LOCAL ESTIMATION OF A BIOMETRIC FUNCTION WITH COVARIATE EFFECTS

ZONGWU CAI* and LIANFEN QIAN

Department of Mathematics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA, E-mail: zcai@uncc.edu
Department of Mathematics, Florida Atlantic University, Boca Raton, FL 33431, USA, E-mail: liqian@fau.edu

ABSTRACT

In the analysis of life tables one biometric function of interest is the life expectancy at age $t$, $e(t) = E(T - t | T > t)$. In this paper $e(t)$ is extended to a regression model – proportional mean residual life regression model with both censoring and explanatory variables. The model usually assumes that the covariate has a log-linear effect on the mean residual life. We consider the proportional mean residual life regression model with a nonparametric risk effect and we discuss estimation of the risk function and its derivatives for two types of baseline mean residual life: parametrized and non-parametrized. In parametric baseline mean residual life case, inference is based on a local likelihood function, while in a nonparametric baseline mean residual life case, we propose a simple approach based on the exponential regression idea to find estimator. This simple method makes the implementation easier. Finally, the consistency and asymptotic normality of the resulting estimators are established. A simulation study is presented to illustrate the estimation procedures.

1. INTRODUCTION

Let $T$ be a non-negative random variable with a mean and a density function $s(\cdot)$. Let $S(t) = 1 - F(t)$ be the survival function. Define

$$e(t) = E(T - t | T > t) = S^{-1}(t) \int_t^\infty S(u) du.$$

*Caï's work was supported, in part, by the National Science Foundation grant DMS 0072400 and funds provided by the University of North Carolina at Charlotte.
For life tables, $e(t)$ is called the life expectancy at age $t$, or more generally a biometric function (Chiang, 1960). It is also called the mean residual life (MRL) function in the reliability literature. In biometry, $e(t)$ is defined via the force of mortality (hazard function) $\lambda(t) = s(t)/S(t)$,

$$e(t) = \int_0^\infty \exp \left\{ - \int_0^t \lambda(t + y) \, dy \right\} \, dy.$$

Like $\lambda(\cdot)$ and $s(t)$, $e(t)$ also determines $S(t)$,

$$S(t) = \frac{e(0)}{e(t)} \exp \left\{ - \int_0^t e^{-1}(u) \, du \right\}.$$

For the detailed interpretation of the MRL function $e(t)$, we refer to the papers by Dasu (1991) and Maguluri and Zhang (1994). There is a vast literature on the nonparametric estimation of $e(t)$ for both uncensored and censored data, see Yang (1978), Csörgő, Csörgő and Horváth (1984), Lee, Park and Sohn (1993), and Li (1997), to name just a few. In demographic and reliability studies, the MRL function may be more important than the hazard function, since the former deals with the entire residual life distribution, whereas the latter relates only the risk of immediate failure. It is obvious that the MRL is an important function in the actuarial work relating to life insurance.

During the past two decades, the biometric function $e(t)$ has been extended into several different directions. For example, Hall and Wellner (1984) introduced a class of survival distributions with linear MRL function, $e(t) = At + B$ ($A > -1$, $B > 0$), which covers a Pareto ($A > 0$), an exponential ($A = 0$) and a rescaled Beta ($-1 < A < 0$) distribution. Oakes and Dasu (1990) proposed a family of semiparametric proportional mean residual life (PMRL) models. Two survivor functions $S(\cdot)$ and $S_0(\cdot)$ are said to have PMRL if

$$e(t) = \theta e_0(t) \quad \text{for all} \quad t \geq 0, \quad \theta > 0. \quad (1.1)$$

Recently, Maguluri and Zhang (1994) extended the model (1.1) to a more general framework with covariate $Z

$$e(t \mid z) = \exp(\psi \, z) \, e_0(t). \quad (1.2)$$

Here, $e_0(t)$ serves as the MRL corresponding to a baseline survivor function $S_0(t)$. They proposed two methods to estimate parameter $\psi$, of which is based on the maximum likelihood equation of the exponential regression model, and the other is based on the underlying proportional hazards structure of the model and Cox's estimating equation.
Estimate of Biometric Function

In this article, we extend the model (1.2) to a more popular and general nonparametric regression model with covariate effect \( Z \)

\[
e(t | z) = \Psi(z) e_0(t),
\]

which is called the proportional mean residual life regression model. Clearly, \( e(t | z) \) is the conditional mean residual life function of \( T - t \) given \( T > t \) and \( Z = z \). When \( \Psi(0) = 1 \), the function \( e_0(t) \) is the conditional MRL function of \( T - t \) given \( T > t \) and \( Z = 0 \), and is called the baseline mean residual function. The model (1.3) looks similar to the proportional hazards model (Cox's model). Following the lead of the Cox’s model, we consider the following PMRL model by taking its reparametrization form

\[
\Psi(z) = \exp(\psi(z)).
\]

Here, \( \psi(\cdot) \) is called the risk function, which is common in the proportional hazards rate analysis literature; see Fan, Gijbels, and King (1997). Thus, the major interest for the PMRL model is to estimate the risk function \( \psi(x) \). We consider two estimation approaches: Local likelihood approach and local likelihood for exponential regression fitting approach. The former discussed in Section 2 is mainly inspired by Fan, Gijbels, and King (1997) for the Cox's model and the latter described in Section 3 comes from Prentice (1973) for the exponential regression fitting.

In many applications, the survival times of studied subjects are not always fully observed, but subject to right-censoring, due to the termination of the study or early withdrawal from the study. Consider the bivariate data \( \{(T_i, Z_i); i = 1, \ldots, n\} \) which form an i.i.d. sample from the population \((T, Z)\). Under the independent censoring model in which we have the i.i.d. censoring times \( C_1, \ldots, C_n \) that are independent of the survival times given the covariates, one only observes the censored data \( Y_i = \min(T_i, C_i) \) and \( \delta_i = I(T_i \leq C_i) \) as well as the associated covariate \( Z_i \). For notational simplicity, it is assumed throughout this paper that the random variables \( T \) and \( C \) are positive and continuous and the covariate \( Z \) remains constant over time.

In this paper, we consider the problem of estimation of risk function in the model (1.3). Section 2 deals with the situation where the baseline mean residual life function is parametrized and discusses inference based on the local likelihood. Also, the consistency and asymptotic normality of the resulting estimator are investigated. For a nonparametric baseline mean residual life function, in Section 3, a simple method based on the exponential regression approach is proposed to estimate risk function and its large sample properties are also derived. In Section 4, we investigate the finite sample behavior of the local likelihood method based on the exponential regression procedure.
2. LOCAL LIKELIHOOD METHOD

2.1. Local likelihood estimation

Let \( s(t \mid z) \) denote the conditional density function of \( T \) given \( Z = z \), and let \( S(t \mid z) = P(T > t \mid Z = z) \) be its conditional survivor function. Under the independent censoring scheme and the usual assumption about uninformative censoring, the conditional likelihood function is

\[
L = \prod_u s(Y_i \mid Z_i) \prod_c S(Y_i \mid Z_i),
\]

where \( \prod_u \) and \( \prod_c \) denote respectively the product over uncensored and censored individuals. This kind of likelihood can be seen often in survival analysis literature (Cox and Oakes, 1984, p. 81).

