes 2} + o(n^{-1}). \qquad (5.11)
\end{aligned}$$

We now use the counting process notation introduced in the proof of Proposition 1 to simplify the expected value in (5.11). Note that

$$\begin{pmatrix} \{\delta - \Lambda_0(Z;\theta_0)\Psi(X)\}U \\ \delta\bar{\lambda}_0'(Z;\theta_0) - \Lambda_0'(Z;\theta_0)\Psi(X) \end{pmatrix} = \int_0^\infty \begin{pmatrix} U \\ \bar{\lambda}_0'(Z;\theta_0) \end{pmatrix} dM(t).$$

By conditioning on X and using the fact that the predictable variation process

$$< M, M > (t) = Y(t)\Psi(X)\lambda_0(t;\theta_0),$$

the expected value in (5.11) can be expressed as

$$\begin{aligned}
&EK_h^2(X-x)\int_0^\infty \begin{pmatrix} U \\ \bar{\lambda}_0'(t;\theta_0) \end{pmatrix}^{\otimes 2} Y(t)\Psi(X)\lambda_0(t;\theta_0)dt \\
&= EK_h^2(X-x)\int_0^\infty \begin{pmatrix} U \\ \bar{\lambda}_0'(t;\theta_0) \end{pmatrix}^{\otimes 2} dN(t) \\
&= EK_h^2(X-x)\delta\begin{pmatrix} U \\ \bar{\lambda}_0'(Z;\theta_0) \end{pmatrix}^{\otimes 2}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\mathrm{Var}\{\ell_n'(\gamma_0)\} &= n^{-1}h^{-1}f(x)\int_{-\infty}^{+\infty} E\Big\{\delta\begin{pmatrix} \underset{\sim}{u} \\ \bar{\lambda}_0'(Z;\theta_0) \end{pmatrix}^{\otimes 2} |X=x\Big\}K^2(u)du + o(n^{-1}h^{-1}) \\
&= n^{-1}h^{-1}f(x)S_1(x;\theta_0) + o(n^{-1}h^{-1}). \qquad (5.12)
\end{aligned}$$

To prove the asymptotic normality, we use the Cramer-Wold device. For any constant vector $b \neq 0$, we need to show

$$\sqrt{nh}\left\{b^T \ell_n'(\gamma_0) - b^T E\ell_n'(\gamma_0)\right\} \xrightarrow{D} \mathrm{N}\{0, f(x) b^T S_1(x;\theta_0) b\}. \tag{5.13}$$

Note that the left-hand side of (5.13) admits the form

$$\sqrt{nh}\, n^{-1} \sum_{i=1}^{n} \{K_h(X_i - x)Y_i - EK_h(X_i - x)Y_i\}.$$

To establish the asymptotic normality, we only need to verify the Lyapounov condition:

$$E \sum_{i=1}^{n} \left|\sqrt{nh}\, n^{-1}\{K_h(X_i - x)Y_i - EK_h(X_i - x)Y_i\}\right|^{2+\eta} = o(1),$$

for some $\eta > 0$. By Condition A (ii), the left-hand side of the above expression is bounded by

$$2(n^{-1}h)^{1+\eta/2}\, n\, E|YK_h(X - x)|^{2+\eta} = O\{(nh)^{-\eta/2}\} \longrightarrow 0.$$

This verifies (5.13). Consequently, by (5.8) and (5.10)

$$\begin{aligned}
&\sqrt{nh}\left\{\widehat{\gamma} - \gamma_0 + S_0(x;\theta_0)^{-1}\bar{b}_n(x)\right\} \\
&= \{-f(x)S_0(x;\theta_0) + o_P(1)\}^{-1}\{\ell_n'(\gamma_0) - E\ell_n'(\gamma_0)\} + o_P(\sqrt{nh}\, h^{p+1}) \\
&\quad \longrightarrow \mathrm{N}\{0, f^{-1}(x) S_0(x;\theta_0)^{-1} S_1(x;\theta_0) S_0(x;\theta_0)^{-1}\}.
\end{aligned} \tag{5.14}$$

It remains to simplify the bias expression $S_0(x;\theta_0)^{-1}\bar{b}_n(x)$, namely to show that

$$S_0(x;\theta_0)^{-1}\bar{b}_n(x) = \begin{pmatrix} b_n(x) \\ 0 \end{pmatrix}, \tag{5.15}$$

where $S_0(x;\theta_0)$, $\bar{b}_n(x)$ and $b_n(x)$ are given respectively by (2.11), (5.10) and (2.13).

