BIOMETRICS xx, 1–??   DOI: xxxxxxxx.x
August 2016

Web-based Supplementary Materials for "FLCRM: Functional Linear Cox Regression Model" by Dehan Kong, Joseph G. Ibrahim, Eunjee Lee and Hongtu Zhu

Dehan Kong, Joseph G. Ibrahim, Eunjee Lee, Hongtu Zhu,

Department of Statistical Sciences, University of Toronto,

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

Department of Biostatistics, University of Michigan, and

Department of Biostatistics, The University of Texas MD Anderson Cancer Center

SUMMARY: In this Web Supplement, we include additional simulation results in Section 1. We include detailed description of the ADNI data analysis in Section 2. Additional real data results are shown in Section 3. We then include the discussion of asymptotic theories and conditions in Section 4. We then state all auxiliary lemmas in Section 5. The proofs of all lemmas are included in Section 6. We include the proofs of Theorems 1 and 2 in Section 7. Finally, we provide proofs of Corollaries 1 and 2 in Section 8.

1. Additional simulation results

In this section, we report the finite sample performance of our method by using AIC as r_n varies from 1 to 10. Tables S1-S2 summarize all results when we set $n = 200$ and censoring rates 0.3 and 0.5. Tables S3-S5 summarize all results when we set $n = 500$ and censoring rates 0.1, 0.3, and 0.5. Tables S6-S8 summarize all results when we set $n = 1000$ and censoring rates 0.1, 0.3 and 0.5.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

We include additional simulation results on the power of our test statistic as the number of functional principle components varies. We have reported the results for two settings. Table S9 summarizes the case with $n = 200$, censoring rate 0.1, and $C_1 = 0.1 * j$ for $j = 0, \ldots, 10$. Table S10 summarizes the case with $n = 200$, censoring rate 0.1, and $C_2 = 0.1 * j$ for $j = 0, \ldots, 10$. Inspecting Tables S9 and S10 that the power of our test statistic is relatively robust to the choice of PV(r_n). There is a little power loss if PV$(r_n) = 0.95$. We also ran additional simulations for many other settings corresponding to different sample sizes, different censoring rates. Since their corresponding findings are similar to those presented here, we omit them for simplicity.

[Table 9 about here.]

[Table 10 about here.]

2. Data Description for ADNI Data Analysis

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). "The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year publicprivate partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org."

2.1 *Hippocampus image preprocessing*

The MRI data, collected across a variety of 1.5 Tesla MRI scanners with protocols individualized for each scanner, includes standard T1-weighted images obtained by using volumetric 3-dimensional sagittal MPRAGE or equivalent protocols with varying resolutions. The typical protocol includes: inversion time (TI) = 1000 ms, flip angle = 8^o, repetition time (TR) = 2400 ms, and field of view (FOV) = 24 cm with a $256 \times 256 \times 170$ acquisition matrix in the $x-$, $y-$, and $z-$dimensions yielding a voxel size of $1.25 \times 1.26 \times 1.2$ mm^3. We adopted a surface fluid registration based hippocampal subregional analysis package (Shi et al., 2013), which uses isothermal coordinates and fluid registration to generate one-to-one hippocampal surface registration. Given the 3D MRI scans, hippocampal substructures were segmented with FIRST (Patenaude et al., 2011) and hippocampal surfaces were automatically reconstructed with the marching cube method (Lorensen and Cline, 1987). We applied an automatic algorithm, topology optimization, to introduce two cuts on a hippocampal surface to convert it into a genus zero surface with two open boundaries. The locations of the two cuts were at the front and back of the hippocampal surface, representing its anterior junction with the amygdala, and its posterior limit as it turns into the white matter of the fornix. Then holomorphic 1-form basis functions were computed (Wang et al., 2010). These induced conformal grids the hippocampal surfaces, which were consistent across subjects. With this conformal grid, we computed the conformal representation of the surface (Shi et al., 2013), i.e., the conformal factor and mean curvature, which represent the intrinsic and extrinsic features of the surface, respectively. The "feature image" of a surface was computed by combining the conformal factor and mean curvature and linearly scaling the dynamic range into $[0, 255]$. Next, we registered the feature image of each surface in the dataset to a common template with an inverse consistent fluid registration algorithm (Shi et al., 2013). With conformal parameterization, we essentially converted a 3D surface registration problem into a 2D image registration problem. The flow induced in the parameter domain establishes high-order correspondences between 3D surfaces. Finally, the radial distance, which retains information on the deformation along the surface normal direction, was computed on the registered surface.