We use the following notation. Denote by \( E_0(t) = \int_0^t e_0^{-1}(u)\,du \), \( e_1(t \mid z) = \frac{\partial}{\partial z} e(t \mid z) \), and \( \xi_0(t) = \log \{ e_0(t) / e_0(0) \} \). Assume temporarily that the baseline mean residual function \( e_0(t) \) has been parametrized as \( e_0(t) = e_0(t; \theta) \) and that \( \psi(z) \) has been parametrized as \( \psi(z) = \psi(z; \beta) \). Therefore, under model (1.3), we have

\[
\log(L) = \sum_{i=1}^n \left[ \delta_i \xi(Y_i; \theta \mid Z_i) - \exp \{-\psi(Z_i; \beta)\} E_0(Y_i; \theta) - \zeta_0(Y_i; \theta) \right],
\]

which

\[
\xi(t; \theta \mid z) = \log \frac{e_1(t; \theta) + \psi^{-1}(z)}{e_0(t; \theta)} \quad \text{and} \quad e_1(t; \theta) = \frac{\partial}{\partial t} e_0(t; \theta).
\]

Maximization of (2.1) leads to the maximum likelihood estimate (MLE) of \( \theta \) and \( \beta \).

Suppose now that \( \psi(z) \) is in nonparametric nature, and assume that the \( p \)th order derivative of \( \psi(Z) \) at \( z \) exists for \( p \geq 1 \). Then by the Taylor's expansion, locally around \( z \), \( \psi(Z) \) can be approximated by

\[
\psi(Z) \approx Z^T \beta,
\]

where \( \beta = (\beta_0, \beta_1, \ldots, \beta_p)^T \) and \( Z = (1, Z - z, \ldots, (Z - z)^p)^T \).

Note that \( \beta \) depends on \( z \). Using the local model (2.2), for given data \( ((\delta_i, Y_i, Z_i)) \), we obtain the local log-likelihood

\[
\ell_n(\beta; \theta) = \sum_{i=1}^n \left[ \delta_i \xi_i(Y_i; \beta, \theta \mid Z_i) - \exp \{-Z_i^T \beta\} E_0(Y_i; \theta) - \zeta_0(Y_i; \theta) \right] K_h(Z_i - z) / n,
\]

where

\[
K_h(Z_i - z) / n
\]
where $Z_i = (1, Z_i - z, \ldots, (Z_i - z)^p)^T$.

$$
\xi_i(t; \beta, \theta | Z) = \log \left[ e_{0}(t; \theta) + \exp (-Z_i^T \beta) \right] - \log e_{0}(t; \theta).
$$

$h = h_n$ is the bandwidth parameter, $K(\cdot)$ is a kernel function and $K_h(\cdot) = h^{-1}K(\cdot/h)$. The local MLEs $\hat{\beta}$ and $\hat{\theta}$ are the roots of the local likelihood equations

$$
\sum_{i=1}^{n} K_h(Z_i - z) \left[ \frac{\delta_i}{m_i(\beta, \theta)} \right] \exp (-Z_i^T \beta) Z_i = 0, \quad (2.4)
$$

and

$$
\sum_{i=1}^{n} K_h(Z_i - z) \left[ \delta_i \xi_i'(Y_i; \beta, \theta | Z) - \exp (-Z_i^T \beta) E_0'(Y_i; \theta) - \zeta_0'(Y_i; \theta) \right] = 0. \quad (2.5)
$$

Here, $m_i(\beta, \theta) = e_{1}(Y_i; \theta) + \exp (-Z_i^T \beta)$, and $\xi_i'(t; \beta, \theta | Z)$ and $e_{0}'(t; \theta)$ denote the gradient vector of $\xi_i(t; \beta, \theta | Z)$ and $e_{0}(t; \theta)$ with respect to parameter vector $\theta$, respectively. For the computational issue of finding the roots of the local likelihood equations (2.4) and (2.5), we refer to the paper by Cai, Fan, and Li (2000) for the details.

It is not clear whether equations (2.4) and (2.5) have unique solutions $\hat{\beta}$ and $\hat{\theta}$. In what follows, let $\hat{\beta}$ and $\hat{\theta}$ be the solutions of (2.4) and (2.5), if they exist. Then, a local polynomial estimator of $\psi^{(v)}(z)$, for $v = 0, \ldots, p$, is

$$
\hat{\psi}^{(v)}(z) = v! \hat{\beta}_v.
$$

As $z$ varies across the range of a set, then an estimated (derivative) curve on that set is obtained.

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

$$
e''_{n}(\beta, \theta) = -n^{-1} \sum_{i=1}^{n} K_h(Z_i - z) \exp (-Z_i^T \beta) \times \begin{bmatrix}
    a_i \left( \frac{Z_i}{-b_i/a_i} \right)^{\beta_2} - \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
    b_i, b_i^T/a_i + C_i \exp (Z_i^T \beta)
\end{bmatrix},
$$

where $A^{\otimes 2}$ denotes $A A^T$ for any vector or matrix $A$,

$$
a_i = E_0(Y_i; \theta) - \frac{\delta_i e_{1}(Y_i; \theta)}{m_i^2(\beta, \theta)}, \quad b_i = E_0'(Y_i; \theta) + \frac{\delta_i e_{1}'(Y_i; \theta)}{m_i^2(\beta, \theta)},
$$

and

$$
C_i = \delta_i \xi_i''(Y_i; \beta, \theta | Z_i) - \exp (-Z_i^T \beta) E_0''(Y_i; \theta) - \zeta_0''(Y_i; \theta).
$$
If

\[ a_i > 0, \quad \text{and} \quad b_i b_i^T / a_i + C_i \exp \{ Z_i^T \beta \} \leq 0, \]

then the local log-likelihood \( L_n(\beta; \theta) \) is strictly concave, which implies that the solution to the equations (2.4) and (2.5) is unique and must be consistent.

2.2. Asymptotic properties

Let \( \beta_0 = (\psi(z), \ldots, \psi^{(p)}(z)/p!)^T \) be the true parameter in the model (2.2), and similarly let \( \theta_0 \) be the true parameter of \( e_0(t; \theta) \). In order to obtain consistency for the local MLEs \( \hat{\beta} \) and \( \hat{\theta} \), the true parameters \( \beta_0 \) and \( \theta_0 \) must solve the asymptotic counterpart of (2.4) and (2.5), respectively, which are the Barlett identities of the local log-likelihood. The first local likelihood equation implies that \( \Psi(z) \) has to satisfy the following equation

\[
E \left[ \frac{\delta}{e_1(Y) + \Psi^{-1}(Z)} \bigg| Z \right] = E \left[ E_0(Y) \big| Z \right]. \tag{2.6}
\]

Proposition 1 (below) shows that under certain conditions, (2.6) holds true and the second local likelihood score equation is justified in Proposition 2 (below). The proof of Proposition 1 is relegated to the Appendix but the proof of Proposition 2 is omitted here since it is similar to that of Proposition 2 in Fan, Gijbels, and King (1997).

**Proposition 1.** Suppose that \( \theta_0 \) is an interior point of the parameter space.

(i) If

\[
E \left[ \left| \frac{\delta}{e_1(Y; \theta_0) + \Psi^{-1}(Z)} \right| \bigg| Z = z \right] < \infty,
\]

and

\[
E \left[ \left| E_0(Y; \theta_0) \right| \bigg| Z = z \right] < \infty,
\]

then (2.6) holds true.

