ASYMPTOTIC THEORY OF THE ESTIMATED PARTIAL LIKELIHOOD

by

Haibo Zhou and Margaret S. Pepe

Department of Biostatistics, University of
North Carolina at Chapel Hill, NC.

Institute of Statistics Mimeo Series No. 2114

April 1993

Asymptotic Theory of The Estimated Partial Likelihood

Haibo Zhou and Margaret S. Pepe*

ABSTRACT

A semi-parametric method, Estimated Partial Likelihood(EPL) method, is proposed for the estimation of the relative risk parameters when auxiliary covariates are present. The EPL method makes full use of the avaiable covariates and auxiliary measurement without imposing the parametric structure on the missing procedures. The EPL function reduces to the partial likelihood function of Cox(1972) when the auxiliary is perfect measurement about true covariates. Since the EPL takes a non-parametric appraoch with respect to the underline missing process, therfore avoides some strong assumptions, e.g. the rare disease assumption(Prentice 1982, Pepe , Self and Prentice 1991). The asymptotic distribution theory is estabilished for the proposed estimator for the case in which the surrogate or mismeasured covariates are categorical. The proposed EPL estimator is found to be consistant and asymptoticaly normally distributed. A consistent variance estimator is provided. An extension of the EPL is presented when the validation sample is a stratified random sample.

KEY WORDS: Missing data; Survival analysis; Relative Risk Parameter Estimation; Stochastic Intergration; Surogate Covariate Data.

* Haibo Zhou is a Postdoctoral Fellow, Department of Biostatistics, The University of North Carolina at Chapel Hill. Chapel Hill, N.C. 27599-7400. Margaret S. Pepe is an associate member, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104. This research was supported in part by National Science Foundation grant No. DMS 89-02211. Part of this work was included in 1992 Department of Statistics, University of Washington Doctoral dissertation by Haibo Zhou. The authors thank C. Ed Davis for the helpful discussion about SOLVD data and insightful comments.

1. INTRODUCTION

Auxiliary Covariate data is a common problem in statistical data analysis. For example, in a substudy of the Studies of Left Ventricular Dysfunction(SOLVD)(SOLVD Investigators, 1991) , the patient's left ventricular ejection fraction(EF) is an importent risk factor in predicting the patient's congestive heart failure. However due to the complication and expensiveness of the measuring procedure, the EF, which is a standardized radionuclide measurement, is only avaiable to a subset of 179 patients in the study. For a total of 1111 patients in the substudy, a nonstandardized across clinical centers measurement of EF is avaiable to everyone.

In the above example, if we denote the the standard EF as X and the nonstadardized EF as Z, then the data we are dealing with is of the form,

$$\begin{array}{ll} \{S_i, \delta_i, X_i, Z_i\} & i \in V \\ \{S_i, \delta_i, Z_i\} & i \in \bar{V} \end{array} \quad (1)$$

where S denotes the observed time event, which is the minimum of potential failure T and censoring time U. δ is the indicator of failure. V is the validation sample where subjects have both X and Z avaiable. $\bar{V}$ is the nonvalidation sample where subjects only have Z avaiable. β, the regression coefficient of X is the primary interest. Let n, n_v and $n_{\bar{v}}$ denote the sample size of the whole, validation and non-validation sample respectively.

The most naive approach to the data in (1) is to discard the non-validation sample and only analyze the validation sample. This could induce a substantial reduction in efficiency. A more conprehensive approach is to parameterize the underline distribution and use a fully parametric approach. However, realistic models are hard to construct, furthermore, this approach may not be robust to model misspecification. The proposed EPL approach is aimed at avoiding the stringent model assumption yet still utilize the information contained in the non-validation sample.

Denote the relative risk functions for the subjects in the validation and non-validation sample respectively as follows:

$$\begin{array}{rcll} r_i(t) & = & r(\beta, X_i(t)) & if \;\; i \in V \\ \bar{r}_i(t) & = & \bar{r}(\beta, Z_i(t)) & if \;\; i \in \bar{V} \end{array}$$

For the observed data we will denote the relative risk function for an individual i as

$$r_i^*(t) = r_i(t) I_{[i \in V]} + \bar{r}_i(t) I_{[i \in \bar{V}]} \quad (2)$$

The inference about regression parameter β can be drawn from the partial likelihood(Prentice & Self, 1983)

$$PL(\beta) = \prod_{i=1}^{n} [\frac{r_i^*(T_i)}{\sum_{j \in \mathcal{R}(T_i)} r_j^*(T_i)}]^{\delta_i} \quad (3)$$

However, as pointed out by Prentice(1982), the induced relative risk function $\bar{r}_i(t)$ can be expressed as

$$\begin{array}{rcl} \bar{r}_i(t) & = & \bar{r}(\beta, Z_i(t)) \\ & = & E(\beta, X_i(t) | T_i \geq t, Z_i(t)) \end{array}$$

where the condition $[T_i \geq t, Z_i(t)]$ usually imply that $\bar{r}(t)$ depend on the baseline hazard and underline distribution of X and Z. Hence the partial likelihood (3) is only useful under some special cases(e.g. under 'rare disease' assumption, Prentice 1982, Pepe, Self and Prentice 1991).

We propose to estimate $\bar{r}(t)$ empirically on the basis of the validation sample observations. So the estimated relative risk function for an individual i given his observed data is:

$$\hat{r}_i^*(t) = r_i(t)I_{[i\in V]} + \hat{\bar{r}}_i(t)I_{[i\in \bar{V}]} \quad (4)$$

where $\hat{\bar{r}}_i(t)$ is the estimated relative risk function for an individual in the non-validation sample with only auxiliary covariate process $Z(t)$ available. Specifically, the empirical estimate of the induced relative risk function denoted by $\hat{\bar{r}}(t)$ is

$$\begin{aligned} \hat{\bar{r}}(t) &= \hat{E}[r(\beta, X(t))|S \geq t, Z(t)] \\ &= \frac{\sum_{i\in V} I_{[S_i\geq t, Z_i(t)=Z(t)]} r(\beta, X_i(t))}{\sum_{i\in V} I_{[S_i\geq t, Z_i(t)=Z(t)]}}, \end{aligned} \quad (5)$$

where Z is assumed discrete with $Prob(Z = z_k) = p_{z_k}$, $k = 1, \cdots, q$ and $\sum_{k=1}^{q} p_{z_k} = 1$. The EPL estimator is the the maximizar of the resulting Estimated Partial Likelihood,

$$EPL(\beta) = \prod_{i=1}^{n} [\frac{\hat{r}_i^*(T_i)}{\sum_{j\in\mathcal{R}(T_i)} \hat{r}_j^*(T_i)}]^{\delta_i} \quad (6)$$

The proposed estimator has the following properties,

1. No parametric assumptions about the underlying distributions are assumed. Hence the EPL estimator is robust with respect to $f(X|T, Z)$;
2. Leave baseline hazard function $\lambda_0(t)$ completely arbitrary;
3. No rare disease assumption(Prentice 1982, Pepe, Self and Prentice 1991) needed in the EPL method;
4. the EPL method is computationally straightforward.
5. Fully utilizes the information about β contained in the non-validation set.
6. Validation set needs not to be a simple random sample of the entire sample. Since the nature of the estimation of $\bar{r}$ in the EPL method, the validation set can be a random sample stratified on Z. Moreover the validation set can be time dependent. i.e. the missingness of the covariate may depend on the Z and time t. The last property is very useful in the large cohort study because the subjects may enter the validation set during the study;

The rest of this artical is orgnized as following: We will first formulate the Cox model(Cox, 1972) into a more general multivariate counting process framework in Section 2 and outline the proofs of the asymptotic properties of the proposed $\hat{\beta}_{EPL}$. The conditions needed in the proofs can be found in Section 2 where some further definitions and notations are given as well. Rigorous proofs of the two main results in this paper are detailed in Section 3 and 4 where consistency of $\hat{\beta}_{EPL}$ is shown and the asymptotic distribution of $\hat{\beta}_{EPL}$ is determined. A consistent estimator of the asymptotic variance-covariance matrix is also presented in Section 4. While all of the above results are obtained assuming that the validation set is a simple random sample, a version of main results with a random validation set stratified on Z can be found in Section 5.

2. THE COUNTING PROCESS FORMULATION OF THE MODEL

In this section we shall first formulate the Cox model in the framework of the multivariate counting processes in Section 2.1. For simplicity the time interval is assumed to be finite. Without lost of generality, we shall be working on the time interval $[0,1]$. The condition that needed in developing the asymptotic theory and some definations are presented in Section 2.2. Section 2.3 outlines the idea of proving the asymptotic theory.

Background theory for this chapter are the theories for multivariate counting processes, stochastic integrals and local martingales. We shall use the basic results from those theories without further comment. A good survey of these theories can be found in Fleming & Harrington(1991). The use of the *local* concept will allow us to avoid making the superfluous integrability conditons. Apart from the background theories, our basic tools are the Inverse Function Theorem(Rudin, 1964), the Inequality of Lenglart and the martingale central limit theorem of Rebolledo. The latter two theorems are not presented in this work. We usually refer to the one given in Andersen & Gill(1982) which is slightly extended with respect to the originals.

2.1. Formulation of The Model

Since we are interested in the asymptotic properties, we shall in fact consider a sequence of models, indexed by $n = 1, 2, ...$ We shall generalize from the censored observation of lifetimes of n subjects to the observation (in the nth model) of an n-component counting process $N^{(n)} = (N_1^{(n)}, ..., N_n^{(n)})$, where $N_j^{(n)}$ counts the observed events in the life of the jth subject, $j = 1, 2, ..., n$, over the time interval $[0,1]$. In this work, only the case of nonrecurrent events will be studied.

Specifically, the multivariate counting processes will be generated as follows,
Let $S_j^{(n)} = min(T_j^{(n)}, U_j^{(n)})$ be the observation of the jth subject, where $T_j^{(n)}$ is the potential failure time , $U_j^{(n)}$ is the potential censoring time and $S_j^{(n)}$ is the observed event time of the jth subject in the study. The counting process $N_j^{(n)}$ is defined by

$$N_j^{(n)}(t) = I_{[S_j^{(n)} \leq t,\ \delta_j^{(n)}=1]} \tag{7}$$

where $\delta_j^{(n)} = I_{[T_j^{(n)} \leq U_j^{(n)}]}$ indicates if the observation is a failure time. So the sample paths of $N_1^{(n)}, \cdots, N_n^{(n)}$ are step functions that satisfy,

1. $N_i^{(n)}(0) = 0, \qquad i = 1, \cdots, n;$
2. jump size +1 only;
3. No two components processes jumping at the same time;
4. At most one jump for each subject in the study.

In our model, properties of stochastic processes, such as being a local martingale or a predictable process, are relative to a right-continuous nondecreasing family $\{\mathcal{F}_t^{(n)} : t \in [0,1]\}$ of sub σ-algebras on the nth probability space $(\Omega^{(n)}, \mathcal{F}^{(n)}, \mathcal{P}^{(n)})$; $\mathcal{F}_t^{(n)}$ is the filtration and can be thought of as the history of everything that happens up to time t(in the nth model).

Our basic assumption is that for each n, $N^{(n)}$ has a random intensity process $\lambda^{(n)} = (\lambda_1^{(n)}, ..., \lambda_n^{(n)})$ such that

$$\lambda_i^{(n)}(t) = Y_i^{(n)}(t)\lambda_0(t)r_i^{*(n)}(t) \tag{8}$$

where $r_i^{*(n)}(t)$ is as defined in (2), $Y_j^{(n)}(t)$ is the at risk process defined by $Y_j^{(n)}(t) = I_{[S_j^{(n)} \geq t]}$ and $\lambda_0(t)$ is a unknown but fixed underlying baseline hazard function.

In our study the covariate processes $X^{(n)}(t)$ and $Z^{(n)}(t)$ are assumed to be predictable and locally bounded. This assumption holds when $X^{(n)}(t)$ and $Z^{(n)}(t)$ are left continuous, with right hand limits and adapted.

By stating that $N^{(n)}$ has intensity process $\lambda^{(n)}$, we mean that the processes $M_i^{(n)}$ defined by

$$M_i^{(n)}(t) = N_i^{(n)}(t) - \int_0^t \lambda_i^{(n)}(u)du, \quad i = 1, ..., n, \quad t \in [0,1] \tag{9}$$

are local martingales on the time interval $[0,1]$. Since the intensity process $\lambda^{(n)}$ is locally bounded as well, the local martingales (9) are in fact local square integrable martingales.

$$< M_i^{(n)}, M_i^{(n)} > (t) = \int_0^t \lambda_i^{(n)}(u)du \quad and \quad < M_i, M_j >= 0, \quad i \neq j, \tag{10}$$

i.e. $M_i^{(n)}$ and $M_j^{(n)}$ are orthogonal when $i \neq j$.

The nice property of the local martingale is that if H_i is locally bounded and $\mathcal{F}_t$-predictable, then $\sum_{i=1}^n \int H_i^{(n)} dM_i^{(n)}$ is a local square integrable martingale with an easy to calculate covariance processes $\sum_{i=1}^n \int H_i^2 d < M_i^{(n)} >$. This is one of the key ingredients in establishing asymptotic properties of $\hat{\beta}_{EPL}$.