2.2 *Demographic information summary*

Among all individuals, 303 were retired and 70 were not; 237 participants were male, and 136 were female; 342 were right-handed, and 31 were left-handed; 300 were married, 45 were widowed, 24 were divorced, and 4 were never married. The participants had an average of 15.7 years of education with a standard deviation 3.0 years. The minimum education length was 4 years and the maximum education length is 20 years. The average age was 75.0 years with a standard deviation of 7.3 years. The youngest person was 55 years old, while the oldest person was 90 years old. We also had genetic information on two alleles of APOE4. For the first allele, 26 had genotype 2, 300 had genotype 3, and 47 had genotype 4. For the second allele, 169 had genotype 3, and 204 had genotype 4. For the ADAS-Cog score, the average score was 11.6 with a standard deviation of 4.5, the lowest score was 2 and the highest score was 27.67. Mild cognitive impairment converters did not differ from mild cognitive impairment noncoverters in gender, handedness, marital status, retirement percentage, and age ($p-$value> 0.05), but as expected, differed from them in APOE4 status as well as baseline cognition ($p-$value< 0.05). Mean follow up time was 99 days longer in converters ($p-$value$= 0.007$).

3. Additional Real Data Results

We have plotted the estimated coefficient functions when $r_n = 17, 18, 19, 21, 22,$ and 23 in Figure S1. From the results, we can see that estimated coefficient functions are quite robust to the choice of r_n.

[Figure 1 about here.]

4. Asymptotic Properties

In this section, we systematically investigate the asymptotic properties of the maximum approximate partial likelihood estimator $\widehat{\eta}$ as well as the asymptotic null distribution of the score statistic T_S. It is assumed that all curves are fully observed just for notational simplicity. Such an assumption has been used in Hall et al. (2006) and Lei (2014), among others. Technically, when functional observations are dense in space, using smoothed curves is as good as using true curves under some smoothness conditions (Hall et al., 2006; Zhang and Chen, 2007).

For simplicity, we shall focus on a finite time interval $[0, \tau]$ with $\tau < \infty$. We consider a re-parametrization of β_j and ξ_{ij} by defining $\beta_{jR} = {\lambda_j}^{1/2}\beta_j$ and $\xi_{ijR} = \xi_{ij}/{\lambda_j}^{1/2}$ and then we have

$$\int_{\mathcal{S}} X_i(s)\beta(s)ds = \sum_{j=1}^{\infty} \xi_{ijR}\beta_{jR} = \sum_{j=1}^{r_n} \xi_{ijR}\beta_{jR} + \sum_{j=r_n+1}^{\infty} \xi_{ij}\beta_j.$$

We also define $\widehat{\xi}_{ijR} = \widehat{\xi}_{ij}/{\lambda_j}^{1/2}$ for $j = 1, \ldots, r_n$. The reason to do re-parametrization is to make the FPC scores serving as predictor variables on a common scale of variabilities. Denote

$\beta_0(s)$ as the true coefficient function. It is assumed that $||\beta_0|| = \{\int_{\mathcal{S}} \beta_0(s)^2 ds\}^{1/2} < \infty$ and $\beta_0(s) = \sum_{j=1}^{\infty} \beta_{j0}\phi_j(s)$. Recall that the hazard function of FLCRM can be rewritten as

$$h_i(t) = h_0^*(t) \exp(\sum_{k=1}^{p} z_{ik}\gamma_k + \sum_{j=1}^{\infty} \xi_{ij}\beta_j), \tag{4.1}$$

where $h_0^*(t) = h_0(t) \exp\{\int_{\mathcal{S}} \mu(s)\beta(s)ds\}$, and the log-approximate partial likelihood function $Q(\eta)$ is given by

$$Q(\eta) = \sum_{i=1}^{n} \int_0^{\tau} \widehat{w}_i^{\mathrm{T}} \eta dN_i(t) - \int_0^{\tau} \log\{\sum_{i=1}^{n} Y_i(t) \exp(\widehat{w}_i^{\mathrm{T}} \eta)\} d\overline{N}(t), \tag{4.2}$$

Therefore, the logarithm of approximate partial likelihood functions for (4.1) and (4.2) are, respectively, given by

$$\begin{aligned} l(\beta(\cdot), \gamma) &= \sum_{i=1}^{n} \int_0^{\tau} (\sum_{k=1}^{p} z_{ik}\gamma_k + \sum_{j=1}^{r_n} \xi_{ijR}\beta_{jR} + \sum_{j=r_n+1}^{\infty} \xi_{ij}\beta_j) dN_i(t) \\ &\quad - \int_0^{\tau} \log\{\sum_{i=1}^{n} Y_i(t) \exp(\sum_{k=1}^{p} z_{ik}\gamma_k + \sum_{j=1}^{r_n} \xi_{ijR}\beta_{jR} + \sum_{j=r_n+1}^{\infty} \xi_{ij}\beta_j)\} d\overline{N}(t), \end{aligned}$$

$$\begin{aligned} Q(\eta_R) &= \sum_{i=1}^{n} \int_0^{\tau} (\sum_{k=1}^{p} z_{ik}\gamma_k + \sum_{j=1}^{r_n} \widehat{\xi}_{ijR}\beta_{jR}) dN_i(t) \\ &\quad - \int_0^{\tau} \log\{\sum_{i=1}^{n} Y_i(t) \exp(\sum_{k=1}^{p} z_{ik}\gamma_k + \sum_{j=1}^{r_n} \widehat{\xi}_{ijR}\beta_{jR})\} d\overline{N}(t), \end{aligned}$$