Since $\int \underset{\sim}{u}^T K(u)du$ is the first row of the matrix $\int \underset{\sim}{u}\underset{\sim}{u}^T K(u)du$, it follows that

$$\left\{\int \underset{\sim}{u}^T K(u)du\right\}\left\{\int \underset{\sim}{u}\underset{\sim}{u}^T K(u)du\right\}^{-1} = (1, 0, \cdots, 0).$$

One can easily verify the above equality by multiplying with the matrix $\int \underset{\sim}{u}\underset{\sim}{u}^T K(u)du$ on both sides. Hence,

$$\left\{\int \underset{\sim}{u}^T K(u)du\right\}\left\{\int \underset{\sim}{u}\underset{\sim}{u}^T K(u)du\right\}^{-1}\left(\int u^{p+1}\underset{\sim}{u} K(u)du\right) = \int u^{p+1}K(u)du.$$

Using this, one can easily verify that

$$S_0(x;\theta_0)\begin{pmatrix} b_n(x) \\ 0 \end{pmatrix} = \bar{b}_n(x),$$

namely (5.15) holds. Combining (5.14) and (5.15), we obtain Theorem 2. $\square$

5.3 Proof of Theorem 4

We first state a simple lemma which will be used throughout this section.

Lemma 1. Suppose that K is bounded and compactly supported. If $g(\cdot)$ is continuous at the point x and $P(t|\cdot)$ is equi-continuous at the point x, then

$$\sup_{0\le t\le\tau} |c_n(t)-c(t)| \xrightarrow{P} 0,$$

provided that $h\to 0$, $nh/\log n\to\infty$, $0<\tau\le+\infty$, where $c_n(t)=n^{-1}\sum_{i=1}^{n} Y_i(t)g(X_i)K_h(X_i-x)$, and $c(t)=f(x)g(x)P(t|x)\int K(u)du$, with $Y_i(t)=I\{Z_i\ge t\}$.

Proof. Upon conditioning on $X_1,\cdots,X_n$ one can apply Theorem 37 in Chapter 2 of Pollard (1984) to get the result. □

Recall the notations of Section 3.3. Denote by $\widehat{\alpha}^*=H^*(\widehat{\beta}^*-\beta^{0*})$ and $U_i^*=(H^*)^{-1}\underset{\sim}{X}_i{}^*$. Then, by (3.10), $\widehat{\alpha}^*$ maximizes

$$\ell_n(\alpha)=n^{-1}\sum_{j=1}^{N} K_h\{X_{(j)}-x\}\left[\underset{\sim}{X}_{(j)}^{*T}\beta^{0*}+U_{(j)}^{*T}\alpha-\log\{\sum_{i\in\mathcal{R}_j}\exp(\underset{\sim}{X}_i{}^{*T}\beta^{0*}+U_i^{*T}\alpha)K_h(X_i-x)\}\right]$$

with respect to α. We will prove a somewhat more general result. Let $\widehat{\alpha}$ maximize

$$\begin{aligned}\ell_n(\alpha,\tau) &= n^{-1}\sum_{j=1}^{N} K_h\{X_{(j)}-x\}I\{Z_{(j)}\le\tau\}\Big[\underset{\sim}{X}_{(j)}^{*T}\beta^{0*}+U_{(j)}^{*T}\alpha \\ &\quad -\log\{\sum_{i\in\mathcal{R}_j}\exp(\underset{\sim}{X}_i{}^{*T}\beta^{0*}+U_i^{*T}\alpha)K_h(X_i-x)\}\Big]. \end{aligned} \tag{5.16}$$

Then our case corresponds to that of $\tau=\infty$.