In the following we shall drop the superscript (n) everywhere. Only β and λ_0 are same in all models(i.e. for each n). Convergence in probability ($\xrightarrow{p}$) and convergence in distribution ($\xrightarrow{d}$) are always relative to the probability measures $\mathcal{P}^{(n)}$ parameterized by β and λ_0.

Let $C(\beta, t)$ be the logarithm of the partial likelihood (3) evaluated at time t, so that according to (8) the partial likelihood (3) for the observed data is

$$C(\beta, t) = \sum_{i=1}^n \int_0^t log\{r_i^*(s)\} dN_i(s) - \int_0^t log\{\sum_{i=1}^n Y_i(s) r_i^*(s)\} d\bar{N}(s), \tag{11}$$

where $\bar{N} = \sum_{i=1}^n N_i$. As we discussed in Section 1, the induced relative risk function $\bar{r}(\beta, t)$ is not known for non-validation set member unless the underlying conditional distribution function $f_{X|S \geq t, Z}(x)$ is specified. In Section 1, we proposed to estimate $\bar{r}(\beta, t)$ empirically for non-validation set members and to draw inference about β from the estimated partial likelihood function (6). Then the proposed estimator $\hat{\beta}_{EPL}$ is the solution to the estimating equation $\frac{\partial}{\partial \beta} \hat{C}(\beta, 1) = 0$, where $\hat{C}(\beta, 1)$ is obtained by substitute $\hat{r}^*(\beta, t)$ for $r^*(\beta, t)$ in (11). The vector of derivatives $\hat{U}(\beta, t)$ of $\hat{C}(\beta, t)$ with respect to β has the form

$$\begin{aligned} \hat{U}(\beta, t) &= \sum_{i=1}^n \int_0^t \frac{\hat{r}_i^{*(1)}(s)}{\hat{r}_i^*(s)} dN_i(s) - \int_0^t \frac{\sum_{i=1}^n Y_i(s) \hat{r}_i^{*(1)}(s)}{\sum_{i=1}^n Y_i(s) \hat{r}_i^*(s)} d\bar{N}(s), \\ &= \sum_{i=1}^n \int_0^t \Delta(\hat{r}_i^*)(s) dN_i(s) \end{aligned} \tag{12}$$

where $\hat{r}_i^{*(1)}(s)$ and $\Delta(\hat{r}_i)(s)$ are defined as

$$\begin{aligned} \hat{r}_i^{*(1)}(s) &= \frac{\partial}{\partial \beta} \hat{r}_i^*(s) \\ \Delta(\hat{r}_i)(s) &= \frac{\hat{r}_i^{(1)}(s)}{\hat{r}_i(s)} - \frac{\sum_{i=1}^n Y_i(s) \hat{r}_i^{(1)}(s)}{\sum_{i=1}^n Y_i(s) \hat{r}_i(s)} \end{aligned}$$

From (8) and (9), we can rewrite $\hat{U}(\beta,t)$ as

$$\hat{U}(\beta,t)=\sum_{i=1}^{n}\int_0^t \Delta(\hat{r}_i^*)(s)dM_i(s)+\sum_{i=1}^{n}\int_0^t \Delta(\hat{r}_i^*)(s)r_i^*(s)Y_i(s)\lambda_0(s)ds \tag{13}$$

2.2. Further Definations and Conditions

Some important notation will be defined in this section which will be used in the proofs. First the notation defined in Section 2.1 will be the same. Unless otherwise stated, all limits are taken as $n \longrightarrow \infty$, which implies $n_v \longrightarrow \infty$ and $n_{\bar{v}} \longrightarrow \infty$ as well, where n_v and $\bar{v}$ are the number of subjects in validation and non-validation sets, respectively. For a matrix A or vector a, we define the norm as $||A|| = sup_{i,j}|a_{ij}|$ and $||a|| = sup_i|a_i|$. For a vector a, define $|a| = (\sum a_i^2)^{\frac{1}{2}} = (a'a)^{\frac{1}{2}}$. We also write the matrix of aa' as $a^{\otimes 2}$ and the matrix $(aa')(aa')'$ as $a^{\otimes 4}$. For the relative risk function r^* (as well as for $\hat{r}^*$, r, $\hat{\bar{r}}$ and $\bar{r}$), we let $r^{*(j)}$ denote the jth derivative of r^* with respect to β, $j=0,1,2$, where $j=0$ represents the function itself.

Some further definitions are:

$$\begin{aligned}
\hat{S}^{(0)} &= \frac{1}{n}\sum_{i=1}^{n}Y_i(t)\hat{r}_i^*(\beta,t)\\
\hat{S}^{(1)} &= \frac{\partial}{\partial\beta}\hat{S}^{(0)} = \frac{1}{n}\sum_{i=1}^{n}Y_i(t)\hat{r}_i^{*(1)}(\beta,t)\\
\hat{S}^{(2)} &= \frac{\partial}{\partial\beta}\hat{S}^{(1)} = \frac{1}{n}\sum_{i=1}^{n}Y_i(t)\hat{r}_i^{*(2)}(\beta,t)\\
\hat{S}^{(3)} &= \frac{1}{n}\sum_{i=1}^{n}Y_i(t)\frac{\hat{r}_i^{*(2)}(\beta,t)}{\hat{r}_i^*(\beta,t)}r_i^*(\beta_0,t)\\
\hat{S}^{(4)} &= \frac{1}{n}\sum_{i=1}^{n}Y_i(t)(\frac{\hat{r}_i^{*(1)}(\beta,t)}{\hat{r}_i^*(\beta,t)})^{\otimes 2}r_i^*(\beta_0,t)\\
\hat{S}^{(5)} &= \frac{1}{n}\sum_{i=1}^{n}Y_i(t)(\frac{\hat{r}_i^{*(2)}(\beta,t)}{\hat{r}_i^*(\beta,t)})^{\otimes 2}r_i^*(\beta_0,t)\\
\hat{S}^{(6)} &= \frac{1}{n}\sum_{i=1}^{n}Y_i(t)(\frac{\hat{r}_i^{*(1)}(\beta,t)}{\hat{r}_i^*(\beta,t)})^{\otimes 4}r_i^*(\beta_0,t)\\
\hat{S}^{(7)} &= \frac{1}{n}\sum_{i=1}^{n}Y_i(t)\frac{\hat{r}_i^{*(1)}(\beta,t)}{\hat{r}_i^*(\beta,t)}r_i^*(\beta_0,t)
\end{aligned}$$

We define $S^{(j)}(\beta,t)$ as the corresponding functions with $r^*(\beta,t)$ substituted for $\hat{r}^*(\beta,t)$ in the above $\hat{S}^{(j)}(\beta,t)$, $j=0,1,2,\cdots,7$. Also, we define

$$\begin{aligned}
s^{(0)} &= E(Y(t)r^*(\beta,t))\\
s^{(1)} &= E(Y(t)r^{*(1)}(\beta,t))\\
s^{(2)} &= E(Y(t)r^{*(2)}(\beta,t))\\
s^{(3)} &= E(Y(t)\frac{r^{*(2)}(\beta,t)}{r^*(\beta,t)}r^*(\beta_0,t))
\end{aligned}$$

$$
\begin{aligned}
s^{(4)} &= E(Y(t)(\frac{r^{*(1)}(\beta,t)}{r^{*}(\beta,t)})^{\otimes 2}r^{*}(\beta_0,t)) \\
s^{(5)} &= E(Y(t)(\frac{r^{*(2)}(\beta,t)}{r^{*}(\beta,t)})^{\otimes 2}r^{*}(\beta_0,t)) \\
s^{(6)} &= E(Y(t)(\frac{r^{*(1)}(\beta,t)}{r^{*}(\beta,t)})^{\otimes 4}r^{*}(\beta_0,t)) \\
s^{(7)} &= E(Y(t)\frac{r^{*(1)}(\beta,t)}{r^{*}(\beta,t)}r^{*}(\beta_0,t))
\end{aligned}
$$

and

$$
\Sigma = \int_0^1 \left[\frac{s^{(4)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} - \left(\frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}\right)^{\otimes 2}\right] s^{(0)}(\beta_0,\omega)\lambda_0(\omega)d\omega \tag{14}
$$

$$
\Sigma_1 = \int_0^1 \left[E\left(Y_i(\omega)\frac{\bar{r}_i^{(1)}(\beta_0,\omega)^{\otimes 2}}{\bar{r}_i(\beta_0,\omega)}\right) - \frac{s^{(1)}(\beta_0,\omega)^{\otimes 2}}{s^{(0)}(\beta_0,\omega)}\right]\lambda_0(\omega)d\omega \tag{15}
$$

$$
\Sigma_2 = E\left\{\int_0^1 \left(\frac{r_j^{(1)}(\beta_0,\omega)}{r_j(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}\right) dM_j(\omega) - \frac{1-\rho}{\rho}Q_j\right\}^{\otimes 2} \tag{16}
$$

$$
\begin{aligned}
Q_j^{\bar{v}} &= \frac{1}{\bar{v}}\sum_{j=1}^{\bar{v}}\int_0^1 \Delta(\bar{r}_i(\beta_0,\omega))\frac{Y_i(\omega)Y_j(\omega)I_{[Z_i=Z_j]}}{p_{z_j}H_{z_j}} \\
&\quad \times [r_j(\beta_0,\omega) - \bar{r}_i(\beta_0,\omega)]\lambda_0(\omega)d\omega \\
Q_j &= \int_0^1 \left(\frac{\bar{r}_j^{(1)}(\beta_0,\omega)}{\bar{r}_j(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}\right) Y_j(\omega) \\
&\quad \times (r_j(\beta_0,\omega) - \bar{r}_j(\beta_0,\omega))\lambda_0(\omega)d\omega
\end{aligned}
$$

We also define $\rho = \lim_{n\to\infty}\frac{n_v}{n}$ and

$$
H_z(t) = Prob(Y(t) = 1|Z = z)
$$

The following list of conditions will be assumed to hold throughout this chapter. There are a number of redundancies in them, and not all conditions are needed for every result, but in this way we hope to avoid too many technical distractions in the theorems and their proofs. Further discussion is deferred till Section 2.5.

(A) $\int_0^1 \lambda_0(t)dt < \infty$

(B) $P(Y(1) = 1|Z = z_k) > 0 \quad k = 1, 2, \cdots, q$

(C) There exists an open subset $\mathcal{B}$, containing β_0, of the Euclidean p space E^p. $r^{(2)}$ with elements $\frac{\partial^2}{\partial\beta_i\partial\beta_j}r(\beta,t)$ exists and is continuous on $\mathcal{B}$ for each $t \in [0,1]$, uniformly in t, and $\bar{r}(\beta,t)$ is bounded away from 0 on $\mathcal{B}\times[0,1]$. $\Sigma(\beta_0)$ is positive definite.

(D)

$$
E\{sup_{\mathcal{B}\times[0,1]}|Y(t)r^{*(j)}(\beta,t)|\} < \infty \quad j = 0,1,2
$$

$$E\{sup_{\mathcal{B}\times[0,1]}|Y(t)(\frac{r^{*(1)}(\beta,t)}{r^*(\beta,t)})^{\otimes 2j}r^*(\beta_0,t)|\} < \infty \quad j=1,2$$

$$E\{sup_{\mathcal{B}\times[0,1]}|Y(t)(\frac{r^{*(2)}(\beta,t)}{r^*(\beta,t)})^{\otimes j}r^*(\beta_0,t)|\} < \infty \quad j=1,2$$

(E)

$$sup_{0\leq t\leq 1}|Z_v^{(K)}(t)| = O_p(1) \ \ as \ \ K=0,1$$

where

$$Z_v^{(K)}(t) \equiv \sqrt{v}\{\frac{1}{v}\sum_{i=1}^{v} I_{[Y_i=1,Z_i=z]}r_i^{(K)}(\beta,t) - E(I_{[Y(t)=1,Z=z]}r^{(K)}(\beta,t))\} \ \ K=0,1$$

Note that the partial derivative conditions on $r^{(j)}$ are satisfied by $s^{(j)}$, $S^{(j)}$, $\hat{S}^{(j)}$ and $\bar{r}$'s as well under regularity conditions. Even though we list condition (E) as one of our assumptions, it is actually fullfilled in a general setting.