(ii) If

\[
E \left[ \left| \delta \xi(Y; \theta_0|Z) \right| \bigg| Z = z \right] < \infty,
\]

\[
E \left[ \left| E_0'(Y; \theta_0) \right| \bigg| Z = z \right] < \infty,
\]

and

\[
E \left[ \left| \xi_0'(Y_i; \theta_0) \right| \bigg| Z = z \right] < \infty.
\]
then
\[ E \left[ \delta \xi'(Y; \theta_0 | Z) - E_0'(Y; \theta_0) \Psi^{-1}(Z) - \zeta_0'(Y; \theta_0) \right] Z = z \right] = 0. \] (2.7)

(iii) If
\[ E \left[ \left\| \delta [\log e_0(Y; \theta_0)]' / (e_1(Y; \theta_0) + \Psi^{-1}(Z)) \right\| Z = z \right] < \infty, \]
and
\[ E \left[ \left\| E_0'(Y; \theta_0) \right\| Z = z \right] < \infty, \]
then
\[ E \left[ \frac{\delta [\log e_0(Y; \theta_0)]'}{e_1(Y; \theta_0) + \Psi^{-1}(Z)} + E_0'(Y; \theta_0) \right] Z = z \right] = 0. \] (2.8)

**Proposition 2.** If \( \theta_0 \) is an interior point of the parameter space, and
\[ E \left[ \left\| \delta [\xi'(Y; \theta_0 | Z)]^{\otimes 2} \right\| Z = z \right] < \infty, \]
\[ E \left[ \left\| E_0'(Y; \theta_0) \right\| Z = z \right] < \infty, \]
\[ E \left[ \left\| \delta \xi''(Y; \theta_0 | Z) \right\| Z = z \right] < \infty, \]
and \[ E \left[ \left\| \zeta_0''(Y; \theta_0) \right\| Z = z \right] < \infty, \]
then
\[ E \left[ \delta \xi''(Y; \theta_0 | Z) - E_0''(Y; \theta_0) \Psi^{-1}(Z) - \zeta_0''(Y; \theta) \right] Z \]
\[ = -E \left[ \delta \left[ \xi'(Y; \theta_0 | Z)] \right]^{\otimes 2} \right| Z. \]

Before we state the main results of this section, we introduce the following notation and list some technical assumptions. Set
\[ H = \text{diag}(1, h, \ldots, h^p)^T, \quad u = (1, u, \ldots, u^p)^T, \]
\[ S_0(z; \theta_0) = \int_{-\infty}^{\infty} E \left[ \delta \left( \frac{\psi^{(u)}(Y; \theta_0 | Z) + 1}{-\xi'(Y; \theta_0 | Z)} \right)^{\otimes 2} \right| Z = z \right] K(u) \, du, \] (2.9)
and
\[ S_1(z; \theta_0) = \int_{-\infty}^{\infty} E \left[ \delta \left( \frac{\psi^{(u)}(Y; \theta_0 | Z) + 1}{-\xi'(Y; \theta_0 | Z)} \right)^{\otimes 2} \right| Z = z \right] K^2(u) \, du. \]
CONDITION A.

(i) $\theta_0$ is an interior point of the parameter space.

(ii) Functions

\[
E \left\{ \delta \left( e_1(Y; \theta_0) + \Psi^{-1}(Z) \right) \right| Z \right\}, \quad E \left\{ E_0(Y; \theta_0) | Z \right\},
\]

\[
E \left\{ \xi_0'(Y; \theta_0) | Z \right\}, \quad E \left\{ E_0'(Y; \theta_0) | Z \right\}, \quad E \left\{ \delta \xi'(Y; \theta_0 | Z) | Z \right\},
\]

\[
E \left\{ \delta \left[ \log e_0(Y; \theta_0) \right| (e_1(Y; \theta_0) + \Psi^{-1}(Z)) | Z \right\},
\]

\[
E \left\{ E_0''(Y; \theta_0) | Z \right\}, \quad E \left\{ \delta \left[ \xi'(Y; \theta_0 | Z) \right]^{\theta_2} | Z \right\},
\]

\[
E \left\{ \delta \xi''(Y; \theta_0 | Z) | Z \right\}, \quad \text{and} \quad E \left\{ \xi_0''(Y; \theta_0 | Z) \right\}
\]

are absolutely finite and continuous at the point $Z = z$.

(iii) For each $z$, there exists functions $N_\varepsilon(t)$ and $M(t)$ with $E N_\varepsilon(Y) < \infty$, $E M(Y) < \infty$, such that

\[
\left| \frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} \xi(t; \theta | z) \right| < N_\varepsilon(t),
\]

\[
\left| \frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} E_0(t; \theta) \right| < M(t), \quad \left| \frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} \xi_0(t; \theta) \right| < M(t),
\]

for all $t$, and all $\theta$ in a neighborhood of $\theta_0$.

(iv) The kernel function $K(\cdot)$ is bounded density with a compact support.

(v) The function $\Psi(\cdot)$ has a continuous $(p+1)$th derivative around the point $z$.

(vi) The density $f(\cdot)$ of $Z$ is continuous at the point $z$ and $f(z) > 0$.

(vii) $n h \to \infty$.

CONDITION B.

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

\[
E \left\{ E_0(Y; \theta_0)^{2+\eta} | Z \right\}, \quad E \left\{ \| E_0'(Y; \theta_0) \|^{2+\eta} | Z \right\},
\]

\[
E \left\{ \| \xi'(Y; \theta_0 | Z) \|^{2+\eta} | Z \right\}, \quad E \left\{ \left( \frac{1}{e_1(Y; \theta_0) + \Psi^{-1}(Z)} \right)^{2+\eta} \right\}.
\]
Estimate of Biometric Function

\[ E \left[ |\zeta_\theta^2(Y; \theta_0) + \delta e_1'(Y; \theta_0) \left| e_1(Y; \theta_0) + \psi^{-1}(Z) \right| \right] \text{ and } E \left( \frac{\delta e_1'(Y; \theta_0)}{e_1(Y; \theta_0) + \psi^{-1}(Z)} \right) \]

are continuous at the point \( Z = z \).

(ii) \( n h^{2p+1} \) is bounded.

We now state the asymptotic consistency in Theorem 3 and the asymptotic normality in Theorem 4, but their proofs are relegated to the Appendix.

**Theorem 3.** Under Condition A, there exists a solution \( \hat{\beta} \) and \( \hat{\theta} \) to the local likelihood equations (2.4) and (2.5) such that

\[ \sqrt{n h} \left( \frac{H(\hat{\beta} - \beta_0) - b_n(z)}{\theta - \theta_0} \right) \xrightarrow{D} N(0, \sigma^2(z)), \quad (2.10) \]

where

\[ b_n(z) = \frac{\psi^{(p+1)}(z)}{(p+1)!} h^{p+1} v^{-1}(K) \int u^{p+1} K(u) \, du, \quad (2.11) \]

\[ \sigma^2(z) = f^{-1}(z) S_0(z; \theta_0)^{-1} S_1(z; \theta_0) S_0(z; \theta_0)^{-1}, \]

and

\[ v(K^j) = \int u u^T K^j(u) \, du, \quad j = 1, 2. \]

**Remark 1.** Note that the bias term \( b_n(z) \) of \( \hat{\beta} \) given in (2.11) has the same expression as that of the least-squares nonparametric regression (Fan and Gijbels, 1996, p. 62). It is not surprising to see that the bias term is independent of the model because the bias term comes from the approximation error.

As an application of Theorem 4, the asymptotic normality of the local polynomial estimation of \( \psi^{(p)}(\cdot) (v = 0, \ldots, p) \) can be obtained easily from (2.10), which is stated as follows.