where $\beta_{r_nR} = (\beta_{1R}, \cdots, \beta_{r_nR})^{\mathrm{T}}$ and $\eta_R = (\beta_{r_nR}, \gamma)$. The function $l(\beta(\cdot), \gamma)$ can also be regarded as a function of η_R, and we define $l(\eta_R)$ as

$$l(\eta_R) = l(\beta(\cdot), \gamma)|_{\{\beta_j = \beta_{j0}, r_n+1 \leqslant j < \infty\}}$$

.

The key ideas of our theoretical development are

(i) to characterize the discrepancy between $Q(\eta_R)$ and $l(\eta_R)$ as well as their first-order and second-order derivatives;

(ii) to prove the consistency and convergence rate of $\widehat{\beta}(s) = \sum_{j=1}^{r_n} \widehat{\beta}_j \widehat{\phi}_j(s)$ and $\widehat{\gamma}$;

(iii) to prove the asymptotic distribution of the score statistic for testing the null effect of a functional predictor.

We need three sets of conditions in order to achieve these developments. Without loss of generality, we assume that the predictor $X(t)$ has been centered. Define $e_i = \sum_{j=r_n+1}^{\infty} \xi_{ij}\beta_{j0}$. With a little abuse of notation, we use C to denote terms that are constant, and C may denote different constants in different places.

The first set of conditions includes the following conditions (A1)-(A5) on the survival data. The conditions (A1)-(A4) can be regarded as a direct extension of some standard conditions in the literature (Fan and Li, 2002; Andersen and Gill, 1982; Murphy and Van der Vaart, 2000).

(A1) $\int_0^\tau h_0(t)dt < \infty$.

(A2) Let $w_{iR} = (\xi_{i1R}, \ldots, \xi_{ir_nR}, z_{i1}, \ldots, z_{ip})^{\mathrm{T}}$. For $d = 0, 1$, and 2, we define

$$S^{(d)}(\eta_R, t) \;=\; n^{-1}\sum_{i=1}^{n} (w_{iR})^{(d)} Y_i(t) \exp(\eta_R^{\mathrm{T}} w_{iR} + e_i),$$

where $(w_{iR})^{(0)} = 1$, $(w_{iR})^{(1)} = w_{iR}$, and $(w_{iR})^{(2)} = (w_{iR})^{\otimes 2}$. Moreover, there exists a neighborhood $\boldsymbol{B}$ of the true value of η_R, denoted as η_{R0}, and a scalar, a vector and a matrix continuous function $s^{(d)}(\eta_R, t) = E\{S^{(d)}(\eta_R, t)\}$ defined on $\boldsymbol{B} \times [0, \tau]$ such that

$$\sup_{t\in[0,\tau], \eta_R \in \boldsymbol{B}} ||S^{(d)}(\eta_R, t) - s^{(d)}(\eta_R, t)|| \longrightarrow^p 0 \quad \text{for} \quad d = 0, 1, 2.$$

(A3) The functions $s^{(d)}$ for $d = 0, 1, 2$ are bounded on $\boldsymbol{B} \times [0, \tau]$ and $s^{(d)}(\cdot, t)$ are continuous in $\eta_R \in \boldsymbol{B}$ uniformly in $t \in [0, \tau]$. Moreover, $s^{(0)}$ is bounded away from 0 on $\boldsymbol{B} \times [0, \tau]$.

(A4) The matrix $\Sigma(\eta_{R0}) = \int_0^\tau v(\eta_{R0}, t)s^{(0)}(\eta_{R0}, t)h_0(t)dt$ is positive definite, where $v(\eta_R, t) = \{s^{(0)}\}^{-1}s^{(2)} - \{s^{(0)}\}^{-2}\{s^{(1)}\}^{\otimes 2}$.

(A5) For any $1 \leqslant k \leqslant p$, z_k is subgaussian.

Here, a random variable Z is said to be subgaussian if there exists some $M > 0$ such that for every $t \in \mathbb{R}$, one has $E(\{\exp(tZ)\}) \leqslant \exp(M^2t^2/2)$. In particular, we call Z as M-subgaussian. (A5*) For any $1 \leqslant k \leqslant p_n$, there exists a constant $M > 0$ such that z_k is M-subgaussian.

The second set of conditions (B1)-(B6) is imposed on the functional predictor, for example the boundedness of the covariance function $K(s,t)$ and the regression operators, that is, $\sum_{j=1}^{\infty} \lambda_j < \infty$ and $\sum_{j=1}^{\infty} \beta_{j0}^2 < \infty$. Conditions (B1) and (B2) are used in Hall and Horowitz (2007). Conditions (B3) and (B4) are used in Hall and Hosseini-Nasab (2006).