2.3. Outline of The Proofs of Asymptotic Theory

Since we define $\hat{\beta}_{EPL}$ as the solution of $\hat{U}(\beta,1)=0$, a Taylor expansion of $\hat{U}(\beta,1)$ about β_0 evaluated at $\hat{\beta}_{EPL}$ gives

$$n^{-\frac{1}{2}}\hat{U}(\beta_0,1) = \{-n^{-1}\frac{\partial}{\partial\beta_*}\hat{U}(\beta_*,1)\}n^{\frac{1}{2}}(\hat{\beta}_{EPL}-\beta_0) \tag{17}$$

where β_* is between $\hat{\beta}$ and β_0. Therefore to prove asymptotic normality of $n^{\frac{1}{2}}(\hat{\beta}_{EPL}-\beta_0)$ it is suffficient to show

$$n^{-\frac{1}{2}}\hat{U}(\beta_0,1) \xrightarrow{d} N(0,\rho\Sigma_1+(1-\rho)\Sigma_2) \text{ as } n\to\infty$$
$$-n^{-1}\frac{\partial}{\partial\beta_*}\hat{U}(\beta_*,1) \xrightarrow{p} \Sigma \text{ as } n\to\infty$$

where Σ, Σ_1, Σ_2 and ρ will be defined later. The difficulty in proving the weak convergence of $n^{-\frac{1}{2}}\hat{U}(\beta_0,1)$ is that the second component in (13) does not disappear as in Andersen & Gill(1982) and Prentice & Self(1983). Therefore we cannot write $\hat{U}(\beta,t)$ as a simple summation of local martingales. Martingale theory which is useful for inference with partial likelihood cannot be invoked here with EPL. In fact the existence of the second component here makes it impossible to write (13) as a summation of uncorrelated terms. In Section 3 we'll give a proof to show that $n^{-\frac{1}{2}}\hat{U}(\beta_0,1)$ still converges to a normal process.

Existence of a unique consistent solution to the EPL equations is obtained as a consequence of the Inverse Function Theorem(Rudin, 1964). Roughly speaking, the Inverse Function Theorem states that a continuously differentiable mapping f is invertible in a neighbourhood of any point x at which the linear transformation $f^{(1)}(x)$ is invertable.

$\hat{\beta}_{EPL}$ is the value at 0 of the inverse function $\frac{1}{n}\hat{U}^{-1}: E^p \to \mathcal{B}$, where $\mathcal{B}$ is an open subset of Euclidean p space E^p and β_0 is contained in $\mathcal{B}$. In Section 2.3 $\frac{1}{n}\hat{U}^{-1}$ is shown to be well defined in an open neighborhood about zero with probability going to one, and $\frac{1}{n}\hat{U}(0)^{-1}$ is shown to be a consistent estimate of β.

3. CONSISTENCY OF $\hat{\beta}_{EPL}$

This section is divided into 2 subsections. In Section 3.1, the preliminary results are given. The consistency theorem is stated and proved in Section 3.2. The principal tool in proving consistency of $\hat{\beta}_{EPL}$ is the Inverse Function Theorem. There are several forms of the Theorem. One form of the theorem is given in Theorem 2.2(Rudin, 1964) for ease of reference. Theorem 1 is an important theorem given by Anderson and Gill(1982). Theorem 5 is the main theorem that states the existence and uniqueness of consistent estimator of β. Theorem 3, 4 and Lemma 1 - 4 are the preliminaries for Lemma 5. Note that stability results in Theorem 3, 4 and asymptotic regularity results in Lemma 1 will also be useful in the asymptotic distribution proof. An analogous proof of consistency with standard likelihood function can be found in Foutz(1977).

3.1. Preliminary Results

The following result which is due to Anderson & Gill will be used below.

Theorem 1 *Let $X; X_1, \cdots,$ be iid random elements of $D_E[0,1]$ (endowed with the Skorohod topology) where elements of $D_E[0,1]$ are right continuous functions on $[0,1]$ with left hand limits taking values in a separable Banach space E (rather than the usual $\Re$). Suppose that*

$$\mathcal{E}||X|| = \mathcal{E}sup_{t\in[0,1]}||X(t)|| < \infty$$

For each n, let $t_1^{(n)} \geq \cdots \geq t_n^{(n)}$ be fixed time instants in $[0,1]$. Let $y_i^{(n)} = I_{[t_i^{(n)},1]}$ and suppose that there exists a distribution function y such that, on $[0,1]$,

$$||\frac{1}{n}\sum_{i=1}^{n} y_i^{(n)} - y|| \longrightarrow 0 \ \ as \ \ n \to \infty$$

Then

$$||\frac{1}{n}\sum_{i=1}^{n} X_i y_i^{(n)} - \mathcal{E}Xy|| \longrightarrow 0 \ \ as \ \ n \to \infty$$

The proof of Theorem 1 can be seen in Anderson and Gill(1982, Theorem III.1).

Theorem 2 (The Inverse Function Theorem) *Suppose f is a mapping from an open set Θ in E^r into E^r, the partial derivatives of f exist and are continuous on Θ, and the matrix of derivatives $f'(\theta^*)$ has inverse $f'(\theta^*)^{-1}$ for some $\theta^* \in \Theta$. Write*

$$\lambda = 1/(4||f'(\theta^*)^{-1}||).$$

Use the continuity of the elements of $f'(\theta^*)$ to fix a neighborhood $\mathbf{U}_\delta$ of θ^* of sufficiently small radius $\delta > 0$ to insure $||f'(\theta) - f'(\theta^*)|| < 2\lambda$, whenever $\theta \in \mathbf{U}_\delta$. Then (a) for every θ_1, θ_2 in $\mathbf{U}_\delta$,

$$|f'(\theta_1) - f'(\theta_2)| \geq 2\lambda|\theta_1 - \theta_2|,$$

and (b) the image set $f(\mathbf{U}_\delta)$ contains the open neighborhood with radius $\lambda\delta$ about $f(\theta^*)$.

Conclusion (a) insures that f is one-to-one on $\mathbf{U}_\delta$ and that f^{-1} is well defined on the image set $f(\mathbf{U}_\delta)$. The theorem is proved in this form in Rudin(1964, p193).

Theorem 3

$$\sup_{\mathcal{B}\times[0,1]} ||\hat{\bar{r}}(\beta,t) - \bar{r}(\beta,t)|| \longrightarrow 0 \ \ a.s.$$

$$\sup_{\mathcal{B}\times[0,1]} ||\hat{r}^*(\beta,t) - r^*(\beta,t)|| \longrightarrow 0 \ \ a.s.$$

PROOF:

$$\hat{\bar{r}}^*(\beta,t) = \frac{\frac{1}{v}\sum_{i=1}^{v} I_{[Y_i(t)=1,Z_i=z]} r_i(\beta,t)}{\frac{1}{v}\sum_{i=1}^{v} I_{[Y_i(t)=1,Z_i=z]}} = \frac{A}{B}$$

Since $Y, \mathbf{X}$ and Z are left continuous processes with right hand limits, if we reverse the time axis, we can consider $I_{[Y_i(t)=1,Z_i=z]} r_i(\beta,t)$ and $I_{[Y_i(t)=1,Z_i=z]}$ as random elements of $D[0,1]$, where the elements of $D[0,1]$ take values in Banach space of continuous function on $\mathcal{B}$. This is also a separable Banach space if we endow this space of functions with the supremum norm. Then by Result 1 and noticing that

$$E\{sup_{[0,1]} I_{[Y(t)=1,Z=z]}\} < \infty$$

we have

$$sup_{\mathcal{B}\times[0,1]}||B - b|| \xrightarrow{p} 0 \tag{18}$$

where $b = E(I_{[Y(t)=1,Z=z]}) = P(Y(t) = 1|Z = z)P(Z = z) = p_z H_z(t)$. Note also that condition (D) implies that

$$E\{sup_{\mathcal{B}\times[0,1]} I_{[Y(t)=1,Z=z]} r(\beta,t)\} < \infty$$

Hence

$$sup_{\mathcal{B}\times[0,1]}||A - a|| \xrightarrow{p} 0 \tag{19}$$

where $a = E(I_{[Y(t)=1,Z=z]} r(\beta,t)) = \bar{r}(\beta,t)P_z H_z(t)$. By Condition (B), we have $b > 0$. Also notice that $\frac{a}{b} = \bar{r}(\beta,t)$, (18) and (19), we have

$$sup_{\mathcal{B}\times[0,1]}||\hat{\bar{r}}(\beta,t) - \bar{r}(\beta,t)|| \xrightarrow{p} 0 \tag{20}$$

Since $\hat{r}_i^*(\beta,t) = r_i(\beta,t)I_{[v]} + \hat{\bar{r}}_i(\beta,t)I_{[\bar{v}]}$, so (20) implies

$$sup_{\mathcal{B}\times[0,1]}||\hat{r}^*(\beta,t) - r^*(\beta,t)|| \xrightarrow{p} 0.$$

The same argument works for $\hat{r}^{*(1)}(\beta,t)$ and $\hat{r}^{*(2)}(\beta,t)$. □

Theorem 4

$$sup_{\mathcal{B}\times[0,1]}||\hat{S}^{(j)} - S^{(j)}|| \xrightarrow{p} 0$$

$$sup_{\mathcal{B}\times[0,1]}||S^{(j)} - s^{(j)}|| \xrightarrow{p} 0$$

$$sup_{\mathcal{B}\times[0,1]}||\hat{S}^{(j)} - s^{(j)}|| \xrightarrow{p} 0$$

where $j = 0, 1, \cdots, 6$.

Lemma 1 $s^{(j)}(\cdot,t)$, $j = 0,1,2,\cdots,6,7$ *are continuous functions of* $\beta \in \mathcal{B}$, *uniformly in* $t \in [0,1]$, $s^{(j)}$, $j = 0,1,2,\cdots,6,7$, *are bounded on* $\mathcal{B}\times[0,1]$; $s^{(0)}(\beta,t)$ *is bounded away from zero on* $\mathcal{B}\times[0,1]$.

Lemma 2

$$sup_{\beta\in\mathcal{B}}||\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1) - (-\Sigma(\beta))|| \xrightarrow{p} 0$$

Lemma 3

$$\frac{1}{n}\hat{U}(\beta_0, 1) \xrightarrow{p} 0 as\ n \to \infty$$

The proofs of Theorem 4, Lemma 1, 2 and 3 are given in Appendix A.

3.2. Consistency of $\hat{\beta}_{EPL}$

Theorem 5 (Consistency Theorem) *There exists a sequence $\{\hat{\beta}_n\}$ such that $\hat{U}_n(\hat{\beta}_n) = 0$ with probability going to one as $n \to \infty$ and*

$$\hat{\beta}_n \xrightarrow{p} \beta_0$$

If $\{\bar{\beta}_n\}$ also satisfies the above conditions then $\hat{\beta}_n = \bar{\beta}_n$ with probability going to one as $n \to \infty$.

PROOF: By Condition (C), $-\Sigma(\beta_0)$ is negative definite, we may define $\lambda = \frac{1}{4\|-\Sigma^{-1}(\beta_0)\|}$. Choose δ sufficiently small that $\|\Sigma(\beta_0) - \Sigma(\beta)\| < \frac{\lambda}{2}$ whenever $|\beta - \beta_0| < \delta$ and that the uniform convergence of $\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta)$ to $-\Sigma(\beta)$ proved in Lemma 2 holds for $|\beta - \beta_0| < \delta$. As in Theorem 2.2, call the neighborhood of β_0 with radius δ as $\mathbf{U}_\delta$.

Also since $\frac{\partial}{\partial\beta}\hat{U}(\beta_0)$ converges to $-\Sigma(\beta_0)$ in probability, it ensures that $\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta_0)$ is negative definite with probability going to one. Let $\lambda_n = \frac{1}{4\|\frac{\partial}{\partial\beta}\hat{U}^{-1}(\beta)\|}$ whenever $\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta)$ is negative definite. Then $\lambda_n \xrightarrow{p} \lambda$. Hence we have

$$\begin{aligned}\|\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta) - \frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta_0)\| &\leq \|\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta) - (-\Sigma(\beta))\| \\ &\quad + \|\Sigma(\beta) - \Sigma(\beta_0)\| + \| - \Sigma(\beta_0) - \frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta_0)\| \\ &< \lambda < 2\lambda_n\end{aligned}$$

with probability going to one as $n \to \infty$ if $|\beta - \beta_0| < \delta$. By the Inverse Function Theorem stated in Theorem 2.2 $\frac{1}{n}\hat{U}$ is a one-to-one function from $\mathbf{U}_\delta$ on to $\frac{1}{n}\hat{U}(\mathbf{U}_\delta)$ and the image set $\frac{1}{n}\hat{U}(\mathbf{U}_\delta)$ contains the open neighborhood of radius $\lambda_n\delta$ about $\frac{1}{n}\hat{U}(\beta_0)$ with probability going to one.

By Lemma 3 we can see that $0 \in \hat{U}(\mathbf{U}_\delta)$ with probability going to 1. Also we have that $|\frac{1}{n}\hat{U}(\beta_0) - 0| < \frac{\lambda\delta}{2}$ with probability going to one as $n \to \infty$.