**Corollary 5.** Under the conditions of Theorem 4, and if \( K(\cdot) \) is symmetric, then

\[ \sqrt{n h} \left| H(\hat{\beta} - \beta_0) - b_n(z) \right| \xrightarrow{D} N \left( 0, \frac{\sigma^2(z; \theta_0)}{f(z)} v^{-1}(K) v(K^2) v^{-1}(K) \right), \]
where
\[
\sigma^2(z; \theta_0) = E \left\{ \frac{\delta}{(\delta(Y; \theta_0) \Psi(z))^2} \middle| Z = z \right\}^{-1}.
\]

Furthermore,
\[
\sqrt{n}h^3 \left\{ \hat{\Psi}(z) - \psi'(z) - \frac{\mu_4(K)}{6 \mu_2(K)} \psi''(z) h^2 \right\} \overset{\mathcal{D}}{\rightarrow} N \left\{ 0, \frac{\sigma^2(z; \theta_0) \nu_2(K)}{f(z) \mu_2^2(K)} \right\},
\]

(2.12)

\[
\text{where } \mu_j(K) = \int u^j K(u) \, du \text{ and } \nu_j(K) = \int u^j K^2(u) \, du.
\]

As a consequence of (2.13), the theoretical optimal bandwidth, which minimizes the asymptotic weighted mean integrated square error,
\[
\int \left[ \left\{ \frac{\mu_4(K)}{6 \mu_2(K)} \psi''(z) h^2 \right\}^2 + \frac{\sigma^2(z; \theta_0) \nu_2(K)}{n h^3 f(z) \mu_2^2(K)} \right] w(z) \, dz,
\]
is given by
\[
\hat{h}_{opt} = \left[ \frac{27 \nu_2(K) \int \sigma^2(z; \theta_0) w(z) f(z) \, dz}{\mu_2^4(K) \int \{ \psi''(z) \}^2 w(z) \, dz} \right]^{1/7} n^{-1/7},
\]
where \( w(\cdot) \) is a nonnegative and known function.

**Remark 2.** Note that when \( \theta_0 \) is known, the local maximum likelihood estimator of \( \beta \) is obtained by maximizing (2.3) with respect to \( \beta \). It can be showed that the bias and variance for the resulting estimator are the same as those given in Corollary 5. This leads to the conclusion that under the conditions given in Corollary 5, \( \hat{\Psi}(z) \) is adaptive in the sense that \( \psi(z) \) can be estimated as well as in the case that \( \theta_0 \) is known.

**Remark 3.** For the parametric linear model \( \psi(z; \beta) = Z^T \beta \), the MLE of \( \beta \) is obtained by maximizing directly the log-likelihood (2.1). In that case, it can be showed along the same lines as the proof of Theorem 4 that
\[
\sqrt{n} \left( \hat{\beta} - \beta_0 \right) \overset{\mathcal{D}}{\rightarrow} N \left\{ 0, \Sigma^{-1} \right\},
\]

where
\[
\Sigma = E \left\{ \delta \left( \frac{Z}{\delta(Y; \theta_0) \psi(z; \beta_0)} \right)^2 \right\},
\]
and
\[
\delta \left( \frac{Z}{\delta(Y; \theta_0) \psi(z; \beta_0)} \right)^2 = \delta \left( \frac{Z}{\delta(Y; \theta_0) \psi(z; \beta_0)} \right)^2.
\]
3. AN EXPONENTIAL REGRESSION APPROACH

In practice, it is difficult to find the solutions of (2.4) and (2.5) since the form of the baseline mean residual function is, in general, unknown. To overcome this difficulty, we propose an approach based on the exponential regression procedure proposed by Prentice (1973) for the survival analysis. Our estimator is the solution of the following equation

\[
U_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \exp \left( -Z_i^T \beta \right) Y_i Z_i K_h(Z_i - z) \sum_{i=1}^{n} \delta_i K_h(Z_i - z) \\
\left/ \sum_{i=1}^{n} \exp \left( -Z_i^T \beta \right) Y_i K_h(Z_i - z) - \frac{1}{n} \sum_{i=1}^{n} \delta_i Z_i K_h(Z_i - z) \right. \\
= 0. \quad (3.1)
\]

It is easy to see that \( U_n(\beta) \) is non-positively definite and (3.1) has a unique solution, denoted by \( \hat{\beta}_s \). If it is known that \( S_0(t) = \exp(-t/\mu_0) \), then the equation (3.1) is the local MLE of this exponential regression model and \( \hat{\beta}_s \) is asymptotically efficient (Prentice, 1973). For the general PMRL regression model (1.3), the estimating equation (3.1) is still asymptotically consistent by the fact that \( U_n(\beta) \to 0 \) in probability. Furthermore, we have asymptotic normality for \( \hat{\beta}_s \) (see Theorem 6).

Note that the function \( \psi(z) \) is not directly estimatible since (3.1) does not involve the intercept \( \beta_0 = \psi(z) \) due to the cancellation. It is not surprising since from the PMRL model (1.3), \( \psi(z) \) is only identifiable to within a constant factor. The identifiability of \( \psi(z) \) is ensured by imposing the condition \( \psi(0) = 0 \). Then the function \( \hat{\psi}(z) = \int_0^z \hat{\psi}(t) dt \) can be estimated by

\[
\hat{\psi}(z) = \int_0^z \hat{\psi}(t) dt.
\]

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

Since the equation (3.1) does not involve \( \beta_0 = \psi(z) \), we need to re-write (3.1). To this end, let

\[
C = \begin{pmatrix} 0 & 1_p \end{pmatrix}, \quad H^* = CHC^T, \quad \beta^* = H^* C \beta, \quad \hat{\beta}^* = H^* C \hat{\beta},
\]

and

\[
Z^* = H^{-1} C Z.
\]

Correspondingly, let

\[
\beta_0^* = H^* C \beta_0, \quad \text{and} \quad u^* = (u, \ldots, u^p)^T.
\]
Then, (3.1) becomes

\[ U_n(b^*) = a_{1n} + \frac{a_{2n}}{a_{3n}} \frac{1}{n} \sum_{i=1}^{n} \exp \left( -Z_i^T \beta^* \right) Y_i Z_i^* K_h(Z_i - z) = 0, \quad (3.2) \]

where

\[ a_{1n} = -\frac{1}{n} \sum_{i=1}^{n} \delta_i Z_i^* K_h(Z_i - z), \quad a_{2n} = \frac{1}{n} \sum_{i=1}^{n} \delta_i K_h(Z_i - z), \]

and

\[ a_{3n} = \frac{1}{n} \sum_{i=1}^{n} \exp \left( -Z_i^T \beta^* \right) Y_i K_h(Z_i - z). \]

Let \( \hat{\beta}_s^* \) be the solution of (3.2). Then, \( \hat{\psi}^{(l)}(z) = l! \hat{\beta}_{s,l}^*/h^l, l \geq 1, \) where \( \hat{\beta}_{s,l}^* \) is the \( l \)-th element of \( \hat{\beta}_s^* \). Set, for \( j = 1, 2, \)

\[ \nu^*(K_j) = \int u^* u^{*,T} K_j(u) \, du, \quad \text{and} \]

\[ \tilde{\nu}_s(K) = \nu^*(K) - \left( \int u^* K(u) \, du \right)^{\Theta_2}. \]

To establish the asymptotic properties of \( \hat{\beta}_s^* \), the following conditions are needed.

**CONDITION C.**

(i) Functions \( E(Y|Z) \) and \( E(\delta|Z) \) are continuous at \( Z = z \).

(ii) The kernel function \( K(\cdot) \) is bounded density with a compact support.

(iii) The function \( \psi(\cdot) \) has a continuous \( (p+1) \)-th derivative around the point \( z \).

(iv) The density \( f(\cdot) \) of \( Z \) is continuous at the point \( z \) and \( f(z) > 0 \).

(v) \( n h \to \infty \).