(B1) $\lambda_j - \lambda_{j+1} \geqslant Cj^{-a-1}$ for $j \geqslant 1$ with $a > 1$.

(B2) $|\beta_{j0}| \leqslant Cj^{-b}$ for $j > 1$.

(B3) For any $C > 0$, there exists an $\epsilon > 0$ such that

$$\sup_{s \in \mathcal{S}} \{E|X(s)|^C\} < \infty \quad \text{and} \quad \sup_{s_1, s_2 \in \mathcal{T}} (E[\{|s_1 - s_2|^{-\epsilon}|X(s_1) - X(s_2)|\}^C]) < \infty.$$

(B4) For each integer $d \geqslant 1$, $\lambda_j^{-d} E(\int_{\mathcal{T}} X_i(t)\phi_j(t)dt)^{2d}$ is bounded uniformly in j.

(B5) $E(\exp\{C||X||\}) < \infty$ for any constant $C > 0$, where $||X|| = \{\int_{\mathcal{S}} X^2(s)ds\}^{1/2}$.

(B6) We assume that $X(\cdot)$ is in a Donsker class.

The third set of conditions (C1), (C2) and (C3) is needed to delineate the diverging speed of the truncation number r_n as well as the decaying rates of λ_j's and β_{j0}'s.

(C1) $r_n^{4a+4}n^{-1} \to 0$.

(C2) $b > a/2 + 1$.

(C3) $r_n^{a/2+2-2b}\log(n) \to 0$.

Condition (C2) is the same as the condition in Hall and Horowitz (2007). Condition (C3) is very weak since under Condition (C2), we have $a/2+2-2b < -a/2$, and $r_n^{a/2+2-2b}\log(n) \to 0$ when r_n diverges faster than $\{\log(n)\}^{2/a}$. Moreover, Condition (C1) gives an upper bound of the diverging speed of r_n, we do not allow r_n diverging too fast in order to guarantee the consistency of our estimates.

Remark: In practice, we use the AIC or percentage of variances to select the r_n, but these methods can not guarantee that the resulting estimates satisfy the multiple conditions imposed (C1)-(C3)

We first focus on the case when p is fixed.

THEOREM 1: *Under conditions (A1)-(A5), (B1)-(B5), and (C1)-(C3) in the supplementary, we have* $||\widehat{\beta}-\beta_0|| = o_P(1)$ *and* $||\widehat{\gamma}-\gamma_0|| = O_P(\alpha_n)$, *where* $\alpha_n = r_n^{-b+1/2} + r_n^{3a/2+3/2}n^{-1/2} + r_n^{3/2-2b}\log(n) = o(1)$.

Theorem 1 establishes the consistency and convergence rate of $\widehat{\beta}(s)$ and $\widehat{\gamma}$.

Remark for Theorem 1: By conditions (C2), we have $b > 3/2$. Combining the fact that $r_n \to \infty$ when $n \to \infty$, we can show that the first term of α_n goes to zero. By condition (C1), since we have $r_n^{2a+2}n^{-1/2} \to 0$, combining the facts that $a > 0$ and $r_n \to \infty$ implies that the second term of α_n goes to zero. It follows from $a/2+2-2b > 3/2-2b$ and condition (C3) that the last term of α_n goes to zero.

THEOREM 2: *Under* H_0, *assume that conditions (A1)-(A5), (B1), (B3)-(B6), and (C1) in the supplementary hold. We have* $(T_S - r_n)/(2r_n)^{1/2} \to^d N(0,1)$ *as* $n \to \infty$.

Theorem 2 establishes the null distribution of our score statistic, which is asymptotically $\chi^2_{r_n}$. We write the results strictly as a normal approximation since $\chi^2_{r_n}$ converges to a Dirac function with point mass at ∞ when $r_n \to \infty$. However, in practice, after we select the truncation number r_n, we can use $\chi^2_{r_n}$ as the null distribution of T_S.

Finally, we consider the situation that p diverges at some polynomial rate of n, denoted by p_n. It can be shown that the asymptotic results in Theorems 1 and 2 are still valid under a set of slightly different conditions.

We have the following corollaries, which are parallel with Theorems 1 and 2.

COROLLARY 1: *Assume $p_n = o(r_n)$, under conditions (A1)-(A4), (A5*), (B1)-(B5), and (C1)-(C3) in the supplementary, we have $||\widehat{\beta} - \beta_0|| = o_P(1)$ and $||\widehat{\gamma} - \gamma_0|| = O_P(\alpha_n)$, where $\alpha_n = r_n^{-b+1/2} + r_n^{3a/2+3/2} n^{-1/2} + r_n^{3/2-2b}\log(n)$.*

As $r_n^{2a+2} n^{-1/2} \to 0$ by Condition (C1), we allow $p_n = o(n^{1/(4a+4)})$.