Consider the inverse function $\frac{1}{n}\hat{U}^{-1} : \frac{1}{n}\hat{U}(\mathbf{U}_\delta) \to \mathbf{U}_\delta$. It is well defined when $\frac{1}{n}\hat{U}$ is one-to-one, i.e. with probability going to one, since $0 \in \hat{U}(\mathbf{U}_\delta)$ with probabilty going to one we may conclude: (A) the root, $\frac{1}{n}\hat{U}^{-1}(0)$, of the estimating equation exists in $\mathbf{U}_\delta$ with probability going to one; (B) since δ may be taken arbitrarily small, $\frac{1}{n}\hat{U}^{-1}(0)$ converges in probability to β_0; (C) by one-to-oneness of $\frac{1}{n}\hat{U}$ on $\mathbf{U}_\delta$, any other sequence $\{\bar{\beta}\}$ of roots to $\hat{U}(\beta) = 0$ necessarily lies outside of $\mathcal{U}_\delta$ with probability going to 1 which implies that the sequence does not converge to β_0. Therefore $\hat{\beta}_{EPL} = \frac{1}{n}\hat{U}^{-1}(0)$ is a unique consistent estimator of β_0. □

4. ASYMPTOTIC NORMALITY OF $\hat{\beta}_{EPL}$

There are three subsection in this section: Section 4.1 presents the preliminary results. Section 4.2 proves the asymptotic normality of $\hat{\beta}_{EPL}$ and a consistent asymptotic covariance matrix estimate is given in Section 4.3.

The asymptotic normality of $\hat{\beta}_{EPL}$ is obtained through three main Theorems. In Theorem 6, the $n^{-\frac{1}{2}}\hat{U}(\beta_0, 1)$ is shown to be equivalent in probability to two independent summations. Theorem 7 further develops the asymptotic distribution of $n^{-\frac{1}{2}}\hat{U}(\beta_0, 1)$. Recall that the second task in showing asymptotic normality of $\hat{\beta}_{EPL}$ is to show

$$-n^{-\frac{1}{2}}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)\,|_{\beta=\beta^*} \xrightarrow{p} \Sigma(\beta_0).$$

This is done in Theorem 8. The main difficulty lies in proving Theorem 6, where we establish five preliminary lemmas for the proof. Lemma 4 - 8 are given in the preliminary section and the proofs of the Lemmas can be found in Appendix B.

4.1. Preliminary Results

Lemma 4

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^1(\hat{r}_i^{*(K)}(\beta_0,\omega) - r_i^{*(K)}(\beta_0,\omega))^2 Y_i(\omega) r_i^*(\beta_0,\omega)\lambda_0(\omega)d\omega \xrightarrow{p} 0 \quad K=0,1$$

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^1(\hat{S}^{(K)}(\beta_0,\omega) - S^{(K)}(\beta_0,\omega))^2 Y_i(\omega) r_i^*(\beta_0,\omega)\lambda_0(\omega)d\omega \xrightarrow{p} 0 \quad K=0,1$$

Lemma 5

$$\begin{aligned} & n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^1 \Delta(\hat{r}_i(\beta_0,\omega))Y_i(\omega)r_i^*(\beta_0,\omega)\lambda_0(\omega)d\omega \\ \stackrel{p}{=}\; & -n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^1 \Delta(r_i^*(\beta_0,\omega))Y_i(\omega)(\hat{r}_i^*(\beta_0,\omega) - r_i^*(\beta_0,\omega))\lambda_0(\omega)d\omega \end{aligned} \tag{21}$$

Lemma 6

$$\begin{aligned} & -n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^1 \Delta(r_i^*(\beta_0,\omega))Y_i(\omega)(\hat{r}_i^*(\beta_0,\omega) - r_i^*(\beta_0,\omega))\lambda_0(\omega)d\omega \\ \stackrel{p}{=}\; & -n^{-\frac{1}{2}}\sum_{i=1}^{\bar{v}}\int_0^1 \frac{\Delta(\bar{r}_i(\beta_0,\omega))}{p_{Z_i}H_{Z_i}(\omega)}Y_i(\omega) \\ & \times\frac{1}{v}\sum_{j=1}^{v} I_{[Z_j=Z_i]}Y_j(\omega)(r_j(\beta_0,\omega) - \bar{r}_i(\beta_0,\omega))\lambda_0(\omega)d\omega \end{aligned}$$

Lemma 7

$$n^{-\frac{1}{2}}\int_0^1\left(\frac{\hat{S}^{(1)}(\beta_0,\omega)}{\hat{S}^{(0)}(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}\right)d\bar{M}(\omega) \xrightarrow{p} 0$$

$$n^{-\frac{1}{2}}\int_0^1\left(\frac{S^{(1)}(\beta_0,\omega)}{S^{(0)}(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}\right)d\bar{M}(\omega) \xrightarrow{p} 0$$

Lemma 8

$$-n^{-\frac{1}{2}}\frac{\bar{v}}{v}\sum_{j=1}^{v}(Q_j^{\bar{v}} - Q_j) \xrightarrow{p} 0$$

where Q_j and $Q_j^{\bar{v}}$ are as defined in Section 2.2.1.

4.2. Asymptotic Normality of $\hat{\beta}_{EPL}$

Now we are ready to prove the asymptotic normality of $\hat{\beta}_{EPL}$.

Theorem 6

$$
\begin{aligned}
n^{-\frac{1}{2}}\hat{U}(\beta_0,1) \ &\overset{p}{=}\ n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^1\left(\frac{r_i^{*(1)}(\beta_0,\omega)}{r_i^*(\beta_0,\omega)}-\frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}\right)dM_i(\omega)\\
&\quad -n^{-\frac{1}{2}}\sum_{i=1}^{v}\frac{\bar{v}}{v}Q_j
\end{aligned}
$$

where Q_j is defined as in Section 2.2.1.

<u>PROOF OF THEOREM 6:</u>

$$
\begin{aligned}
& n^{-\frac{1}{2}}\hat{U}(\beta_0,1)\\
=\ & n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^1\left(\frac{\hat{r}_i^{*(1)}(\beta_0,\omega)}{\hat{r}_i^*(\beta_0,\omega)}-\frac{\hat{S}^{(1)}(\beta_0,\omega)}{\hat{S}^{(0)}(\beta_0,\omega)}\right)dM_i(\omega)\\
& +n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^1\left(\frac{\hat{r}_i^{*(1)}(\beta_0,\omega)}{\hat{r}_i^*(\beta_0,\omega)}-\frac{\hat{S}^{(1)}(\beta_0,\omega)}{\hat{S}^{(0)}(\beta_0,\omega)}\right)Y_i(\omega)r_i^*(\beta_0,\omega)\lambda_0(\omega)d\omega
\end{aligned}
\tag{22}
$$

Note that

$$
W(t)=n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t\left(\frac{\hat{r}_i^{*(1)}(\beta_0,\omega)}{\hat{r}_i^*(\beta_0,\omega)}-\frac{r_i^{*(1)}(\beta_0,\omega)}{r_i^*(\beta_0,\omega)}\right)dM_i(\omega)
$$

is a local Martingale with covariance at $t=1$

$$
\begin{aligned}
<W>(1)\ &=\ \int_0^1\sum_{i=1}^{n}\frac{1}{n}\left(\frac{\hat{r}_i^{*(1)}(\beta_0,\omega)}{\hat{r}_i^*(\beta_0,\omega)}-\frac{r_i^{*(1)}(\beta_0,\omega)}{r_i^*(\beta_0,\omega)}\right)^2Y_i(\omega)r_i^*(\beta_0,\omega)\lambda_0(\omega)d\omega\\
&=\ \int_0^1\frac{1}{n}\sum_{i=1}^{\bar{v}}\left(\frac{\hat{\bar{r}}_i^{(1)}(\beta_0,\omega)}{\hat{\bar{r}}_i(\beta_0,\omega)}-\frac{\bar{r}_i^{(1)}(\beta_0,\omega)}{\bar{r}_i(\beta_0,\omega)}\right)^2Y_i(\omega)\bar{r}_i(\beta_0,\omega)\lambda_0(\omega)d\omega\\
&=\ \int_0^1\sum_{k=1}^{q}\left(\frac{\hat{\bar{r}}_{z_k}^{(1)}(\beta_0,\omega)}{\hat{\bar{r}}_{z_k}(\beta_0,\omega)}-\frac{\bar{r}_{z_k}^{(1)}(\beta_0,\omega)}{\bar{r}_{z_k}(\beta_0,\omega)}\right)^2\bar{r}_{z_k}(\beta_0,\omega)\\
&\quad\ \times\frac{1}{n}\sum_{i=1}^{\bar{v}}I_{[Z_i=z_k]}Y_i(\omega)\lambda_0(\omega)d\omega\\
&\xrightarrow{p}\ 0
\end{aligned}
$$

by Theorem 3 and the fact that $\frac{1}{n}\sum_{i=1}^{\bar{v}}I_{[Z_i=z_k]}Y_i(\omega)$ converges to $(1-\rho)p_{z_k}H_{z_k}(\omega)$ uniformly in ω and Condition A. Hence $W(\cdot)\xrightarrow{p}0$ by Lenglart Inequality. The above result together with Theorem 4, the first term of (22) equals to

$$
n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^1\left(\frac{r_i^{*(1)}(\beta_0,\omega)}{r_i^*(\beta_0,\omega)}-\frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}\right)dM_i(\omega)
$$

in probability.

The second term in (22) by Lemma 6

$$
\begin{aligned}
&= -n^{-\frac{1}{2}} \sum_{i=1}^{\bar{v}} \int_0^1 \left(\frac{\bar{r}_i^{(1)}(\beta_0,\omega)}{\bar{r}_i(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right) Y_i(\omega) \\
&\quad \times \frac{1}{v} \sum_{j=1}^{v} I_{[Z_j=Z_i]} Y_i(\omega) \frac{(r_j(\beta_0,\omega) - \bar{r}_i(\beta_0,\omega))}{P_{Z_i} H_{Z_i}(\omega)} \lambda_0(\omega) d\omega + o_p(1) \\
&= -n^{-\frac{1}{2}} \frac{\bar{v}}{v} \sum_{j=1}^{\bar{v}} Q_j^{\bar{v}} + o_p(1)
\end{aligned}
$$

Furthermore by Lemma 8, the Theorem holds. □

Theorem 7

$$
n^{-\frac{1}{2}} \hat{U}(\beta_0, 1) \xrightarrow{D} N_p(0, (1-\rho)\Sigma_1 + \rho\Sigma_2)
$$

where the variance-covariance matrices Σ_1 and Σ_2 are as defined in (15) and (16).

PROOF: From Theorem 6, we have

$$
\begin{aligned}
& n^{-\frac{1}{2}} \hat{U}(\beta_0, 1) \\
&= n^{-\frac{1}{2}} \int_0^1 \sum_{i=1}^{\bar{v}} \left(\frac{\bar{r}_i^{(1)}(\beta_0,\omega)}{\bar{r}_i(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right) dM_i(\omega) \\
&\quad + n^{-\frac{1}{2}} \sum_{j=1}^{v} \left[\int_0^1 \left(\frac{r_j^{(1)}(\beta_0,\omega)}{r_j(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right) dM_i(\omega) - \frac{\bar{v}}{v} Q_j \right] + o_p(1) \qquad (23)
\end{aligned}
$$

The first term and the second term on the right hand side are independent. By the martingale central limit theorem, the first term in (23) converges weakly to a continuous normal process. The covariance process of this normal process evaluated at $t = 1$ is $(1-\rho)\Sigma_1$. i.e.

$$
\begin{aligned}
< W > (t) &= \int_0^t n^{-1} \sum_{i=1}^{\bar{v}} \left(\frac{\bar{r}_i^{(1)}(\beta_0,\omega)}{\bar{r}_i(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right)^2 Y_i(\omega) \bar{r}_i(\beta_0,\omega) \lambda_0(\omega) d\omega \\
&\xrightarrow{p} (1-\rho) \int_0^t \left[E\left(\frac{\bar{r}_i^{(1)}(\beta_0,\omega)^{\otimes 2}}{\bar{r}_i(\beta_0,\omega)} Y_i(\omega) \right) - \frac{s^{(1)}(\beta_0,\omega)^{\otimes 2}}{s^{(0)}(\beta_0,\omega)} \right] \lambda_0(\omega) d\omega \\
< W > (1) &\equiv (1-\rho)\Sigma_1
\end{aligned}
$$

The second term in $n^{-\frac{1}{2}} \hat{U}(\beta_0, 1)$ is also a summation of iid terms from subjects in the validation sample. By the Central Limit Theorem, it converges to a normal distribution with mean

$$
E\left\{ \int_0^1 \left(\frac{r_j^{(1)}(\beta_0,\omega)}{r_j(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right) dM_j(\omega) - \frac{\bar{v}}{v} Q_j \right\} \qquad (24)
$$

and covariance

$$
\rho Var\left\{ \int_0^1 \left(\frac{r_j^{(1)}(\beta_0,\omega)}{r_j(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right) dM_j(\omega) - \frac{1-\rho}{\rho} Q_j \right\} \qquad (25)
$$

The first term in the mean expression (25) is a local martingale and expected value of a local martingale is zero. The second term is also zero, since

$$
\begin{aligned}
& E\left\{-\frac{\bar{v}}{v}Q_j\right\} \\
= & -\frac{\bar{v}}{v}E\int_0^1 Y_j(\omega)(r_j(\beta_0,\omega)-\bar{r}_j(\beta_0,\omega))\left(\frac{\bar{r}_j^{(1)}(\beta_0,\omega)}{\bar{r}_j(\beta_0,\omega)}-\frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}\right)\lambda_0(\omega)d\omega \\
= & -\frac{\bar{v}}{v}\int_0^1 E\left\{\left(\frac{\bar{r}_j^{(1)}(\beta_0,\omega)}{\bar{r}_j(\beta_0,\omega)}-\frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}\right)\right. \\
& \times \left. E\left[Y_j(\omega)r_j(\beta_0,\omega)-Y_j(\omega)\bar{r}_j(\beta_0,\omega)\,|_{Y_j(\omega)=1,Z_j=z_j}\right]\right\}\lambda_0(\omega)d\omega \\
\equiv & \ 0
\end{aligned}
$$

The covariance matrix can be expressed as

$$\rho\Sigma_2\equiv\rho E\left\{\int_0^1\left(\frac{r_j^{(1)}(\beta_0,\omega)}{r_j(\beta_0,\omega)}-\frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}\right)dM_j(\omega)-\frac{1-\rho}{\rho}Q_j\right\}^{\otimes 2}.$$

Since the two terms in $n^{-\frac{1}{2}}\hat{U}(\beta_0,\omega)$ are from the validation and non-validation sets respectively the limiting distribution of $n^{-\frac{1}{2}}\hat{U}(\beta_0,\omega)$ is normal with mean zero and covariance matrix $(1-\rho)\Sigma_1+\rho\Sigma_2$, where Σ_1 and Σ_2 are as defined in (15) and (16). □

The second step in proving the asymptotic nomality of $n^{\frac{1}{2}}(\hat{\beta}_{EPL}-\beta_0)$ is to show that $-n^{-1}\frac{\partial}{\partial\beta}\hat{U}(\beta,\cdot)|_{\beta=\beta^*}$ converges to some positive definite quantity. The theorem below gives the limit of $-n^{-1}\frac{\partial}{\partial\beta}\hat{U}(\beta,\cdot)|_{\beta=\beta^*}$.