Let

$$N_i(t)=I\{Z_i\le t,\delta_i=1\} \qquad \text{and} \qquad Y_i(t)=I\{Z_i\ge t\}.$$

Put

$$S_{n,0}(\alpha,u)=n^{-1}\sum_{i=1}^{n} Y_i(u)\exp(\underset{\sim}{X}_i{}^{*T}\beta^{0*}+U_i^{*T}\alpha)K_h(X_i-x).$$

Then,

$$\begin{aligned}\ell_n(\alpha,\tau) &= \int_0^\tau n^{-1}\sum_{i=1}^{n} K_h(X_i-x)(\underset{\sim}{X}_i{}^{*T}\beta^{0*}+U_i^{*T}\alpha)dN_i(u) \\ &\quad -\int_0^\tau \log\{nS_{n,0}(\alpha,u)\}n^{-1}\sum_{i=1}^{n} K_h(X_i-x)dN_i(u). \end{aligned} \tag{5.17}$$

The proof of Theorem 4 consists of the following main steps:

Step 1: consistency

$$\widehat{\alpha}^* \xrightarrow{P} 0. \tag{5.18}$$

Step 2: Asymptotic normality of $\ell_n'(0,\tau)$:

$$\sqrt{nh}\,\{\ell_n'(0,\tau) - b_n(\tau)\} \xrightarrow{D} \mathrm{N}\{0, \Sigma(\tau, x)\}, \tag{5.19}$$

where $\ell_n'(0,\tau) = \left.\frac{\partial \ell_n(\alpha,\tau)}{\partial \alpha}\right|_{\alpha=0}$, and

$$b_n(\tau) = f(x)\Psi(x)\frac{\psi^{(p+1)}(x)}{(p+1)!}\Lambda(\tau, x)bh^{p+1} \tag{5.20}$$

with $\Lambda(\tau, x)$ defined by (3.9) and

$$\Sigma(\tau, x) = f(x)\Psi(x)\Lambda(\tau, x)B. \tag{5.21}$$

Step 3: For any $\widehat{\alpha}^{**} \xrightarrow{P} 0$,

$$\ell_n''(\widehat{\alpha}^{**}, \tau) \xrightarrow{P} -f(x)\Psi(x)\Lambda(\tau, x)A \equiv \Sigma_1(\tau, x). \tag{5.22}$$

Once Steps $1-3$ are established we proceed as follows. Since $\widehat{\alpha}^*$ maximizes $\ell_n(\alpha,\tau)$, it follows from a Taylor expansion around 0 that

$$0 = \ell_n'(\widehat{\alpha}^*, \tau) = \ell_n'(0,\tau) + \ell_n''(\widehat{\alpha}_n^{**}, \tau)\widehat{\alpha}^*,$$

where $\widehat{\alpha}_n^{**}$ lies between 0 and $\widehat{\alpha}^*$, and hence $\widehat{\alpha}_n^{**} \xrightarrow{P} 0$. Thus, by (5.22)

$$\widehat{\alpha}^* + \Sigma_1(\tau, x)^{-1}b_n(\tau) = -\ell_n''(\widehat{\alpha}_n^{**}, \tau)^{-1}\{\ell_n'(0,\tau) - b_n(\tau)\} + o_P(h^{p+1}).$$

By (5.19), (5.22) and Slutsky's theorem, we conclude that

$$\sqrt{nh}\left\{\widehat{\alpha}^* + \Sigma_1(\tau, x)^{-1}b_n(\tau)\right\} \xrightarrow{D} \mathrm{N}\{0, \Sigma_1(\tau, x)^{-1}\Sigma(\tau, x)\Sigma_1(\tau, x)^{-1}\}. \tag{5.23}$$

From (5.23), we obtain (3.11). Thus it remains to prove (5.18), (5.19) and (5.22).

Proof of (5.18).

Let the filtration $\mathcal{F}_{nt}$ be the statistical information accruing during the time $[0,t]$, namely,

$$\mathcal{F}_{nt} = \sigma\{X_i, N_i(u), Y_i(u), i = 1, \cdots, n,\ 0 \le u \le t\}.$$

Then, under the independent censoring scheme,

$$M_i(t) = N_i(t) - \int_0^t Y_i(u)\exp\{\psi(X_i)\}\lambda_0(u)du \tag{5.24}$$

is an $\mathcal{F}_{nt}$-martingale. Let

$$S_{n,0}^*(u) = n^{-1} \sum_{i=1}^{n} K_h(X_i - x) Y_i(u) \exp\{\psi(X_i)\}$$

and

$$S_{n,1}^*(u) = n^{-1} \sum_{i=1}^{n} K_h(X_i - x) Y_i(u) \exp\{\psi(X_i)\} U_i^*.$$