**CONDITION D.**

(i) There exists an \( \eta > 0 \) such that \( E(|Y|^{2+\eta}|Z) \) is finite and continuous at \( Z = z \).

(ii) \( n h^{2p+3} \) is bounded.

**THEOREM 6.** Under Condition C, we have

\[ \hat{\beta}_s^* - \beta_0^* \overset{P}{\to} 0. \]

Suppose in addition that Condition D holds, then

\[ \sqrt{n h} \left\{ \hat{\beta}_s^* - \beta_0^* - b_n^*(z) \right\} \overset{D}{\to} N \left\{ 0, \sigma_n^*(z)^2 \right\}. \]
where
\[ b^*_n(z) = \frac{\psi^{(p+1)}(z)}{(p+1)!} \int K(u) \, du, \tag{3.3} \]
and
\[ \sigma^*_n(z)^2 = \frac{E \left[ \left( \frac{\delta - \frac{E(\delta | Z = z)}{E(Y | Z = z)} Y}{E(\delta | Z = z)} \right)^2 \mid Z = z \right] \tilde{v}_*(K) \tilde{v}_*(K^2) \tilde{v}_*(K). \tag{3.4} \]

Note that the brief proof of Theorem 6 can be found nowhere but in the Appendix. It can be seen easily that the asymptotic variance is somewhat different from that in Theorem 4.

4. SIMULATION STUDY

In this section, we conduct a simulation study to illustrate our proposed methods in Sections 2 and 3. We consider the case of a single covariate \( Z \) and study two examples: one with unit exponential as the baseline distribution and the other with Hall-Wellner-type baseline distribution with \( e_0(t) = 3.1 t + 1 \). For Hall-Wellner-type baseline mean residual life distribution, the proportional mean residual life regression model (1.3) is equivalent to the following transformed regression model

\[ \log E_0(T) = -\log(A + \psi^{-1}(Z)) + \log \epsilon, \]

where \( \epsilon \) has the standard exponential distribution and \( A \) is the slope of the baseline mean residual life function.

In this simulation study, the risk function is taken to be \( \psi(z) = z(1 - z) \). The censoring random variable \( C \) is independent of \( Z \) and \( T \) and its distribution is indicated in the following tables. Three sample sizes and three bandwidths are considered, and the Epanechnikov kernel \( K(u) = 0.75(1 - u^2)I(|u| \leq 1) \) is used. The estimator \( \hat{\psi}^{(v)}(\cdot) \) is assessed via the square-Root of Average Square Errors (RASE)

\[ \text{RASE}_v = \left[ \frac{1}{n_{\text{grid}}} \sum_{k=1}^{n_{\text{grid}}} \left( \hat{\psi}^{(v)}(z_k) - \psi^{(v)}(z_k) \right)^2 \right]^{1/2}, \]

where \( \{z_k, k = 1, \ldots, n_{\text{grid}} \} \) are the grid points at which the function \( \psi^{(v)}(\cdot) \) is estimated. The grid point is taken from 0.05 to 0.90 with increment 0.01. The covariate \( Z \) is generated from \( U(0, 1) \) and the censoring random variable \( C \) is simulated from \( U(0, 5) \) and \( U(0, 10) \) for the standard exponential baseline residual mean function, and \( U(0, 1) \) and \( U(0, 5) \) for the Hall-Wellner baseline residual mean function. We simulate data and compute
Table 1.
Simulation results for the exponential distribution with 100 replicates and three sample sizes and three bandwidths

<table>
<thead>
<tr>
<th></th>
<th>n = 250</th>
<th>n = 400</th>
<th>n = 700</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Censoring: (U(0, 5)), Censoring rate: 15%-30%</td>
<td>Censoring: (U(0, 10)), Censoring rate: 15%-20%</td>
<td></td>
</tr>
<tr>
<td>(h = 0.5)</td>
<td>0.218 ± 0.164, 0.184</td>
<td>0.159 ± 0.120, 0.124</td>
<td>0.113 ± 0.089, 0.097</td>
</tr>
<tr>
<td>(h = 1.0)</td>
<td>0.147 ± 0.105, 0.116</td>
<td>0.120 ± 0.091, 0.097</td>
<td>0.095 ± 0.070, 0.080</td>
</tr>
<tr>
<td>(h = 1.5)</td>
<td>0.131 ± 0.098, 0.107</td>
<td>0.116 ± 0.090, 0.094</td>
<td>0.088 ± 0.064, 0.082</td>
</tr>
</tbody>
</table>

Figure 1. The estimated curves of \(\psi(z)\) and \(\psi'(z)\) for the exponential baseline residual mean life function. True function — solid curve; estimated function — dashed curve.

the estimate of \(\psi^{(0)}(z)\) based on (3.2). For both examples, we repeat the simulation 100 times.

For the standard exponential baseline residual mean function, Table 1 reports the results on RASEs of the local likelihood estimate of the risk function \(\psi(\cdot)\) based on 100 replicates. In each cell, the first, second and third numbers represent the mean, standard deviation and median of RASE for three sample sizes \(n = 250, 400, 700\), respectively. Figure 1 represents the estimated \(\hat{\psi}(\cdot)\) and \(\hat{\psi}'(\cdot)\) from a random sample with \(n = 500, h = 0.5, Z \sim U(-3, 3), C \sim U(0, 5)\) and the censoring rate 7.8%, which shows that the local estimation method performs reasonably well.

For the Hall-Wellner baseline residual mean function, we consider three sample sizes \(n = 500, 700, 1000\). A summary of the simulation results
Table 2.
Simulation results for the Hall-Wellner distribution with 100 replicates and three sample sizes and three bandwidths

<table>
<thead>
<tr>
<th></th>
<th>n = 500</th>
<th>n = 700</th>
<th>n = 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>h = 0.5</td>
<td>0.111 ± 0.093, 0.087</td>
<td>0.101 ± 0.070, 0.084</td>
<td>0.1072 ± 0.0745, 0.1038</td>
</tr>
<tr>
<td>h = 1.0</td>
<td>0.113 ± 0.069, 0.106</td>
<td>0.094 ± 0.058, 0.089</td>
<td>0.091 ± 0.053, 0.093</td>
</tr>
<tr>
<td>h = 1.5</td>
<td>0.137 ± 0.080, 0.128</td>
<td>0.124 ± 0.082, 0.118</td>
<td>0.140 ± 0.071, 0.146</td>
</tr>
</tbody>
</table>

Censoring: U(0, 5), Censoring rate: 9%-20%

<table>
<thead>
<tr>
<th></th>
<th>n = 500</th>
<th>n = 700</th>
<th>n = 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>h = 0.5</td>
<td>0.173 ± 0.118, 0.159</td>
<td>0.147 ± 0.097, 0.122</td>
<td>0.141 ± 0.104, 0.119</td>
</tr>
<tr>
<td>h = 1.0</td>
<td>0.160 ± 0.123, 0.131</td>
<td>0.114 ± 0.076, 0.102</td>
<td>0.121 ± 0.081, 0.107</td>
</tr>
<tr>
<td>h = 1.5</td>
<td>0.185 ± 0.121, 0.186</td>
<td>0.148 ± 0.096, 0.141</td>
<td>0.147 ± 0.087, 0.144</td>
</tr>
</tbody>
</table>

Figure 2. The estimated curves of $\psi(z)$ and $\psi'(z)$ for the Hall-Wellner baseline residual mean life function. True function – solid curve; estimated function – dashed curve.

on 100 RASE-values of the local estimate of $\psi(\cdot)$ based on 100 simulations is given in Table 2. The estimated curves of $\psi(\cdot)$ and $\psi'(\cdot)$ are depicted in Figure 2 for $n = 700$, $h = 1.0$, $Z \sim U(-3, 3)$, and $C \sim U(0, 5)$ with the censoring rate 5.8%. Note that the estimation for the risk function seems to be somewhat underestimated by the trapezoidal rule of Tibshirani and Hastie (1987).