Theorem 8

$$-n^{-1}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)|_{\beta=\beta^*}\xrightarrow{p}\Sigma(\beta_0)\ \ as\ n\to\infty$$

where β^* lies between $\hat{\beta}_{EPL}$ and β_0, and $\Sigma(\beta_0)$ is positive definite.

<u>PROOF:</u> Recall from Lemma 2 that $-n^{-1}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)\xrightarrow{p}\Sigma(\beta)$ for any $\beta\in\mathcal{B}$ and that $\Sigma(\beta_0)$ is positive definite, where

$$\Sigma(\beta_0)=\int_0^1\left[\frac{s^{(2)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}-\left(\frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)}\right)^{\otimes 2}\right]s^{(0)}(\beta_0,\omega)\lambda_0(\omega)d\omega$$

From Lemma 1, $\Sigma(\beta)$ is continuous in β.

$$
\begin{aligned}
& \left|-\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)|_{\beta=\beta^*}-\Sigma(\beta_0)\right| \\
\leq & \left|-\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)|_{\beta=\beta^*}-\Sigma(\beta^*)\right|+|\Sigma(\beta^*)-\Sigma(\beta_0)|
\end{aligned}
$$

The first term on the right hand side of the inequality converges to zero in probability as $n\to\infty$ by Lemma 2. The second term converges to zero as well by the continuity of Σ and the fact that β^* is between $\hat{\beta}_{EPL}$ and β_0, and that $\hat{\beta}_{EPL}$ is consistent for β_0 by Theorem 2.5 . Therefore

$$-n^{-1}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)|_{\beta=\beta^*}\xrightarrow{p}\Sigma(\beta_0)\ \text{ as } n\to\infty$$

where $\Sigma(\beta_0)$ is positive definite. □

Theorem 9 (Asymptotic Normality Theorem)

$$\sqrt{n}(\hat{\beta}_{EPL} - \beta_0) \xrightarrow{D} N_p(0, \Sigma_{EPL}(\beta_0))$$

where $\Sigma_{EPL}(\beta_0) = \Sigma^{-1}(\beta_0)((1-\rho)\Sigma_1 + \rho\Sigma_2)\Sigma^{-1}(\beta_0)^T$ *as* $n \to \infty$.

PROOF:

$$n^{-\frac{1}{2}}(\hat{\beta}_{EPL} - \beta_0) = [-n^{-1}\frac{\partial}{\partial\beta}\hat{U}(\beta^*, 1)]^{-1} n^{-\frac{1}{2}}\hat{U}(\beta_0, 1)$$

by Theorem 7 and 8, the result is straightforward. □

4.3. Asymptotic Variance Estimator

As in the ordinary partial likelihood setting, the variance estimator needs an estimator for the baseline hazard function (or cumulative hazard $\Lambda_0(t) = \int_0^t \lambda_0(\omega)d\omega$). A continuous estimator obtained by linear interpolation between failure times of

$$\hat{\Lambda}_0(t) = \sum_{T_i \le t} \frac{\delta_i}{\sum_{j \in \mathcal{R}(T_i)} e^{\hat{\beta}x}}$$

for the underlying cumulative hazard was suggested by Breslow (1972, 1974) in the ordinary partial likelihood with relative risk $r = e^{\beta x}$ where $\hat{\beta}$ is the partial likelihood estimate. In this work, we propose to estimate $d\Lambda_0(t)$ by

$$d\hat{\hat{\Lambda}}_0(t) = \frac{\sum_{i=1}^n dN_i(t)}{\sum_{i=1}^n Y_i(t)\hat{r}_i^*(\hat{\beta}_{EPL}, t)} = \frac{1}{\hat{S}^{(0)}(\hat{\beta}_{EPL}, t)} \frac{1}{n} \sum_{i=1}^n dN_i(t) \tag{26}$$

Then estimation of Σ_{EPL} is done through empirically estimating each component in Σ_1, Σ_2 and Σ. That is, the expectation in Σ_1, Σ_2 and Σ are estimated by taking the average in the observed data, r^* is estimated by $\hat{r}^*$, $s^{(j)}$ are estimated by $\hat{S}^{(j)}$. We define

$$\begin{aligned}
\hat{\Sigma}_1(\beta) &= \int_0^1 \left(\sum_{k=1}^q \frac{\hat{\hat{r}}_k^{(1)}(\beta,t)^{\otimes 2}}{\hat{\hat{r}}_k(\beta,t)} \frac{1}{n} \sum_{j=1}^n Y_j(t) I_{[Z_j = Z_k]} - \frac{\hat{S}^{(1)}(\beta,t)^{\otimes 2}}{\hat{S}^{(0)}(\beta,t)} \right) d\hat{\hat{\Lambda}}_0(t) \\
\hat{\Sigma}_2(\beta) &= \frac{1}{v} \sum_{j=1}^v \left\{ \int_0^1 \left(\frac{r_i^{(1)}(\beta,t)}{r_i(\beta,t)} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} \right) dM_i(t) \right. \\
&\quad - \frac{1-\rho}{\rho} \int_0^1 \int_0^1 \left(\frac{\hat{\hat{r}}_i^{(1)}(\beta,t)}{\hat{\hat{r}}_i(\beta,t)} - \frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} \right) \\
&\quad \left. \times Y_j(t)(r_i(\beta,t) - \hat{\hat{r}}_i(\beta,t) d\hat{\hat{\Lambda}}_0(t) \right\}^{\otimes 2} \\
\hat{\Sigma}(\beta) &= \int_0^1 \left(\hat{S}^{(4)}(\beta,t) - \frac{\hat{S}^{(1)}(\beta,t)^{\otimes 2}}{\hat{S}^{(0)}(\beta,t)} \right) d\hat{\hat{\Lambda}}_0(t)
\end{aligned} \tag{27}$$

Then the estimator of $\Sigma_{EPL}(\beta_0)$ is given by

$$\hat{\Sigma}_{EPL}(\hat{\beta}_{EPL}) = \hat{\Sigma}(\hat{\beta}_{EPL})^{-1}((1-\rho)\hat{\Sigma}_1(\hat{\beta}_{EPL}) + \rho\hat{\Sigma}_2(\hat{\beta}_{EPL}))(\hat{\Sigma}(\hat{\beta}_{EPL})^{-1})^T \tag{28}$$

The following theorem shows that $\hat{\Sigma}_{EPL}(\hat{\beta}_{EPL})$ is a consistent variance estimator of $\Sigma_{EPL}(\beta_0)$.

Theorem 10

$$\hat{\Sigma}_{EPL}(\hat{\beta}_{EPL}) \xrightarrow{p} \Sigma_{EPL}(\beta_0) \quad as\ n \to \infty$$

The proof of the Theorem is provided in Appendix C.

5. ASYMPTOTIC THORY WITH STRATIFIED VALIDATION SET

Suppose that the validation set is not a simple random sample of the entire study sample as we assumed thus far in this chapter. Rather it is a random sample stratified on the auxiliary Z. Let d_i, $i = 1, \cdots, q$ and $0 \leq d_i \leq 1$, denotes the predecided probability that goes to the validation set for a subject with auxiliary z_i in the study. i.e. $100 \times d_i$ percent subjects who have the auxiliary value z_i will be randomly chosen to be in the validation set; and the other $100 \times (1 - d_i)$ percent subjects will be in the non-validation set. Hence the validation size n_v in this case is

$$v = \sum_{i=1}^{q} d_i \times p_i \times n$$

where n is the entire sample size and $p_i = Prob(Z = z_i)$.

Let $\hat{\beta}^s_{EPL}$ denotes the EPL estimator obtained from the above stratified set up. Since the subjects both in the validation set and non-validation set are still indenpendent and identically distributed, the proof of the asympotic theory in the Section 2.3 and 2.4 are still valid here except that the asymtotic variance of $\hat{\beta}^s_{EPL}$ is different with $\hat{\beta}_{EPL}$. The large sample theory of $\hat{\beta}^s_{EPL}$ are given by following theorems.

Theorem 11 *The $\hat{\beta}^s_{EPL}$ is consistent estimator of β.*

PROOF: See proof of Theorm 5. □

Theorem 12

$$\sqrt{n}(\hat{\beta}^s_{EPL} - \beta) \xrightarrow{D} N_p(0, \Sigma^s_{EPL}) as\ n \to \infty.$$

where $\Sigma^s_{EPL} = \Sigma^{-1}((1 - \sum_{i=1}^{q} d_i \times p_i)\Sigma_1 + \sum_{i=1}^{q} d_i \times p_i \Sigma^s_2)(\Sigma^{-1})^T$ *and*

$$\Sigma^s_2 \equiv E\left\{ \int_0^1 \left(\frac{r_j^{(1)}(\beta,\omega)}{r_j(\beta,\omega)} - \frac{s^{(1)}(\beta,\omega)}{s^{(0)}(\beta,\omega)} \right) dM_j(\omega) - \frac{1 - \sum_{i=1}^{q} d_i \times p_i}{\sum_{i=1}^{q} d_i \times p_i} Q_j \right\}^{\otimes 2}.$$

Σ, Σ_1 *and* Q_j *are the same as defined in Section 2.2*

PROOF: The proof of this theorem is analogous to the proof of Theorem 9. □

The proposed estimator of the asymptotic variance $\Sigma^s_{EPL}(\beta)$ is given by

$$\hat{\Sigma}^s_{EPL}(\hat{\beta}^s_{EPL}) = \hat{\Sigma}(\hat{\beta}^s_{EPL})^{-1}((1 - \sum_{i=1}^{q} \hat{d}_i \times \hat{p}_i)\hat{\Sigma}_1(\hat{\beta}^s_{EPL}) + \rho\hat{\Sigma}^s_2(\hat{\beta}^s_{EPL}))(\hat{\Sigma}(\hat{\beta}^s_{EPL})^{-1})^T \qquad (29)$$

where $\hat{d}_i$ and $\hat{p}_i$ are the empirical estimates from the sample. i.e. $\hat{p}_i$ is the proportion of subjects in the entire sample that have the value z_i; $\hat{d}_i$ is the ratio of the number of subjects in the validation set with z_i devided by the number of subjects in the entire sample who have z_i.

The following theorem shows that $\hat{\Sigma}^s_{EPL}(\hat{\beta}^s_{EPL})$ is a consistent variance estimator of $\Sigma^s_{EPL}(\beta)$.

Theorem 13

$$\hat{\Sigma}^s_{EPL}(\hat{\beta}^s_{EPL}) \xrightarrow{p} \Sigma^s_{EPL}(\beta) \quad as\ n \to \infty$$

<u>PROOF:</u> The proof of this theorem is analogous to the proof of Theorem 10. □

4. DISCUSSION

The finite condition (A) is easy to justify in practice. The infinite interval case cannot be derived from the finite interval case by a simple mapping. This is because $\int_0^\infty \lambda_0(t)dt = \infty$ in general. Some extra conditions have to be made ensuring that the contribution to the test statistics from data on (τ, ∞) can be made arbitrarily small, uniformly in n, by taking τ large enough. Anderson and Gill(1982) outlined the proof of such extension when the covariate is bounded in the case of ordinary partial likelihood with complete covariate data(Theorem 4.2 in Anderson and Gill, 1982). A reference of the tightness condition on $[\tau, \infty)$ can be found in Fleming and Harrington(1991).

Recall that the empirical estimation of induced relative risk function $\bar{r}$ requires for a subject with Z at time t in the non-validation set that there exist subjects still at risk and sharing the same auxiliary Z in the validation set. Hence condition (B) is a natural assumption ensuring that EPL can be carried out in large samples.