Then, by (5.24), we can write

$$\ell_n(\alpha, t) - \ell_n(0, t) = A_n(\alpha, t) + X_n(\alpha, t), \tag{5.25}$$

where

$$A_n(\alpha, t) = \int_0^t S_{n,1}^*(u)^T \alpha \lambda_0(u) du - \int_0^t \log \left\{ \frac{S_{n,0}(\alpha, u)}{S_{n,0}(0, u)} \right\} S_{n,0}^*(u) \lambda_0(u) du,$$

and

$$X_n(\alpha, t) = \int_0^t n^{-1} \sum_{i=1}^{n} K_h(X_i - x) \left[U_i^{*T} \alpha - \log \left\{ \frac{S_{n,0}(\alpha, u)}{S_{n,0}(0, u)} \right\} \right] dM_i(u).$$

By Lemma 1, we have

$$\begin{aligned} A_n(\alpha, t) &= f(x)\Psi(x)\Lambda(t, x) \left[\nu_1^T \alpha - \log \left\{ \int \exp(\underset{\sim}{u}^{*T} \alpha) K(u) du \right\} \right] + o_P(1) \\ &\equiv A(\alpha, t) + o_P(1). \end{aligned} \tag{5.26}$$

The process $X_n(\alpha, \cdot)$ is a locally square integrable martingale with the predictable variation process

$$\begin{aligned} B_n(t) &\equiv \langle X_n(\alpha, \cdot), X_n(\alpha, \cdot) \rangle (t) \\ &= \sum_{i=1}^{n} \int_0^t n^{-2} K_h^2(X_i - x) \left[U_i^{*T} \alpha - \log \left\{ \frac{S_{n,0}(\alpha, u)}{S_{n,0}(0, u)} \right\} \right]^2 Y_i(u) \exp\{\psi(X_i)\} \lambda_0(u) du. \end{aligned}$$

By using Lemma 1, it can be calculated that

$$EX_n(\alpha, t)^2 = EB_n(t)^2 = O(n^{-1}h^{-1}) \longrightarrow 0, \qquad 0 \le t \le \tau.$$

This together with (5.25) and (5.26) leads to

$$\ell_n(\alpha, \tau) - \ell_n(0, \tau) = A(\alpha, \tau) + o_P(1). \tag{5.27}$$

Clearly, $A(\alpha, \tau)$ is strictly concave, having the maximizer at the point $\alpha = 0$. Since $\widehat{\alpha}^*$ maximizes the concave function $\ell_n(\alpha, \tau) - \ell_n(0, \tau)$, by the concavity lemma [see e.g. Andersen and Gill (1982)],

$$\widehat{\alpha}^* \xrightarrow{P} 0.$$

Proof of (5.19).

Let

$$S_{n,1}(\alpha,u) = n^{-1}\sum_{i=1}^{n} Y_i(u)\exp(\underset{\sim}{X}_i^{*T}\beta^{0*} + U_i^{*T}\alpha)K_h(X_i - x)U_i^{*}.$$

Then, by Lemma 1, we have

$$\sup_{0\le u\le\tau}\left\|\frac{S_{n,1}(0,u)}{S_{n,0}(0,u)} - \nu_1\right\| \xrightarrow{P} 0. \tag{5.28}$$

By (5.24), we can express $\ell_n'(0,\tau)$ as

$$\ell_n'(0,\tau) = U_n(\tau) + B_n(\tau), \tag{5.29}$$

where

$$U_n(\tau) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(X_i - x)\left\{U_i^{*} - \frac{S_{n,1}(0,u)}{S_{n,0}(0,u)}\right\} dM_i(u),$$

and

$$B_n(\tau) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(X_i - x)\left\{U_i^{*} - \frac{S_{n,1}(0,u)}{S_{n,0}(0,u)}\right\} Y_i(u)\exp\{\psi(X_i)\}\lambda_0(u)du.$$

Note that

$$\begin{aligned} B_n(\tau) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} K_h(X_i - x)&\left\{U_i^{*} - \frac{S_{n,1}(0,u)}{S_{n,0}(0,u)}\right\} \\ &\times Y_i(u)\left[\exp\{\psi(X_i)\} - \exp\{\psi(x) + \underset{\sim}{X}_i^{*T}\beta^{0*}\}\right]\lambda_0(u)du. \end{aligned} \tag{5.30}$$