COROLLARY 2: *Under H_0, assume that conditions (A1)-(A4), (A5*), (B1), (B3)-(B6), and (C1) hold, and $p_n = o(r_n)$. We have $(T_S - r_n)/(2r_n)^{1/2} \to^d N(0,1)$ as $n \to \infty$.*

5. Lemmas

We need to introduce some notation before we present lemmas. Denote $||a||$ as the l_2 norm of a vector a and $||A||_F$ as the Frobenius norm of a matrix A. For a process $X(\cdot)$, denote $||X|| = \{\int_{\mathcal{S}} X^2(s)ds\}^{1/2}$. Also denote $\widehat{\Delta} = |||\widehat{K} - K||| = [\int\int_{\mathcal{S}^2}\{\widehat{K}(s_1,s_2) - K(s_1,s_2)\}^2 ds_1 ds_2]^{1/2}$, where K is the covariance function of a process $X(\cdot)$ and $\widehat{K}$ is an estimate of K. Let $\delta_j = \min_{1\leqslant k\leqslant j}(\lambda_k - \lambda_{k+1})$, the minimum spacing between the eigenvalues up to the $(j+1)$-th eigenvalue. Recall that

$$\begin{aligned} l(\eta_R) &= \sum_{i=1}^n \int_0^\tau (\sum_{k=1}^p z_{ik}\gamma_k + \sum_{j=1}^{r_n} \xi_{ijR}\beta_{jR} + e_i) dN_i(t) \\ &\quad - \int_0^\tau \log\{\sum_{i=1}^n Y_i(t)\exp(\sum_{k=1}^p z_{ik}\gamma_k + \sum_{j=1}^{r_n}\xi_{ijR}\beta_{jR} + e_i)\} d\overline{N}(t), \end{aligned}$$

$$\begin{aligned} Q(\eta_R) &= \sum_{i=1}^{n} \int_0^\tau (\sum_{k=1}^{p} z_{ik}\gamma_k + \sum_{j=1}^{r_n} \widehat{\xi}_{ijR}\beta_{jR}) dN_i(t) \\ &\quad - \int_0^\tau \log\{\sum_{i=1}^{n} Y_i(t) \exp(\sum_{k=1}^{p} z_{ik}\gamma_k + \sum_{j=1}^{r_n} \widehat{\xi}_{ijR}\beta_{jR})\} d\overline{N}(t), \end{aligned} \tag{5.1}$$

where $\beta_{r_nR} = (\beta_{1R}, \cdots, \beta_{r_nR})^{\mathrm{T}}$, $\eta_R = (\beta_{r_nR}, \gamma)$ and $e_i = \sum_{j=r_n+1}^{\infty} \xi_{ij}\beta_{j0}$. We also define

$$\alpha_n = r_n^{-b+1/2} + r_n^{3a/2+3/2} n^{-1/2} + r_n^{3/2-2b} \log(n). \tag{5.2}$$

We first present several lemmas that are used in the proof of the main theoretical results. The first lemma is the same as Lemma 3.3 of Hall and Hosseini-Nasab (2009).

LEMMA 1: *Assume that with probability 1, X is left-continuous at each point (or right-continuous at each point), and that Conditions (B3) and (B4) holds. Then, for each $C > 0$,*

$$E(\widehat{\Delta}^C) < constant * n^{-C/2}. \tag{5.3}$$

The second lemma is the same as Theorem 3 of Hall and Hosseini-Nasab (2006).

LEMMA 2: *Under Conditions (B3) and (B4), we have*

$$||\widehat{\phi}_j - \phi_j|| \leqslant 8^{1/2}\delta_j^{-1}|||\widehat{K} - K||| \quad \textit{for any } j. \tag{5.4}$$

As noted by a referee, the sign of $\widehat{\phi}_j$ actually is not estimable. So here we assume that the signs of $\widehat{\phi}_j$ and ϕ_j have been aligned.

We now restrict η_R in the set $V = \{\eta_R : ||\eta_R - \eta_{R0}|| = O(\alpha_n)\}$. Recall the definition of $\eta = (\beta_n^{\mathrm{T}}, \gamma^{\mathrm{T}})^{\mathrm{T}}$, where $\beta_n = (\beta_1, \cdots, \beta_{r_n})^{\mathrm{T}}$. We have the following lemma:

LEMMA 3: *If $||\eta_R - \eta_{R0}|| = O(\alpha_n)$, then there exists a constant C such that $||\eta|| \leqslant C$.*

For the constant C in Lemma 3, we have $V \subseteq V^*$, where $V^* = \{\eta : ||\eta|| \leqslant C\}$. Then we state the following three lemmas 4, 5, and 6, which hold uniformly for all $\eta \in V^*$, and thus hold uniformly for all $\eta_R \in V$.