From Tables 1 and 2, one can observe that the median of RASE is always less than the mean. Furthermore, as we can see from Tables 1 and 2 and Figures 1 and 2, the estimate for the exponential baseline mean residual function performs better than that for the Hall-Wellner. This is perfectly natural, since the former estimate is efficient when the underlying baseline distribution is exponential.
APPENDIX: PROOFS

We still use the same notation as in Sections 2 and 3.

**Proof of Proposition 1.** We employ the martingale technique. Let \( N(t) = I(Y \leq t, \delta = 1) \), \( X(t) = I(Y \geq t) \) and let \( \mathcal{F}_t = \sigma(Z, N(u), X(u), 0 \leq u \leq t) \) be the history up to time \( t \). Set

\[
M(t) = N(t) - \int_0^t X(u) \frac{e_1(u; \theta_0) + \Psi^{-1}(Z)}{e_0(u; \theta_0)} \, du. \tag{A.1}
\]

Then, \( M(t) \) is an \( \mathcal{F}_t \)-martingale (Fleming and Harrington, 1991, p. 19). Since

\[
\begin{align*}
\int_0^\infty \frac{1}{e_1(t; \theta_0) + \Psi^{-1}(Z)} \, dM(t) &= \int_0^\infty \frac{1}{e_1(t; \theta_0) + \Psi^{-1}(Z)} \, dN(t) - \int_0^\infty I(Y \geq t) \frac{1}{e_0(t; \theta_0)} \, dt \\
&= \frac{1}{e_1(Y; \theta_0) + \Psi^{-1}(Z)} - E_0(Y; \theta_0)
\end{align*}
\]

and \( \{e_1(t; \theta_0) + \Psi^{-1}(Z)\}^{-1} \) is \( \mathcal{F}_t \)-measurable, (2.6) follows by taking the conditional expectation of the above equality with respect to \( \delta \) and \( Y \) given \( Z = z \). To prove (2.7), note that

\[
\int_0^\infty \xi'(t; X(t; \theta_0)|Z) \, dM(t) = \delta \xi'(Y; \theta_0|Z) - E_0'(Y; \theta_0) \Psi^{-1}(Z) - \zeta_0(Y; \theta_0).
\]

Since \( \xi'(Y; \theta_0|Z) \) is \( \mathcal{F}_t \)-measurable, (2.7) follows by taking the conditional expectation of the above equality with respect to \( \delta \) and \( Y \) given \( Z = z \). (2.8) follows immediately from the facts that

\[
\int_0^\infty \frac{[\log e_0(t; \theta_0)]'}{e_1(t; \theta_0) + \Psi^{-1}(Z)} \, dM(t) = \delta \frac{[\log e_0(Y; \theta_0)]'}{e_1(Y; \theta_0) + \Psi^{-1}(Z)} + E_0'(Y; \theta_0)
\]

and \( \{\log e_0(Y; \theta_0)\}' / \{e_1(Y; \theta_0) + \Psi^{-1}(Z)\} \) is \( \mathcal{F}_t \)-measurable. This proves the proposition. \( \square \)

**Proof of Theorem 3.** Let \( \alpha = \mathbf{H}(\beta - \beta_0), \hat{\alpha} = \mathbf{H}(\hat{\beta} - \beta_0) \) and \( U_i = \mathbf{H}^{-1} Z_i \). Denote by \( m_i(\alpha, \theta) = e_1(Y_i; \theta) + \exp(-Z_i^T \beta_0 - U_i^T \alpha) \) and \( \xi_i(\alpha, \theta|Z_i) = \log m_i(\alpha, \theta) - \log e_0(Y_i; \theta) \) . Put

\[
\ell_n(\alpha, \theta) = \sum_{i=1}^n \left[ \delta_i \delta_i(\alpha, \theta|Z_i) - \exp(-Z_i^T \beta_0 - U_i^T \alpha) E_0(Y_i; \theta) - \zeta_0(Y_i; \theta) \right] \\
\times K_h(Z_i - z)/n. \tag{A.2}
\]
Then the problem is equivalent to showing that there exists a solution \( \hat{\alpha} \) and \( \hat{\theta} \) to the likelihood equation

\[
\begin{align*}
\frac{\partial \ell_n(\alpha, \theta)}{\partial \alpha} &= n^{-1} \sum_{i=1}^{n} \left[ E_0(Y_i; \theta) - \frac{s_i}{m_i(\alpha, \theta)} \right] \exp \left( -Z_i^T \beta_0 - U_i^T \alpha \right) \\
&\quad \times K_h(Z_i - z) U_i = 0 \\
\frac{\partial \ell_n(\alpha, \theta)}{\partial \theta} &= n^{-1} \sum_{i=1}^{n} K_h(Z_i - z) \left[ \delta_i \xi'_i(Y_i; \alpha, \theta | Z_i) - \xi'_0(Y_i; \theta) \\
&\quad - \exp \left( -Z_i^T \beta_0 - U_i^T \alpha \right) \xi'_0(Y_i; \theta) \right] = 0,
\end{align*}
\]

such that

\[
\hat{\alpha} \xrightarrow{p} 0 \quad \text{and} \quad \hat{\theta} - \theta_0 \xrightarrow{p} 0.
\]

To this end, let \( y = (\alpha^T, \theta^T)^T \) and \( y_0 = (0^T, \theta_0^T)^T \). Denote by \( S_\epsilon \) the sphere centered at \( y_0 \) with radius \( \epsilon \). If it is showed that for any sufficiently small \( \epsilon \), the probability

\[
\sup_{y \in S_\epsilon} \ell_n(y) \leq C_n(y_0) = \ell_n(0, \theta_0)
\]

(A.4)

tends to one, then \( \ell_n(y) \) has a local maximum in the interior of \( S_\epsilon \). Since the likelihood equation (A.3) must be satisfied at a local maximum, it then follows that for any \( \epsilon > 0 \), with probability tending to one, the likelihood equation has a solution \( \{\hat{\alpha}(\epsilon), \hat{\theta}(\epsilon)\} \) within \( S_\epsilon \). Let \( (\hat{\alpha}, \hat{\theta}) \) be the closest root to \( y_0 \). Then

\[
P \left\{ \|\hat{\alpha}\|^2 + \|\hat{\theta} - \theta_0\|^2 \leq \epsilon \right\} \to 1.
\]

This in turn implies that

\[
H(\hat{\beta} - \beta_0) \xrightarrow{p} 0 \quad \text{and} \quad \hat{\theta} - \theta_0 \xrightarrow{p} 0.
\]

To finish the proof, it suffices to establish (A.4). To this effect, denote by \( y_j \) and \( y_{0j} \) the \( j^{th} \) elements of \( y \) and \( y_0 \), respectively. By the Taylor's expansion around the point \( y_0 \), we have

\[
\ell_n(y) - \ell_n(y_0) = \ell'_n(y_0)^T (y - y_0) + \frac{1}{2} (y - y_0)^T \ell''_n(y_0)(y - y_0) + R_n(y^*)
\]

(A.5)

with \( y^* \) lying between \( y_0 \) and \( y \), where

\[
R_n(y) = \frac{1}{6} \sum_{j,k,l} (y_j - y_{0j})(y_k - y_{0k})(y_l - y_{0l}) \frac{\partial^3}{\partial y_j \partial y_k \partial y_l} \ell_n(y).
\]