By Taylor's expansion we have

$$\exp\{\psi(X_i)\} - \exp\{\psi(x) + \underset{\sim}{X}_i^{*T}\beta^{0*}\} = \Psi(x)\frac{\psi^{(p+1)}(x)}{(p+1)!}(X_i - x)^{p+1} + o_P(h^{p+1}). \tag{5.31}$$

Thus, by Lemma 1, (5.28) and (5.31) we obtain

$$\begin{aligned} B_n(\tau) &= f(x)\Psi(x)\frac{\psi^{(p+1)}(x)}{(p+1)!}\Lambda(\tau,x)\int_{-\infty}^{+\infty} K(u)(\underset{\sim}{u}^{*} - \nu_1)u^{p+1}du\, h^{p+1} + o_P(h^{p+1}) \\ &= b_n(\tau) + o_P(h^{p+1}). \end{aligned} \tag{5.32}$$

We now treat the process $U_n(t)$, using the martingale central limit theorem [see e.g. Theorem 5.3.5 of Fleming and Harrington (1991)]. The predictable variation process $U_n^{*}(t) = \sqrt{nh}\, U_n(t)$ is

$$< U_n^{*}, U_n^{*} > (t) = n^{-1}h\sum_{i=1}^{n}\int_0^{t} K_h^2(X_i - x)\left\{U_i^{*} - \frac{S_{n,1}(0,u)}{S_{n,0}(0,u)}\right\}^{\otimes 2} Y_i(u)\exp\{\psi(X_i)\}\lambda_0(u)du.$$

By Lemma 1

$$\begin{aligned} < U_n^*, U_n^* > (t) &= f(x)\Psi(x)\Lambda(t,x) \int K^2(u)(\underset{\sim}{u}^* - \nu_1)^{\otimes 2} du + o_P(1) \\ &= \Sigma(t,x) + o_P(1). \end{aligned}$$

Write the ℓ^{th} element of the vector $U_n^*(t)$ as

$$\frac{\sqrt{nh}}{n} \sum_{i=1}^n \int_0^t K_h(X_i - x) H_{n,\ell}(u) dM_i(u).$$

Then, $H_{n,\ell}(u)$ is a bounded random vector. To prove the asymptotic normality we need to check the Lindeberg condition:

$$\sum_{i=1}^n \int_0^t n^{-1} h K_h^2(X_i - x) H_{n,\ell}^2(u) I\left\{\sqrt{h/n}|K_h(X_i - x)H_{n,\ell}(u)| > \varepsilon\right\}$$
$$\times Y_i(u) \exp\{\psi(X_i)\}\lambda_0(u) du \xrightarrow{P} 0, \qquad \text{for all } \varepsilon > 0.$$

The last statement is valid since the random variable $K\{(X_i - x)/h\}H_{n,\ell}(x)$ is bounded and hence when n is large enough, the set indicator becomes empty. This establishes

$$\sqrt{nh} U_n(t) \xrightarrow{D} \mathrm{N}\{0, \Sigma(t,x)\}, \qquad 0 < t \le \tau.$$

This together with (5.29) and (5.32) proves (5.19).

<u>Proof of (5.22)</u>. Let

$$S_{n,2}(\alpha, u) = n^{-1} \sum_{i=1}^n Y_i(u) \exp(\underset{\sim}{X}_i^{*T} \beta^{0*} + U_i^{*T} \alpha) K_h(X_i - x) U_i^* U_i^{*T}.$$

Then,

$$\ell_n''(\alpha, t) = -\int_0^t \frac{S_{n,2}(\alpha,u)S_{n,0}(\alpha,u) - S_{n,1}(\alpha,u)S_{n,1}(\alpha,u)^T}{S_{n,0}(\alpha,u)^2} n^{-1} \sum_{i=1}^n K_h(X_i - x) dN_i(u).$$

For any consistent estimator $\widehat{\alpha}^{**} \xrightarrow{P} 0$, since the random vector U_i^* and the other involved random variables are bounded, by the continuity, it can be easily shown that

$$\ell_n''(\widehat{\alpha}^{**}, \tau) = \ell_n''(0, \tau) + o_P(1). \tag{5.33}$$