LEMMA 4: *Under Conditions (B1)-(B5), we have*

$$\sup_{||\eta|| \leqslant C} \max_{1 \leqslant i \leqslant n} |\rho_i| = O_p(r_n^{a+3/2} n^{-1/2} \{\log(n)\}^{1/2} + r_n^{1/2-b} \{\log(n)\}^{1/2}),$$

where $\rho_i = \nu_i + e_i$*, in which* $\nu_i = \eta_R(w_{iR} - \widehat{w}_{iR}) = \sum_{j=1}^{r_n} (\xi_{ijR} - \widehat{\xi}_{ijR})\beta_{jR} = \sum_{j=1}^{r_n} (\xi_{ij} - \widehat{\xi}_{ij})\beta_j$ *and* $e_i = \sum_{j=r_n+1}^{\infty} \xi_{ij}\beta_{j0}$*. Moreover, under additional Conditions (C1) and (C3), we have*

$$\sup_{||\eta|| \leqslant C} \max_{1 \leqslant i \leqslant n} |\rho_i| = o_p(1).$$

LEMMA 5: *Let* $\pi_i = \eta_R^{\mathrm{T}} \widehat{w}_{iR} + \rho_i = \sum_{k=1}^{p} z_{ik}\gamma_k + \sum_{j=1}^{r_n} \xi_{ijR}\beta_{jR} + \sum_{j=r_n+1}^{\infty} \xi_{ij}\beta_{j0}$*. For any fixed positive integer* d*, under Conditions (A5) and (B5), we have* $E[\{\exp(\pi_i)\}^{2d}] = O(1)$ *uniformly for* $||\eta|| \leqslant C$.

Let $\widehat{w}_{iR} = (\widehat{\xi}_{i1R}, \ldots, \widehat{\xi}_{ir_nR}, z_{i1}, \ldots, z_{ip})^{\mathrm{T}}$. For $d = 0, 1$, and 2, we define

$$S^{(d)*}(\eta_R, t) = n^{-1} \sum_{i=1}^{n} (\widehat{w}_{iR})^{(d)} Y_i(t) \exp(\eta_R^{\mathrm{T}} \widehat{w}_{iR}),$$

where $(\widehat{w}_{iR})^{(0)} = 1$, $(\widehat{w}_{iR})^{(1)} = \widehat{w}_{iR}$, and $(\widehat{w}_{iR})^{(2)} = (\widehat{w}_{iR})^{\otimes 2}$.

LEMMA 6: *Under Conditions (A1)-(A5), (B1)-(B5), and (C1)-(C3), we have*

$$|S^{(0)}(\eta_R, t) - S^{(0)*}(\eta_R, t)| = O_p(n^{-1/2} r_n^{a+3/2} + r_n^{-b} + r_n^{1-2b} \log(n)), \quad (5.5)$$

$$||S^{(1)}(\eta_R, t) - S^{(1)*}(\eta_R, t)|| = O_p(\alpha_n), \quad (5.6)$$

$$||S^{(2)}(\eta_R, t) - S^{(2)*}(\eta_R, t)||_F = O_p(r_n \alpha_n) \quad (5.7)$$

uniformly for $t \in [0, \tau]$ *and* $||\eta|| \leqslant C$.

LEMMA 7: *Under Conditions (A1)-(A5), (B1)-(B5), and (C1)-(C3), we have*

$$\sup_{||\eta||\leqslant C} ||\partial_{\eta_R} Q(\eta_R) - \partial_{\eta_R} l(\eta_R)|| = O_p(n\alpha_n), \tag{5.8}$$

$$\sup_{||\eta||\leqslant C} ||\partial^2_{\eta_R} Q(\eta_R) - \partial^2_{\eta_R} l(\eta_R)||_F = O_p(nr_n\alpha_n), \tag{5.9}$$

where $\partial_{\eta_R} = \partial/\partial\eta_R$.