The regularity condition (C) on boundedness and continuity allows the interchange of orders of various limiting operations.

Condition (D) ensures the asymptotic stability of $S^{(j)}, j = 0, 1, , \cdots, 7$. Note that the uniformity of convergence in β is not needed in the asymptotic normality proofs. However the consistency proof needs the uniformity of convergence in β. Prentice and Self(1983) considered a general form of relative risk function r, rather than $e^{\beta X}$ as in Anderson and Gill. The asymptotic stability assumption we have assumed in this work is not more than that in Prentice and Self(1983). This makes sense because after all the relative risk r^* in this work is of a general form.

Finally Condition (E), the uniformly boundedness of $Z_v^{(k)}(t)$, is needed to ensure the asymptotic normality. By modern empirical theory this assumption is satisfied in some general set ups. For example if $r(\beta, x) = e^{\beta x}$ and x is time independent, then it can be shown that the Condition (E) holds. A survey of these results can be found in Wellner[10].

Appendix A. Proof of Priliminary Results for Consistency

<u>PROOF OF THEOREM 4:</u> The proof of this theorem is only demonstrated for $j = 0$. For the other cases of j, the proofs are analogous.

$$\begin{aligned}
\hat{S}^{(0)} - S^{(0)} &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\hat{r}_i^*(\beta, t) - \frac{1}{n}\sum_{i=1}^{n} Y_i(t)r_i^*(\beta, t) \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)(\hat{r}_i^*(\beta, t) - r_i^*(\beta, t)) \\
&= \frac{1}{n}\sum_{i=1}^{v} Y_i(t)(\hat{\bar{r}}_i(\beta, t) - \bar{r}_i(\beta, t)) \\
&= \sum_{k=1}^{q}(\hat{\bar{r}}_k(\beta, t) - \bar{r}_k(\beta, t))\frac{1}{n}\sum_{i=1}^{v} Y_i(t)I_{[Z_i = z_k]}
\end{aligned}$$

By Theorem 3,

$$\|\hat{\bar{r}}_k(\beta,t) - \bar{r}_k(\beta,t)\|_{\mathcal{B}\times[0,1]} \xrightarrow{a.s.} 0 \text{ for any } k = 1,\cdots,q.$$

Since

$$\|\hat{S}^{(0)} - S^{(0)}\| \le \sum_{k=1}^{q} \|\hat{\bar{r}}_k(\beta,t) - \bar{r}_k(\beta,t)\| \|\frac{1}{n}\sum_{i=1}^{v} Y_i(t) I_{[Z_i = z_k]}\|$$

and

$$\|\frac{1}{n}\sum_{i=1}^{v} Y_i(t) I_{[Z_i = z_k]}\| \le 1$$

Then

$$sup_{\mathcal{B}\times[0,1]}\|\hat{S}^{(0)} - S^{(0)}\| \xrightarrow{p} 0.$$

The same argument works for $\hat{S}^{(j)}$, where $j = 1,2,\cdots,6$.

The second convergence in the Lemma can be shown using the same types of arguments as in Theorem 3. As in Theorem 3, we can treat $Y_i(t)r_i^*(\beta,t)$ as random elements of $D[0,1]$ (after reversing the time axis!). By condition (D) and Theorem 1, we can show that

$$sup_{\mathcal{B}\times[0,1]}\|S^{(j)} - s^{(j)}\| \xrightarrow{p} 0$$

for $j = 0,1,2,\cdots,6,7$.

The third expression of the theorem is an immediate consequence of the first two results. □

<u>PROOF LEMMA 1:</u> Recall that

$$s^{(0)} = E(Y(t)r(X(t)\beta)$$

Since $r^*(X(t)\beta)$ is continuous in β on $\mathcal{B}$, for any $\beta \to \beta^*$ in $\mathcal{B}$, we have

$$sup_t|Y(t)r^*(X(t)\beta)| \xrightarrow{p} sup_t|Y(t)r^*(X(t)\beta^*)| \text{ as } \beta \to \beta^*.$$

From Condition (D)

$$E(sup_{[0,1]\times\mathcal{B}}Y(t)r^*(X(t)\beta)) < \infty$$

Hence by the Dominated Convergence Theorem(Chung, 1977), $s^{(0)}(\beta) \xrightarrow{p} s^{(0)}(\beta^*)$ as $\beta \to \beta^*$, so $s^{(0)}$ is continuous function of $\beta \in \mathcal{B}$ for each β uniformly in $t \in [0,1]$. The same argument applies to $s^{(j)}$ $j = 1,2,\cdots,6,7$.

Also by Condition (D), $s^{(j)}$, $j = 0,1,2,\cdots,6,7$, are bounded on $\mathcal{B}\times[0,1]$.

Since $\bar{r}(\beta,t)$ is bounded away from zero, we have

$$\begin{aligned} \bar{r}(\beta,t) &= E(r(\beta,t)|Y(t) = 1, Z = z) \\ &= E(Y(t)r(\beta,t)|Z = z)/P(Y(t) = 1|Z = z) \end{aligned}$$

By Condition (B) and (C) that $P(Y(t) = 1|Z = z) > 0$ and $\bar{r}(\beta,t)$ is bounded away from zero, we have $E(Y(t)r(\beta,t)|Z = z)$ bounded away from zero. Hence $s^{(0)} = E(E(Y(t)r(\beta,t)|Z = z))$ is bounded away from zero. □

<u>PROOF LEMMA 2:</u>

$$\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1) = \frac{1}{n}\int_0^1 \sum_{i=1}^{n} \left[\frac{\hat{r}_i^{*(2)}(\beta,t)\hat{r}_i^*(\beta,t) - \hat{r}_i^{*(1)}(\beta,t)^{\otimes 2}}{\hat{r}_i^*(\beta,t)^2} \right.$$

$$- \frac{\hat{S}^{(2)}(\beta,t)\hat{S}^{(0)}(\beta,t) - \hat{S}^{(1)}(\beta,t)^{\otimes 2}}{\hat{S}^{(0)}(\beta,t)^2} \Bigg] dN_i(t)$$

$$= \int_0^1 \frac{1}{n}\sum_{i=1}^n \left[\frac{\hat{r}_i^{*(2)}(\beta,t)}{\hat{r}_i^*(\beta,t)} - \left(\frac{\hat{r}_i^{*(1)}(\beta,t)}{\hat{r}_i^*(\beta,t)}\right)^{\otimes 2} - \frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} + \left(\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right)^{\otimes 2} \right] dN_i(t)$$

Define

$$A(\beta,1) = \int_0^1 \frac{1}{n}\sum_{i=1}^n \left[\frac{\hat{r}_i^{*(2)}(\beta,t)}{\hat{r}_i^*(\beta,t)} - \left(\frac{\hat{r}_i^{*(1)}(\beta,t)}{\hat{r}_i^*(\beta,t)}\right)^{\otimes 2} - \frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} + \left(\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right)^{\otimes 2} \right] Y_i(t) r_i^*(\beta_0,t)\lambda_0(t)dt$$

Then using (9) one can write

$$\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1) - A(\beta,1) = \int_0^1 \frac{1}{n}\sum_{i=1}^n \left[\frac{\hat{r}_i^{*(2)}(\beta,t)}{\hat{r}_i^*(\beta,t)} - \left(\frac{\hat{r}_i^{*(1)}(\beta,t)}{\hat{r}_i^*(\beta,t)}\right)^{\otimes 2} - \frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} + \left(\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right)^{\otimes 2} \right] dM_i(t)$$

Since (C) ensures the $\hat{r}_i^{*(j)}(\beta,t)$ and $\hat{S}^{(j)}(\beta,t)$ to be predictable and locally bounded for $\beta \in \mathcal{B}$, therefore $\frac{\partial}{\partial\beta}\hat{U}(\beta,1) - A(\beta,1)$ is a local square integrable martingale with variance process $B(\beta,1)$ given by

$$B(\beta,1) = \int_0^1 \frac{1}{n^2}\sum_{i=1}^n \left[\frac{\hat{r}_i^{*(2)}(\beta,t)}{\hat{r}_i^*(\beta,t)} - \left(\frac{\hat{r}_i^{*(1)}(\beta,t)}{\hat{r}_i^*(\beta,t)}\right)^{\otimes 2} - \frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} + \left(\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right)^{\otimes 2} \right]^2 \lambda_i(t)dt$$

We would like to show that $B(\beta,1) \xrightarrow{p} 0$, so that $\frac{1}{n}\frac{\partial}{\partial\beta}\hat{U}(\beta,1)$ and $A(\beta,1)$ would be shown to converge in probability to the same limit. Expanding the squared term in the above expression gives

$$\begin{aligned} B(\beta,1) = & \int_0^1 \frac{1}{n^2}\sum_{i=1}^n \left(\frac{\hat{r}_i^{*(2)}(\beta,t)}{\hat{r}_i^*(\beta,t)}\right)^{\otimes 2} r_i^*(\beta_0,t)\lambda_0(t)dt \\ & + \int_0^1 \frac{1}{n^2}\sum_{i=1}^n \left(\frac{\hat{r}_i^{*(1)}(\beta,t)}{\hat{r}_i^*(\beta,t)}\right)^{\otimes 4} r_i^*(\beta_0,t)\lambda_0(t)dt \\ & + \int_0^1 \frac{1}{n}\left(\frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right)^{\otimes 2} S^{(0)}(\beta_0,t)\lambda_0(t)dt \\ & + \int_0^1 \frac{1}{n}\left(\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right)^{\otimes 4} S^{(0)}(\beta_0,t)\lambda_0(t)dt \\ & + \text{ the interaction terms of the 4 components in above} \end{aligned}$$

The final two integrals in this expression converges in probability to zero uniformly in β, in view of Lemma 1, 3, on $S^{(j)}$ along with the finite interval Condition (A). The interaction term will converged to zero if the first two terms do upon applying the Schwarz inequality and the convergence just noted in the last two terms.

The first two terms above are $\frac{1}{n}\hat{S}^{(5)}$, $\frac{1}{n}\hat{S}^{(6)}$ respectively. By Theorem 4 and Lemma 1 and Condition (A), the first two terms converge to zero uniformly in $\mathcal{B}$. It now follows that $||B(\beta,1)||_{\mathcal{B}}\xrightarrow{p}0$ so that by an inequality of Lenglart (Appendix 1 of Anderson and Gill, 1982) $\frac{\partial}{\partial\beta}\hat{U}(\beta,1)$ converges in probability to the same limit as does $A(\beta,1)$, uniformly in $\beta\in\mathcal{B}$.

From Lemma 1 and 2 and Condition (A)

$$\begin{aligned} A(\beta,1) &= \int_0^1[\hat{S}^{(3)}(\beta,t)-\frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}S^{(0)}(\beta_0,t) \\ &\quad -\hat{S}^{(4)}(\beta,t)+(\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)})^{\otimes 2}S^{(0)}(\beta_0,t)]\lambda_0(t)dt \\ &\xrightarrow{p} \int_0^1[s^{(3)}(\beta,t)-\frac{s^{(2)}(\beta,t)}{s^{(0)}(\beta,t)}s^{(0)}(\beta_0,t) \\ &\quad -s^{(4)}(\beta,t)+(\frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)})^{\otimes 2}s^{(0)}(\beta_0,t)]\lambda_0(t)dt \\ &\equiv -\Sigma(\beta) \quad \text{uniformly in } \beta \end{aligned}$$

since $s^{(3)}(\beta_0,t)=s^{(2)}(\beta_0,t)$, $\Sigma(\beta)$ at $\beta=\beta_0$ equals

$$\int_0^1[\frac{s^{(4)}(\beta_0,t)}{s^{(0)}(\beta_0,t)}-(\frac{s^{(1)}(\beta_0,t)}{s^{(0)}(\beta_0,t)})^{\otimes 2}]s^{(0)}(\beta_0,t)\lambda(t)dt=\Sigma(\beta_0)$$

which is positive definite by Condition (C). □

<u>PROOF LEMMA 3:</u> Recall

$$\frac{1}{n}\hat{U}(\beta_0,1)=\frac{1}{n}\sum_{i=1}^{n}\int_0^1\left(\frac{\hat{r}_i^{*(1)}(\beta_0,t)}{\hat{r}_i^{*}(\beta_0,t)}-\frac{\hat{S}^{(1)}(\beta_0,t)}{\hat{S}^{(0)}(\beta_0,t)}\right)dN_i(t)$$