First, by recalling the definition of \( \beta_0 \) and (A.3), we have

\[
\ell'_n(y_0)
\]
\[
\frac{E_0(Y; \theta; \theta_0) - \frac{\delta_{\theta_0}}{m_0(\theta_0)}}{\varphi(z)} + \frac{\varphi_{\theta_0}(z)}{m_0(\theta_0)} + \frac{\varphi(z)}{m_0(\theta_0)} - \frac{\varphi_{\theta_0}(z)}{m_0(\theta_0)} = \sum_{i=1}^{n} \left( E_0(Y_i; \theta_0) - \frac{\delta_{\theta_0}}{m_0(\theta_0)} \right) U_i \quad \times \quad \frac{\varphi(z)}{m_0(\theta_0)} - \frac{\varphi_{\theta_0}(z)}{m_0(\theta_0)} - \frac{\varphi(z)}{m_0(\theta_0)} - \frac{\varphi_{\theta_0}(z)}{m_0(\theta_0)} \right) 
\]

where \( \varphi(z) = \int \mu(K) dK \). This, in conjunction with (2.6) and (2.7), implies that \( \ell_n^{(\eta)}(\gamma_0) \xrightarrow{p} 0 \). Thus, with probability tending to one,

\[
\| \ell_n^{(\eta)}(\gamma_0)^T (\gamma - \gamma_0) \| \leq \epsilon. \quad (A.6)
\]

Next, we derive the limit of \( \ell_n^{(\eta)}(\gamma_0) \). To this end, some algebraic calculation yields

\[
\ell_n^{(\eta)}(\gamma_0) = -n^{-1} \sum_{i=1}^{n} K_h(z_i - z) \exp (-Z_i^T \beta_0) \left( E_0(Y; \theta_0) - \frac{\delta_{\theta_0}}{m_0(\theta_0)} \right) U_i \left( \frac{\delta_{\theta_0}}{m_0(\theta_0)} + E_0(Y; \theta_0) \right) - \left( \frac{\delta_{\theta_0}}{m_0(\theta_0)} + E_0(Y; \theta_0) \right)^T U_i \right) 
\]

where \( m(\theta) = e_1(Y; \theta) + \Psi^{-1}(Z) \), and

\[
\tau(Y; \theta_0 | z) = E_0(Y; \theta_0) + \frac{\delta_{\theta_0}}{m_0(\theta_0)} \nu(K) - \tau(Y; \theta_0 | z)^T \Psi^{-1}(Z) \left( E_0(Y; \theta_0) - \Psi^{-1}(Z) \delta_{\theta_0}(Y; \theta_0 | Z) - \delta_{\theta_0}(Y; \theta_0 | Z) \right) \]

It follows by (2.6), (2.7) and Proposition 2 that

\[
\ell_n^{(\eta)}(\gamma_0) = -f(z)[S_0(z; \theta_0) - S_2(z; \theta_0)] + o_P(1),
\]

where

\[
S_2(z; \theta_0) = \begin{pmatrix} \eta(Y; \theta_0 | z)^T & \eta(Y; \theta_0 | z) \end{pmatrix}
\]
 Estimate of Biometric Function

\[ \eta(Y; \theta_0 | z) = \Psi^{-1}(z) \mu(K) E \left[ \frac{\delta \{ \log \epsilon_0(Y; \theta_0) \}'}{\epsilon_1(Y; \theta_0) + \Psi^{-1}(Z)} + E_0(Y; \theta_0) \right]_{Z = z} \]

By Proposition 1 (iii), \( \eta(Y; \theta_0 | z) = 0 \). Thus, we have
\[ \ell_n'(\gamma_0) = -f(z)S_0(z; \theta_0) + o_P(1), \quad (A.7) \]
and hence, with probability tending to one,
\[ (\gamma - \gamma_0)^T \ell_n''(\gamma_0)(\gamma - \gamma_0) < -a f(z) \epsilon^2, \quad \text{for all } \gamma \in S_\epsilon, \quad (A.8) \]
where \( a \) is the smallest eigenvalue of \( S_0(z; \theta_0) \). By Condition A (iii),
\[ |R_n(\gamma)| \leq C \epsilon^3 n^{-1} \sum_{i=1}^n (N_{Y_i} + M_{Y_i}) = C \epsilon^3 (EN_{Y_i} + EM_{Y_i}) + o_P(1) \quad (A.9) \]
for some constant \( C > 0 \). Substituting (A.6), (A.8) and (A.9) into (A.5), we conclude with probability tending to one when \( \epsilon \) is small enough,
\[ \ell_n(\gamma) \leq \ell_n(\gamma_0) \}
This completes the proof of Theorem 1.

Proof of Theorem 4. We continue to use the notation introduced in the proof of Theorem 1. By the Taylor's expansion and Condition A(iii), we have
\[ 0 = \ell_n' (\tilde{\gamma}) = \ell_n'(\gamma_0) + \ell_n''(\gamma_0)(\tilde{\gamma} - \gamma_0) + O_P(\|\tilde{\gamma} - \gamma_0\|^2). \]

Hence, by (A.7),
\[ (\tilde{\gamma} - \gamma_0) = -\{\ell_n''(\gamma_0) + o_P(1)\}^{-1} \ell_n'(\gamma_0) \]
\[ = \{f(z)S_0(z; \theta_0) + o_P(1)\}^{-1} \ell_n'(\gamma_0). \quad (A.10) \]

Thus, it suffices to establish the asymptotic normality of \( \ell_n'(\gamma_0) \). We first compute the mean and variance of \( \ell_n'(\gamma_0) \). To this end, by the Taylor's expansion again,
\[ \exp\{-\psi(Z_i)\} \leq \exp(-Z_i^T \beta_0) = \Psi^{-1}(z) \frac{\psi^{(p+1)}(z)}{(p+1)!} (Z_i - z)^{p+1} (1 + o_P(1)). \quad (A.11) \]
By (A.2), (2.6) and (2.7), \( E \ell_n'(\gamma_0) \) becomes

\[
EK_h(Z - z)[\Psi^{-1}(Z) - \exp(-Z^T \beta_0)] \left( \begin{array}{c}
\frac{-z}{m(\theta_0)} \exp(-Z^T \beta_0) \\
\frac{\delta}{m(\theta_0)} m(0, \theta_0) + E_0'(Y; \theta_0)
\end{array} \right). 
\]

It follows from (A.11) and Proposition 1 (iii) that

\[
E \ell_n'(\gamma_0) = f(z) \tilde{b}_n(z) + o(h^{p+1}), 
\]

where \( \tilde{b}_n(z) \)

\[
= -\Psi^{-1}(z) \frac{\psi^{(p+1)}(z)}{(p + 1)!} \left[ \frac{\delta}{m(\theta_0)} \exp(-Z^T \beta_0) \right] \left[ \int u u^{p+1} K(u) \, du \right]. 
\]

By the same token, the variance of \( \ell_n'(\gamma_0) \) becomes

\[
n \text{Var}[\ell_n'(\gamma_0)] = EK_h^2(Z - z) \left( \begin{array}{c}
\left( E_0(Y; \theta_0) - \frac{\delta}{m(\theta_0)} \Psi^{-1}(Z) \right) U \\
\delta \Psi^{-1}(Y; \theta_0) \end{array} \right) \left( \begin{array}{c}
\Psi^{-1}(Z) U \\
-\delta \Psi^{-1}(Y; \theta_0)
\end{array} \right) + o(1). 
\]