Before we introduce Lemma 8, we define

$$\begin{aligned}
Q_T(\eta) &= \sum_{i=1}^n \int_0^\tau w_i^{\mathrm{T}}\eta dN_i(t) - \int_0^\tau \log\{\sum_{i=1}^n Y_i(t)\exp(w_i^{\mathrm{T}}\eta)\} d\overline{N}(t), \\
S_T(\eta) &= \frac{\partial Q_T(\eta)}{\partial\eta} = \sum_{i=1}^n \int_0^\tau w_i dN_i(t) - \int_0^\tau \frac{\sum_{i=1}^n w_i Y_i(t)\exp(w_i^{\mathrm{T}}\eta)}{\sum_{i=1}^n Y_i(t)\exp(w_i^{\mathrm{T}}\eta)} d\overline{N}(t), \\
I_T(\eta) &= -\frac{\partial S_T(\eta)}{\partial\eta} = \int_0^\tau \left[\frac{\sum_{i=1}^n w_i^{\otimes 2} Y_i(t)\exp(w_i^{\mathrm{T}}\eta)}{\sum_{i=1}^n Y_i(t)\exp(w_i^{\mathrm{T}}\eta)} - \left\{\frac{\sum_{i=1}^n w_i^{\mathrm{T}} Y_i(t)\exp(w_i^{\mathrm{T}}\eta)}{\sum_{i=1}^n Y_i(t)\exp(w_i^{\mathrm{T}}\eta)}\right\}^{\otimes 2}\right] d\overline{N}(t).
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
Q(\eta) &= \sum_{i=1}^n \int_0^\tau \widehat{w}_i^{\mathrm{T}}\eta dN_i(t) - \int_0^\tau \log\{\sum_{i=1}^n Y_i(t)\exp(\widehat{w}_i^{\mathrm{T}}\eta)\} d\overline{N}(t), \\
S(\eta) &= \frac{\partial Q(\eta)}{\partial\eta} = \sum_{i=1}^n \int_0^\tau \widehat{w}_i dN_i(t) - \int_0^\tau \frac{\sum_{i=1}^n \widehat{w}_i Y_i(t)\exp(\widehat{w}_i^{\mathrm{T}}\eta)}{\sum_{i=1}^n Y_i(t)\exp(\widehat{w}_i^{\mathrm{T}}\eta)} d\overline{N}(t), \\
I(\eta) &= -\frac{\partial S(\eta)}{\partial\eta} = \int_0^\tau \left[\frac{\sum_{i=1}^n \widehat{w}_i^{\otimes 2} Y_i(t)\exp(\widehat{w}_i^{\mathrm{T}}\eta)}{\sum_{i=1}^n Y_i(t)\exp(\widehat{w}_i^{\mathrm{T}}\eta)} - \left\{\frac{\sum_{i=1}^n \widehat{w}_i^{\mathrm{T}} Y_i(t)\exp(\widehat{w}_i^{\mathrm{T}}\eta)}{\sum_{i=1}^n Y_i(t)\exp(\widehat{w}_i^{\mathrm{T}}\eta)}\right\}^{\otimes 2}\right] d\overline{N}(t).
\end{aligned}$$

LEMMA 8: *Under Conditions (A1)-(A5), (B1), (B3)-(B6), and (C1), we have*

$$\sup_{||\gamma||\leqslant C} ||S_T(0,\gamma) - S(0,\gamma)|| = o_p(n^{1/2}),$$

$$\sup_{||\gamma||\leqslant C} ||I_T(0,\gamma) - I(0,\gamma)||_F = o_p(n^{1/2}).$$

LEMMA 9: *Let* a_i *be a* $r_n \times 1$ *random vector with mean* $E(a_i) = 0$ *and* $E(a_i a_i^{\mathrm{T}}) = I_{r_n}$*, where* I_{r_n} *is a* $r_n \times r_n$ *identity matrix. Let* $\widetilde{a}_n = n^{-1/2}\sum_{i=1}^n a_i$*. It is assumed that* $E\{(a_i^{\mathrm{T}} a_i)^2\} = o(nr_n)$*.*

We have

$$(\widetilde{a}_n^{\mathrm{T}}\widetilde{a}_n - r_n)/(2r_n)^{1/2} \to^d N(0,1).$$

6. Proofs of Lemmas

Since Lemmas 1 and 2 are directly copied from Hall and Hosseini-Nasab (2009) and Hall and Hosseini-Nasab (2006), we refer readers to their proofs for details.

Proof of Lemma 3:

Recall that $\eta_R = (\beta_{r_nR}, \gamma)$ and $\eta = (\beta_n^{\mathrm{T}}, \gamma^{\mathrm{T}})^{\mathrm{T}}$, where $\beta_n = (\beta_1, \cdots, \beta_{r_n})^{\mathrm{T}}$ and $\beta_{r_nR} = (\beta_{1R}, \cdots, \beta_{r_nR})^{\mathrm{T}}$. Since $\lambda_j = \sum_{k=j}^{\infty}(\lambda_k - \lambda_{k+1}) \geqslant \sum_{k=j} k^{-a-1} = Cj^{-a}$, we have $\sum_{j=1}^{r_n} \lambda_j^{-1} \leqslant C\sum_{j=1}^{r_n} j^a = O(r_n^{a+1})$. By using Conditions (C1)-(C3), we have

$$\begin{aligned} ||\eta - \eta_0|| &\leqslant ||\gamma - \gamma_0|| + \sum_{j=1}^{r_n} \lambda_j^{-1}{}^{1/2} ||\beta_{r_nR} - \beta_{r_nR0}|| = O(\alpha_n) + O(r_n^{a/2+1/2}\alpha_n) \\ &= O(r_n^{2a+2}n^{-1/2} + r_n^{a/2-b+1} + r_n^{a/2+2-2b}\log(n)) = o(1). \end{aligned}$$

As $||\eta_0|| < \infty$, there exists a constant C such that $||\eta|| \leqslant C$.