Define

$$A(\beta_0,1)=\frac{1}{n}\sum_{i=1}^{n}\int_0^1\left(\frac{\hat{r}_i^{*(1)}(\beta_0,t)}{\hat{r}_i^{*}(\beta_0,t)}-\frac{\hat{S}^{(1)}(\beta_0,t)}{\hat{S}^{(0)}(\beta_0,t)}\right)Y_i(t)r_i^*(\beta_0,t)\lambda_0 dt$$

then by similar argument as in Lemma 2, $\hat{U}(\beta_0,1)-A(\beta_0,1)$ is a local square integrable martingale with covariance process $B(\beta_0,1)$ given by

$$\begin{aligned} B(\beta_0,1) &= \int_0^1\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{\hat{r}_i^{*(1)}(\beta_0,t)}{\hat{r}_i^{*}(\beta_0,t)}-\frac{\hat{S}^{(1)}(\beta_0,t)}{\hat{S}^{(0)}(\beta_0,t)}\right)^{\otimes 2}Y_i(t)r_i^*(\beta_0,t)\lambda_0 dt \\ &= \int_0^1\left[\frac{1}{n}\hat{S}^{(3)}(\beta_0,t)+\frac{1}{n}\left(\frac{\hat{S}^{(1)}(\beta_0,t)}{\hat{S}^{(0)}(\beta_0,t)}\right)^{\otimes 2}S^{(0)}(\beta_0,t)\right. \\ &\quad \left.-2\frac{1}{n}\hat{S}^{(7)}(\beta_0,t)\frac{\hat{S}^{(1)}(\beta_0,t)}{\hat{S}^{(0)}(\beta_0,t)}\lambda_0(t)dt\right. \end{aligned}$$

Each term converges to zero by noting Theorem 4 and Lemma 1 and Condition (A). Hence $\frac{1}{n}\hat{U}(\beta_0, 1)$ converges to the same limit as $A(\beta_0, 1)$. Observe that

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{r}_i^{*(1)}(\beta_0,t)}{\hat{r}_i^{*}(\beta_0,t)}Y_i(t)r_i^{*}(\beta_0,t) \\
= \; & \frac{1}{n}\sum_{i=1}^{v}Y_i(t)r_i^{*(1)}(\beta_0,t) + \frac{1}{n}\sum_{i=1}^{\bar{v}}Y_i(t)\frac{\hat{\bar{r}}_i^{(1)}(\beta_0,t)}{\hat{\bar{r}}_i(\beta_0,t)}\bar{r}_i(\beta_0,t) \\
= \; & \frac{1}{n}\sum_{i=1}^{v}Y_i(t)r_i^{*(1)}(\beta_0,t) + \frac{\bar{v}}{n}\sum_{k=1}^{q}\frac{n_q}{\bar{v}}\frac{1}{n_q}\sum_{j=1}^{\bar{v}}Y_j(t)\frac{\hat{\bar{r}}_k^{(1)}(\beta_0,t)}{\hat{\bar{r}}_k(\beta_0,t)}\bar{r}_k(\beta_0,t)I_{[Z_j=z_k]} \\
\xrightarrow{p} \; & \rho E(Y(t)r^{(1)}(\beta_0,t)) \\
& +(1-\rho)\sum_{k=1}^{q}p_{z_k}H_{z_k}(t)\bar{r}_q^{(1)}(\beta_0,t) \text{ uniformly in t by Theorem 3 and (D)} \\
= \; & E(Y(t)r^{(1)}(\beta_0,t)) \\
= \; & s^{(1)}(\beta_0,t)
\end{aligned}
$$

Hence

$$A(\beta_0, 1)\xrightarrow{p}\int_0^1(s^{(1)}(\beta_0,t) - s^{(1)}(\beta_0,t))\lambda_0(t)dt = 0$$

$\hat{U}(\beta_0, 1)\xrightarrow{p}0$. □

Appendix B. Proof of Priliminary Results for Normality

PROOF LEMMA 4:

The first expression

$$
\begin{aligned}
= \; & n^{-\frac{1}{2}}\sum_{i=1}^{\bar{v}}\int_0^1(\hat{\bar{r}}_i^{(K)}(\beta_0,\omega) - \bar{r}_i^{(K)}(\beta_0,\omega))^2Y_i(\omega)\bar{r}_i(\beta_0,\omega)\lambda_0(\omega)d\omega \\
= \; & n^{-\frac{1}{2}}\int_0^1\sum_{k=1}^{q}(\hat{\bar{r}}_{z_k}^{(K)}(\beta_0,\omega) - \bar{r}_{z_k}^{(K)}(\beta_0,\omega))^2\bar{r}_{z_k}(\beta_0,\omega) \\
& \times\sum_{i=1}^{\bar{v}}I_{[Z_i=z_k]}Y_i(\omega)\lambda_0(\omega)d\omega \\
= \; & \sum_{k=1}^{q}\int_0^1\sqrt{n}(\hat{\bar{r}}_{z_k}^{(K)}(\beta_0,\omega) - \bar{r}_{z_k}^{(K)}(\beta_0,\omega))(\hat{\bar{r}}_{z_k}^{(K)}(\beta_0,\omega) - \bar{r}_{z_k}^{(K)}(\beta_0,\omega))\bar{r}_k(\beta_0,\omega) \\
& \times\frac{1}{n}\sum_{i=1}^{\bar{v}}I_{[Z_i=z_k]}Y_i(\omega)\lambda_0(\omega)d\omega
\end{aligned}
$$

where q is the index value of Z, the first term in above expression is $O_p(1)$ by condition (E), the second term converges to zero uniformly in ω by Theorem 3, Condition (D) implies that $\bar{r}_k(\beta_0,\omega)$ is bounded uniformly in ω, the last term is bounded uniformly in ω. Finally with Condition (A) the first expression converges to zero in probability.

The second expression in the Lemma

$$
\begin{aligned}
&= \sqrt{n}\int_0^1 (\hat{S}^{(K)}(\beta_0,\omega) - S^{(K)}(\beta_0,\omega))^2 S^{(0)}(\beta_0,\omega)\lambda_0(\omega)d\omega \\
&= \sqrt{n}\int_0^1 [\frac{1}{n}\sum_{i=1}^{n} Y_i(\omega)(\hat{r}_i^{*(K)}(\beta_0,\omega) - r_i^{*(K)}(\beta_0,\omega))] \\
&\quad \times (\hat{S}^{(k)}(\beta_0,\omega) - S^{(k)}(\beta_0,\omega))S^{(0)}(\beta_0,\omega)\lambda_0(\omega)d\omega \\
&= \sqrt{n}\int_0^1 [\sum_{k=1}^{q}(\hat{\bar{r}}_{z_k}^{(K)}(\beta_0,\omega) - \bar{r}_{z_k}^{(K)}(\beta_0,\omega))] \\
&\quad \times \frac{1}{n}\sum_{i=1}^{\bar{v}} I_{[Z_i=z_k]}Y_i(\omega)(\hat{S}^{(k)}(\beta_0,\omega) - S^{(k)}(\beta_0,\omega))S^{(0)}(\beta_0,\omega)\lambda_0(\omega)d\omega \\
&\xrightarrow{p} 0
\end{aligned}
$$

by Theorem 4, Lemma 1, Condition (A) and a similar argument as was used above for the first expression. □

<u>PROOF OF LEMMA 5:</u> By Taylor expansion

$$
\begin{aligned}
f(x,y) &= f(x_0,y_0) + \left.\frac{\partial f(x,y)}{\partial x}\right|_{(x_0,y_0)}(x-x_0) \\
&\quad + \left.\frac{\partial f(x,y)}{\partial y}\right|_{(x_0,y_0)}(y-y_0) + O((x-x_0)^2 + (y-y_0)^2)
\end{aligned}
$$

if $\frac{\partial^2}{\partial x^2}f$, $\frac{\partial^2}{\partial y^2}f$, and $\frac{\partial^2}{\partial x \partial y}f$ are finite. We have

$$
\frac{\hat{r}^{*(1)}}{\hat{r}^{*}} = \frac{\hat{r}^{*(1)}}{r^{*}} - \frac{r^{*(1)}(\hat{r}^{*} - r^{*})}{r^{*2}} + O[(\hat{r}^{*} - r^{*})^2 + (\hat{r}^{*(1)} - r^{*(1)})^2]
$$

$$
\frac{\hat{S}^{(1)}}{\hat{S}^{0}} = \frac{\hat{S}^{(1)}}{S^{0}} - \frac{S^{(1)}(\hat{S}^{0} - S^{0})}{S^{0 2}} + O[(\hat{S}^{0} - S^{0})^2 + (\hat{S}^{(1)} - S^{(1)})^2]
$$

With the above expansion, the left side of (21) can be expressed as

$$
\begin{aligned}
&-n^{-\frac{1}{2}}\int_0^1 \sum_{i=1}^{n}\left(\frac{r_i^{*(1)}(\beta_0,\omega)}{r_i^{*}(\beta_0,\omega)} - \frac{S^{(1)}(\beta_0,\omega)}{S^{(0)}(\beta_0,\omega)}\right)Y_i(\omega) \\
&\times (\hat{r}_i^{*}(\beta_0,\omega) - r_i^{*}(\beta_0,\omega))\lambda_0(\omega)d\omega + \text{ the remainder}
\end{aligned}
$$

From Lemma 4, the remainder goes to zero in probability. Therefore the result holds. □

<u>PROOF OF LEMMA 6:</u> Notice that $\hat{r}_i^{*}(\beta_0,\omega) - r_i^{*}(\beta_0,\omega) = 0$ for $i \in V$, the left side of (21) is

$$
left = -n^{-\frac{1}{2}}\sum_{i=1}^{\bar{v}}\int_0^1 \Delta(\bar{r}_i(\beta_0,\omega))Y_i(\omega)(\hat{\bar{r}}_i(\beta_0,\omega) - \bar{r}_i(\beta_0,\omega))\lambda_0(\omega)d\omega
$$

while the right hand side of (21) is

$$
\begin{aligned}
right &= -n^{-\frac{1}{2}}\sum_{i=1}^{\bar{v}}\int_0^1 \Delta(\bar{r}_i(\beta_0,\omega))Y_i(\omega) \\
&\quad \times \frac{\frac{1}{v}\sum_{j=1}^{v} I_{[Z_j=Z_i]}Y_j(\omega)}{p_{Z_i}H_{Z_i}(\omega)}(\hat{\bar{r}}_i(\beta_0,\omega) - \bar{r}_i(\beta_0,\omega))\lambda_0(\omega)d\omega
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& left - right \\
= \quad & -n^{-\frac{1}{2}} \sum_{i=1}^{\bar{v}} \int_0^1 \Delta(\bar{r}_i(\beta_0,\omega)) Y_i(\omega)(\hat{\bar{r}}_i(\beta_0,\omega) - \bar{r}_i(\beta_0,\omega)) \\
& \times \frac{\frac{1}{v}\sum_{j=1}^{v} I_{[Z_j=Z_i]} Y_j(\omega) - p_{Z_i} H_{Z_i}(\omega)}{p_{Z_i} H_{Z_i}(\omega)} \lambda_0(\omega) d\omega \\
= \quad & -\sum_{k=1}^{q} \int_0^1 \Delta(\bar{r}_{z_k}(\beta_0,\omega)) \sqrt{n}(\hat{\bar{r}}_{z_k}(\beta_0,\omega) - \bar{r}_{z_k}(\beta_0,\omega)) \\
& \times (\frac{1}{v}\sum_{j=1}^{v} I_{[Z_j=Z_k]} Y_j(\omega) - p_{Z_k} H_{Z_k}(\omega)) \frac{\frac{1}{n}\sum_{i=1}^{\bar{v}} I_{[Z_i=z_k]} Y_i(\omega)}{p_{z_k} H_{z_k}(\omega)} \lambda_0(\omega) d\omega \\
\xrightarrow{p} \quad & 0
\end{aligned}
$$

by noting (18) and Condition (E). □

<u>PROOF OF LEMMA 7:</u> Since $\bar{M}(\omega)$ is a local martingale and $\frac{\hat{S}^{(1)}}{\hat{S}^{(0)}} - \frac{s^{(1)}}{s^{(0)}}$ is predictable, then

$$
W = n^{-\frac{1}{2}} \int_0^1 \left(\frac{\hat{S}^{(1)}(\beta_0,\omega)}{\hat{S}^{(0)}(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right) d\bar{M}(\omega)
$$

is a martingale with covariance process $< W > (1)$.

$$
\begin{aligned}
& < W > (1) \\
= \quad & \int_0^1 \frac{1}{n} \left[\frac{\hat{S}^{(1)}(\beta_0,\omega)}{\hat{S}^{(0)}(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right]^2 d < \bar{M}, \bar{M} > \\
= \quad & \int_0^1 \left(\frac{\hat{S}^{(1)}(\beta_0,\omega)}{\hat{S}^{(0)}(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right)^2 S^{(0)}(\beta_0,\omega)\lambda_0(\omega) dw
\end{aligned}
$$

By Theorem 4 and Condition A,

$$
< W > (1) \xrightarrow{p} 0
$$

Therefore the Lemma holds by Lenglart Inequality. The proof of the second expression in Lemma is analogous. □