Next, the counting process introduced in the proof of Proposition 1 is employed here to simplify the calculation of the first term on the right-hand side of (A.14). Since

\[
\left( \begin{array}{c}
\left( E_0(Y; \theta_0) - \frac{\delta}{m(\theta_0)} \Psi^{-1}(Z) \right) U \\
\delta \Psi^{-1}(Y; \theta_0) \end{array} \right) 
\]

\[
= -\int_0^\infty \left( \frac{\Psi^{-1}(Z) U}{e_t(Y; \theta_0)} \right) dM(t) 
\]

by (A.1), conditioning on \( Z \), one has

\[
< M, M > (t) = X(t) \frac{e_t(Y; \theta_0) + \Psi^{-1}(Z)}{e_0(t; \theta_0)} , 
\]

so that the first term on the right-hand side of (A.14) becomes

\[
EK_h^2(Z - z) \int_0^\infty \left( \frac{\Psi^{-1}(Z) U}{e_t(Y; \theta_0) + \Psi^{-1}(Z)} \right) dM(t) 
\]

\[
= EK_h^2(Z - z) \int_0^\infty \left( \frac{\Psi^{-1}(Z) U}{e_t(Y; \theta_0) + \Psi^{-1}(Z)} \right) dN(t) 
\]
\[
E K_h^2 (Z - z) \delta \left( \frac{\psi^{-1}(z) u}{\epsilon_1(r; \theta_0) + \psi^{-1}(z)} \right) \delta \left( \frac{-\zeta'(Y; \theta_0|Z)}{-\delta_i'(Y; \theta_0|Z)} \right).
\]

Therefore, we have

\[
(n h) \text{Var} \{ \ell_n^*(\gamma_0) \} = \int_{-\infty}^{\infty} E \left\{ \delta \left( \frac{\psi^{-1}(z) u}{\epsilon_1(r; \theta_0) + \psi^{-1}(z)} \right) \delta \left( \frac{-\zeta'(Y; \theta_0|Z)}{-\delta_i'(Y; \theta_0|Z)} \right) \right\} K^2(u) \, du + o(1)
\]

\[
= f(z) S_1(z; \theta_0) + o(1).
\]

To accomplish the proof, the Cramér-Wold device is utilized. Namely, it is to show that for any constant vector \( b \neq 0 \),

\[
\sqrt{nh} \left\{ b^T \ell_n^*(\gamma_0) - b^T E \ell_n^*(\gamma_0) \right\} \stackrel{D}{\longrightarrow} N \left( 0, f(z) b^T S_1(z; \theta_0) b \right).
\]

(A.15)

Note that the left-hand side of (A.15) can be expressed as follows

\[
\sqrt{nh} n^{-1} \sum_{i=1}^{n} \left\{ K_h(Z_i - z) X_i - E K_h(Z_i - z) X_i \right\},
\]

where

\[
X_i = b^T \begin{pmatrix}
E_0(Y_i; \theta_0) - \frac{\delta_i}{m_0(0; \theta_0)} \\
\delta_i \zeta'(Y_i; 0; \theta_0|Z_i) - \exp(-Z_i^T \beta_0) E_0'(Y_i; \theta_0) - \zeta_0'(Y_i; \theta_0)
\end{pmatrix}.
\]

To establish (A.15), it suffices to check the Lyapounov condition

\[
E \sum_{i=1}^{n} \left| \sqrt{nh} \left\{ K_h(Z_i - z) X_i - E K_h(Z_i - z) X_i \right\} \right|^{2+\eta} = o(1)
\]

for some \( \eta > 0 \), which can be verified easily. By Condition B (i), the left-hand side of the above expression is bounded by

\[
2(n^{-1}h)^{1+\eta/2} n E |X_i| K_h(Z - z)|^{2+\eta} = O ((nh)^{-\eta/2}) \rightarrow 0,
\]

which implies that (A.15) holds true. Therefore, by (A.10) and (A.12)

\[
\sqrt{nh} \left\{ \widehat{\gamma} - \gamma_0 - S_0(z; \theta_0)^{-1} \widehat{b}_n(z) \right\} \rightarrow N \left\{ 0, f^{-1}(z) S_0(z; \theta_0)^{-1} S_1(z; \theta_0) S_0(z; \theta_0)^{-1} \right\}.
\]

It remains to show that

\[
S_0(z; \theta_0)^{-1} \widehat{b}_n(z) = \begin{pmatrix} b_n(z) \\ 0 \end{pmatrix},
\]

(A.16)
where $S_0(z; \theta_0)$, $\tilde{b}_n(z)$ and $b_n(z)$ are given respectively by (2.9), (A.13) and (2.10). Using the following fact

$$
\mu^T(K) \nu^{-1}(K) \left( \int u^{p+1} u K(u) \, du \right) = \int u^{p+1} K(u) \, du,
$$

and Proposition 1 (iii), one can verify that

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

so that (A.16) holds. Therefore, the theorem is proved. \qed

**Proof of Theorem 6.** By the Taylor’s expansion,

$$
0 = U_n(\beta^*_0) = U_n(\beta_0^*) + U'_n(\beta_0^*) \left( \beta^*_s - \beta_0^* \right) + R_n.
$$

It is easy to show that

$$
U'_n(\beta_0^*) = -E(\delta \mid Z = z) f(z) \tilde{v}_s(K) + o_P(1).
$$

Using the same techniques as those employed in the proof of Theorem 1 in Cai, Fan, and Li (2000), we can prove that

$$
R_n = o_p \left( (n h)^{-1/2} \right).
$$

Then,

$$
\sqrt{n h} \left( \beta^*_s - \beta_0^* \right) = \{E(\delta \mid Z = z) f(z) \tilde{v}_s(K)\}^{-1} \sqrt{n h} U_n(\beta_0^*) + o_P(1).
$$

It is easy to see that

$$
a_{2n} = f(z) E(\delta \mid Z = z) + o_P(1), \quad \text{and} \quad a_{3n} = f(z) E(Y \mid Z = z) + o_P(1),
$$

so that

$$
U_n(\beta_0^*) = V_n(\beta_0^*) + o_p \left( (n h)^{-1/2} \right),
$$

where

$$
V_n(\beta_0^*) = \frac{1}{n} \sum_{i=1}^n \left[ -\delta_i + \frac{E(\delta \mid Z = z)}{E(Y \mid Z = z)} \exp \left( -Z_i^T \beta^*_s \right) Y_i \right] Z_i^T K_h(Z_i - z).
$$

Some algebraic calculation gives

$$
E[V_n(\beta_0^*)] = E(\delta \mid Z = z) f(z) \frac{\psi(p+1)(z)}{(p + 1)!} h^{p+1} \int u^{p+1} K(u) \, du + o(h^{p+1}),
$$

where $\psi(p+1)(z)$ is the $(p+1)$th derivative of $\psi(z)$. The proof is completed.
Estimate of Biometric Function

\[
\begin{align*}
n h \text{Var}[V_n(\beta_0^*)] &= E \left[ \left( \delta - \frac{E(\delta | Z = z)}{E(Y | Z = z)} \right)^2 | Z = z \right] f(z) \nu(K^2) + o(1).
\end{align*}
\]

Therefore,

\[
\begin{align*}
\sqrt{n h} \left( \beta_0^* - \beta_0^* - b_n^*(z) \right) \\
= \sqrt{n h} \left( \frac{V_n(\beta_0^*) - E(V_n(\beta_0^*)))}{E(\delta | Z = z) f(z)} \right) + o_P(1).
\end{align*}
\]

Hence, the theorem is proved. □

REFERENCES