By using Lemma 1,

$$\frac{S_{n,2}(0,u)S_{n,0}(0,u) - S_{n,1}(0,u)S_{n,1}(0,u)^T}{S_{n,0}(0,u)^2} - A \xrightarrow{P} 0$$

uniformly in $0 \le u \le \tau$, where $A = \int \underset{\sim}{u}^* \underset{\sim}{u}^{*T} K(u) du - \nu_1 \nu_1^T$ was defined in (3.8). Thus, by Lemma 1,

$$
\begin{aligned}
\ell_n''(0, \tau) &= -A \int_0^\tau n^{-1} \sum_{i=1}^n K_h(X_i - x) dN_i(u) + o_P(1) \\
&= -A n^{-1} \sum_{i=1}^n K_h(X_i - x) N_i(\tau) + o_P(1) \\
&= -A f(x) E\{N(\tau) | X = x\} + o_P(1), \qquad (5.34)
\end{aligned}
$$

with $N(\tau) = I\{Z \le \tau, \delta = 1\}$, as in the proof of Proposition 1.

Recall that the compensator of $N(t)$ is $\Psi(X) \int_0^t Y(u) \lambda_0(u) du$. Hence,

$$
E\{N(\tau) | X = x\} = \Psi(x) \Lambda(\tau, x). \qquad (5.35)
$$

Combining (5.33) — (5.35), we conclude that

$$
\ell_n''(\widehat{\alpha}^{**}, \tau) = -f(x) \Psi(x) \Lambda(\tau, x) A + o_P(1).
$$

This proves (5.22) and completes the proof of Theorem 4. □

REFERENCES

Aitkin, M. and Clayton, D.G. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *Appl. Statist.* **29** 156 – 163.

Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993). *Statistical models based on counting processes.* Springer-Verlag, New York.

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

Breslow, N.E. (1972). Contribution to the discussion on the paper by D.R. Cox, "Regression and life tables". *J. Royal Statist. Soc. B* **34** 216 – 217.

Breslow, N.E. (1974). Covariance analysis of censored survival data. *Biometrics* **30** 89 – 99.

Cox, D.R. (1972). Regression models and life-tables (with discussion). *J. Royal Statist. Soc. B* **4** 187 – 220.

Cox, D.R. (1975). Partial likelihood. *Biometrika* **62** 269 – 276.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications.* Chapman and Hall, London.

Fleming, T.R. and Harrington, D.P. (1991). *Counting processes and survival analysis.* Wiley, New York.

Gentleman, R. and Crowley, J. (1991). Local full likelihood estimation for the proportional hazards model. *Biometrics* **47** 1283 – 1296.

Hastie, T. and Tibshirani, R. (1990a). *Generalized additive models.* Chapman and Hall, London.

Hastie, T. and Tibshirani, R. (1990b). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics* **46** 1005 – 1016.

Hjort, J. (1995). Dynamic likelihood hazard rate estimation. *Biometrika* to appear.

Kooperberg, C., Stone, C. J. and Truong, Y. (1995a). Hazard regression. *Jour. of the Amer. Statist. Assoc.* **90** 78 – 94.

Kooperberg, C., Stone, C. J. and Truong, Y. (1995b). The L_2 rate of convergence for hazard regression. *Scand. J. Statist.* **22** 143 – 157.

Lehmann, E.L. (1983). *Theory of point estimation.* Wadsworth & Brooks/Cole, Pacific Grove, California.

Li, G. and Doss, H. (1995). An approach to nonparametric regression for life history data using local linear fitting. *Ann. Statist.* **23** 787 – 823.

Marron, J.S. and Padgett, W.J. (1987). Asymptotically optimal bandwidth selection from randomly right-censored samples. *Ann. Statist.* **15** 152 – 1535.

Müller, H.G. and Wang, J.L. (1990). Analyzing changes in hazard functions: An alternative to change-point models. *Biometrika* **77** 610 – 625.

Müller, H.G. and Wang, J.L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50** 61 – 76.

O'Sullivan, F. (1988). Nonparametric estimation of relative risk using splines and cross-validation. SIAM Jour. Sci. Statist. Comp. **9** 531 – 542.

Pollard, D. (1984). *Convergence of stochastic processes.* Springer-Verlag, New York.

Stute, W. and Wang, J.L. (1993). A strong law under random censorship. *Ann. Statist.* **21** 1591 – 1607.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Jour. Amer. Statist. Assoc.* **82** 559 – 567.

Wong, W.H. (1986). Theory of partial likelihood. *Ann. Statist.* **14** 88 – 123.