<u>PROOF OF LEMMA 8:</u>

$$
\begin{aligned}
& -n^{-\frac{1}{2}} \frac{\bar{v}}{v} \sum_{j=1}^{v} Q_j^{\bar{v}} \\
= \quad & -\int_0^1 n^{-\frac{1}{2}} \sum_{j=1}^{v} \sum_{i=1}^{\bar{v}} \left(\frac{\bar{r}_i^{(1)}(\beta_0,\omega)}{\bar{r}_i(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right) Y_i(\omega) \\
& \times \frac{1}{v} I_{[Y_j(\omega), Z_j=Z_i]} \frac{(r_j(\beta_0,\omega) - \bar{r}_i(\beta_0,\omega))}{p_{Z_i} H_{Z_i}(\omega)} \lambda_0(\omega) d\omega \\
= \quad & -\int_0^1 n^{-\frac{1}{2}} \sum_{i=1}^{\bar{v}} \left(\frac{\bar{r}_i^{(1)}(\beta_0,\omega)}{\bar{r}_i(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right) / (p_{Z_i} H_{Z_i}(\omega)) Y_i(\omega) \\
& \times \frac{1}{v} \sum_{j=1}^{v} I_{[Y_j(\omega), Z_j=Z_i]} (r_j(\beta_0,\omega) - \bar{r}_i(\beta_0,\omega)) \lambda_0(\omega) d\omega
\end{aligned}
$$

$$= -\int_0^1 \sum_{k=1}^{q} \left(\frac{\bar{r}_{z_k}^{(1)}(\beta_0,\omega)}{\bar{r}_{z_k}(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right) / (p_{z_k} H_{z_k}(\omega))$$
$$\times \sqrt{n}(\frac{1}{v}\sum_{j=1}^{v} I_{[Y_j(\omega)=1, Z_j=z_k]}(r_j(\beta_0,\omega) - \bar{r}_{z_k}(\beta_0,\omega)))$$
$$\times n^{-1}\sum_{i=1}^{\bar{v}} I_{[Y_i(\omega)=1, Z_i=z_k]}\lambda_0(\omega)d\omega$$

Observe

$$\frac{1}{n}\sum_{i=1}^{\bar{v}} I_{[Y_i(\omega)=1, Z_i=z_k]} \xrightarrow{p} (1-\rho)p_{z_k}H_{z_k}(\omega) \text{ uniformly in } \omega$$

Moreover with

$$Z^v = \sqrt{n}(\frac{1}{v}\sum_{j=1}^{v} I_{[Y_j(\omega)=1, Z_j=z_k]}(r_j(\beta_0,\omega) - \bar{r}_{z_k}(\beta_0,\omega)))$$

as was assumed in Condition (E) that $sup_{0\leq\omega\leq 1}|Z^v| = O_p(1)$. Hence it is clear that

$$-n^{-\frac{1}{2}}\frac{\bar{v}}{v}\sum_{j=1}^{v} Q_j^{\bar{v}}$$
$$= -\int_0^1 \sum_{k=1}^{q} \left(\frac{\bar{r}_{z_k}^{(1)}(\beta_0,\omega)}{\bar{r}_{z_k}(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right)$$
$$\times \frac{\bar{v}}{v} n^{-\frac{1}{2}} \sum_{j=1}^{v} I_{[Y_j(\omega)=1, Z_j=z_k]}(r_j(\beta_0,\omega) - \bar{r}_{z_k}(\beta_0,\omega))\lambda_0(\omega)d\omega + o_p(1)$$
$$= -\frac{\bar{v}}{v} n^{-\frac{1}{2}} \int_0^1 \sum_{j=1}^{v} Y_j(\omega)(r_j(\beta_0,\omega) - \bar{r}_j(\beta_0,\omega))$$
$$\times \left(\frac{\bar{r}_j^{(1)}(\beta_0,\omega)}{\bar{r}_j(\beta_0,\omega)} - \frac{s^{(1)}(\beta_0,\omega)}{s^{(0)}(\beta_0,\omega)} \right) \lambda_0(\omega)d\omega + o_p(1)$$
$$= -n^{-\frac{1}{2}}\frac{\bar{v}}{v}\sum_{j=1}^{v} Q_j + o_p(1)$$

Therefore

$$-n^{-\frac{1}{2}}\frac{\bar{v}}{v}\sum_{j=1}^{v}(Q_j^{\bar{v}} - Q_j) \xrightarrow{p} 0.$$

The result holds. □

APPENDIX C: PROOFS OF VARIANCE ESTIMATOR

PROOF: First, we will show that

$$\hat{\Sigma}_1(\beta) \xrightarrow{p} \Sigma_1(\beta) \text{ as } n \to \infty$$

for any β. Note that

$$\hat{\Sigma}_1(\beta) - \Sigma_1(\beta) = \int_0^1 \left[\left(E(\frac{\bar{r}^{(1)}(\beta,t)^{\otimes 2}}{\bar{r}(\beta,t)} Y(t)) \right. \right.$$

$$
\begin{aligned}
&- \sum_{k=1}^{q} \frac{\hat{\bar{r}}_{z_k}^{(1)}(\beta,t)^{\otimes 2}}{\hat{\bar{r}}_{z_k}(\beta,t)} \frac{1}{n}\sum_{j=1}^{n} Y_j(t) I_{[Z_j=z_k]} \Bigg) \\
&+ \left(\frac{\hat{S}^{(1)}(\beta,t)^{\otimes 2}}{\hat{S}^{(0)}(\beta,t)} - \frac{s^{(1)}(\beta,t)^{\otimes 2}}{s^{(0)}(\beta,t)} \right) \Bigg] d\Lambda_0(\omega) \\
&+ \int_0^1 \sum_{k=1}^{q} \frac{\hat{\bar{r}}_{z_k}^{(1)}(\beta,t)^{\otimes 2}}{\hat{\bar{r}}_{z_k}(\beta,t)} \frac{1}{n}\sum_{j=1}^{n} Y_j(t) I_{[Z_j=z_k]} \\
&\times (\lambda_0(t)dt - \frac{1}{\hat{S}^{(0)}(\beta,t)} \sum_{i=1}^{n} dN_i(t))
\end{aligned} \tag{30}
$$

By Theorem 3 and (18), we have

$$
\begin{aligned}
& \sum_{k=1}^{q} \frac{\hat{\bar{r}}_{z_k}^{(1)}(\beta,t)^{\otimes 2}}{\hat{\bar{r}}_{z_k}(\beta,t)} \frac{1}{n}\sum_{j=1}^{n} Y_j(t) I_{[Z_j=z_k]} \\
\xrightarrow{p} \;& \sum_{k=1}^{q} \frac{\bar{r}_{z_k}^{(1)}(\beta,t)^{\otimes 2}}{\bar{r}_{z_k}(\beta,t)} P(Y(t)=1, Z=z_k) \\
= \;& E\left(\frac{\bar{r}_{z_k}^{(1)}(\beta,t)^{\otimes 2}}{\bar{r}_{z_k}(\beta,t)} Y(t) \right) \quad \text{uniformly in } t
\end{aligned} \tag{31}
$$

With Theorem 7 and above result, the first integration in (30) converges to 0 in probability. By (26) the second integration in (30) can be expressed as

$$
\begin{aligned}
& \int_0^1 \sum_{k=1}^{q} \frac{\hat{\bar{r}}_{z_k}^{(1)}(\beta,t)^{\otimes 2}}{\hat{\bar{r}}_{z_k}(\beta,t)} \frac{1}{n}\sum_{j=1}^{n} Y_j(t) I_{[Z_j=z_k]} (d\Lambda_0(t) \\
& - \frac{1}{\hat{S}^{(0)}(\beta,t)} S^{(0)}(\beta,t) d\Lambda_0(t) - \frac{1}{\hat{S}^{(0)}(\beta,t)} \frac{1}{n}\sum_{i=1}^{n} dM_i(t)) \\
= \;& \int_0^1 \sum_{k=1}^{q} \frac{\hat{\bar{r}}_{z_k}^{(1)}(\beta,t)^{\times 2}}{\hat{\bar{r}}_{z_k}(\beta,t)} \frac{1}{n}\sum_{j=1}^{n} Y_j(t) I_{[Z_j=z_k]} ((1 - \frac{S^{(0)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}) d\Lambda_0(t) \\
& - \frac{1}{\hat{S}^{(0)}(\beta,t)} \frac{1}{n}\sum_{i=1}^{n} dM_i(t))
\end{aligned} \tag{32}
$$

By Theroem 7 and (31), the first term in (32) converges to zero for any β. The second term is a local martingale. It can be shown as in the proof of Lemma 2.2 that this local martingale converges to zero. This is done by showing the covariance process of this local martingale equals to

$$
\begin{aligned}
& \frac{1}{n^2} \int_0^1 \left(\sum_{k=1}^{q} \frac{\hat{\bar{r}}_{z_k}^{(1)}(\beta,\omega)^{\otimes 2}}{\hat{\bar{r}}_{z_k}(\beta,\omega)} \frac{1}{n}\sum_{j=1}^{n} Y_j(\omega) I_{[Z_j=z_k]} \right. \\
& \left. - \frac{1}{\hat{S}^{(0)}(\beta,\omega)} \right)^2 S^{(0)}(\beta,\omega) \\
= \;& \frac{1}{n} \int_0^1 \left(\frac{1}{n}\sum_{j=1}^{n} Y_j(\omega) \frac{\hat{\bar{r}}_{z_k}^{(1)}(\beta,\omega)^{\otimes 2}}{\hat{\bar{r}}_{z_k}(\beta,\omega)} - \frac{1}{\hat{S}^{(0)}(\beta,\omega)} \right)^2 S^{(0)}(\beta,\omega)
\end{aligned}
$$

$$\xrightarrow{p} 0$$

By noticing that $\hat{S}^{(0)}(\beta,t)\xrightarrow{p}s^{(0)}(\beta,t)$ uniformly in t, $s^{(0)}(\beta,t)$ is bounded away from zero and that

$$\frac{1}{n}\sum_{j=1}^{n}Y_j(\omega)\frac{\hat{\bar{r}}_{z_k}^{(1)}(\beta,\omega)^{\otimes 2}}{\hat{\bar{r}}_{z_k}(\beta,\omega)}\xrightarrow{p}E(Y(t)\frac{\hat{\bar{r}}^{(1)}(\beta,t)^{\otimes 2}}{\hat{\bar{r}}(\beta,t)})\leq s^{(4)}(\beta,t)$$

where $s^{(4)}(\beta,t)$ by Lemma 2.1 is uniformly bounded on $[0,1]\times\mathcal{B}$. Therefore by Lenglart Inequality, we have that the second term in (32) converges to zero in probability for any β. This implies

$$|\hat{\Sigma}_1(\beta)-\Sigma_1(\beta)|\xrightarrow{p}0.$$

Similarly we can show that

$$|\hat{\Sigma}_2(\beta)-\Sigma_2(\beta)|\xrightarrow{p}0$$
$$|\hat{\Sigma}(\beta)-\Sigma(\beta)|\xrightarrow{p}0$$

With the convergence of $\hat{\Sigma}_1(\beta)$, $\hat{\Sigma}_2(\beta)$, and $\hat{\Sigma}(\beta)$, we have

$$|\hat{\Sigma}_{EPL}(\beta)-\Sigma_{EPL}(\beta)|\xrightarrow{p}0. \tag{33}$$

Hence

$$\begin{aligned}
& |\hat{\Sigma}_{EPL}(\hat{\beta}_{EPL})-\Sigma_{EPL}(\beta_0)| \\
\leq\;& |\hat{\Sigma}_{EPL}(\hat{\beta}_{EPL})-\Sigma_{EPL}(\hat{\beta}_{EPL})|+|\Sigma_{EPL}(\hat{\beta}_{EPL})-\Sigma_{EPL}(\beta_0)| \\
\xrightarrow{p}\;& 0 \text{ as } n\to\infty
\end{aligned}$$

by the fact that (33) holds and that Σ_{EPL} is continuous in β and $\hat{\beta}_{EPL}\xrightarrow{p}\beta_0$ as $n\to\infty$. □

References

[1] Andersen, P.K. and Gill, R.D. (1982), "Cox's Regression Model for Counting Processes: A Large Sample Study", *The Annals of Statistics*, 10, 1100-1120.

[2] Caroll, R. J. and Wand, M. P. (1991), "Semiparametric Estimation in Logistic Measurement Error Models", *J. R. Statist. Soc. B*, 53, 573-585

[3] Fleming, T.R. and Harrington, D.P. (1991), *Counting Processes and Survival Analysis*, Wiley, New York.

[4] Foutz, R. V. (1977), "On the Unique Consistent Solution to the Likelihood Equations", *Journal of the American Statistical Association*, 72, 147-148.

[5] GAUSS System, Version 2.1 (1991), Kent, WA: Aptech Systems Inc.

[6] Kalbfleisch, J. D. (1974), "Some Efficiency Calculations for Survival Distributions", *Biometrika*, 61, 31-38.

[7] Prentice, R. L. (1982), "Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model" *Biometrika* 69, 2, 331-342

[8] Rudin, W (1964), *Principles of Mathematical Analysis*, New York: McGraw-Hill Book Co..

[9] The SOLVD Investors (1991) "Effect of Enalapril on Surbival In patients with reduced Left Ventricular Ejection Fraction and Congestive Heart Failure", *New England Iournal of Madicine* , 325:293-302(August 1), 1991

[10] Jon Wellner Manuscript. "Empirical Processes In Action: A Review"

[11] Haibo Zhou (1992), "Auxiliary Covariate Data in Failure Time Regression Analysis" *Ph.D Dissertation, Unversity of Washington,*