Proof of Lemma 4:

Since $\max_{1\leqslant i\leqslant n}|\rho_i| \leqslant \max_{1\leqslant i\leqslant n}|\nu_i| + \max_{1\leqslant i\leqslant n}|e_i|$, it is sufficient to control the order of $\max_{1\leqslant i\leqslant n}|\nu_i|$ and $\max_{1\leqslant i\leqslant n}|e_i|$. Since $(\sum_{j=1}^{r_n}|\beta_j|^2)^{1/2} \leqslant ||\eta|| \leqslant C$, it follows from Lemma 2 that

$$\begin{aligned} |\nu_i| \leqslant ||X_i|| * \sum_{j=1}^{r_n} ||\widehat{\phi}_j - \phi_j|||\beta_j| &\leqslant ||X_i|| \sum_{j=1}^{r_n}(||\widehat{\phi}_j - \phi_j||^2)^{1/2}(\sum_{j=1}^{r_n}|\beta_j|^2)^{1/2} \\ &\leqslant C||X_i||(\sum_{j=1}^{r_n}\delta_j^{-2}|||\widehat{K} - K|||^2)^{1/2}. \end{aligned}$$

Moreover, it follows from Condition (B1) and Lemma 1 that

$$\sum_{j=1}^{r_n}\delta_j^{-2} \leqslant \sum_{j=1}^{r_n} j^{2a+2} = O(r_n^{2a+3}) \text{ and } |||\widehat{K} - K||| = O_p(n^{-1/2}).$$

By condition (B5), we have $\max_{1\leqslant i\leqslant n}||X_i|| = O_p(\{\log(n)\}^{1/2})$. Thus, we have

$$\sup_{||\eta||\leqslant C}\max_{1\leqslant i\leqslant n}|\nu_i| = O_p(r_n^{a+3/2}n^{-1/2}(\log(n))^{1/2}).$$

We consider $\max_{1\leqslant i\leqslant n}|e_i|$ as follows. Since $(\sum_{j=r_n+1}^{\infty}\beta_{j0}^2)^{1/2} \leqslant (\sum_{j=r_n+1}^{\infty}j^{-2b})^{1/2} = O(r_n^{1/2-b})$ for $b > 1/2$, we have

$$\begin{aligned}|e_i|/r_n^{1/2-b} &= |\sum_{j=r_n+1}^{\infty}\xi_{ij}\beta_{j0}|/r_n^{1/2-b} \leqslant \{(\sum_{j=r_n+1}^{\infty}\xi_{ij}^2)^{1/2}(\sum_{j=r_n+1}^{\infty}\beta_{j0}^2)^{1/2}\}/r_n^{1/2-b} \\ &\leqslant ||X_i||\{(\sum_{j=r_n+1}^{\infty}\beta_{j0}^2)^{1/2}\}/r_n^{1/2-b} = O(||X_i||).\end{aligned}$$

Therefore, it follows from Condition (B5) that $E\{\exp(|e_i|/r_n^{1/2-b})\} < \infty$, which indicates $\max_{1\leqslant i\leqslant n}|e_i|/r_n^{1/2-b} = O_p(\{\log(n)\}^{1/2})$, i.e. $\max_{1\leqslant i\leqslant n}|e_i| = O_p(r_n^{1/2-b}\{\log(n)\}^{1/2})$.

Finally, it follows from Conditions (C1) and (C3) that

$$\sup_{||\eta||\leqslant C}\max_{1\leqslant i\leqslant n}|\rho_i| = O_p(r_n^{a+3/2}n^{-1/2}\{\log(n)\}^{1/2} + r_n^{1/2-b}\{\log(n)\}^{1/2}) = o_p(1),$$

which finishes the proof.

Proof of Lemma 5:

For any fixed positive integer d, we have

$$\begin{aligned}E[\{\exp(\pi_i)\}^{2d}] &= E[\exp\{2d(\sum_{k=1}^{p}z_{ik}\gamma_k + \sum_{j=1}^{r_n}\xi_{ijR}\beta_{jR} + \sum_{j=r_n+1}^{\infty}\xi_{ij}\beta_{j0})\}] \\ &\leqslant E[\prod_{k=1}^{p}\exp(2dz_{ik}\gamma_k) \times \exp\{2d(\sum_{j=1}^{r_n}\xi_{ij}^2)^{1/2}(\sum_{j=1}^{r_n}\beta_j^2)^{1/2}\} \\ &\quad \times \exp\{2d(\sum_{j=r_n+1}^{\infty}\xi_{ij}^2)^{1/2}(\sum_{j=r_n+1}^{\infty}\beta_{j0}^2)^{1/2}\}] \\ &\leqslant \prod_{k=1}^{p}[E\{\exp(2dz_{ik}\gamma_k)\}]^{1/2} \times \exp(2d||X_i|| \times ||\eta||) \times \exp(2d||X_i|| \times ||\beta_0||)\end{aligned}$$

By condition (A5) and (B5), we have $E[\{\exp(\pi_i)\}^{2d}] = O(1)$ uniformly for $||\eta|| \leqslant C$.

Proof of Lemma 6:

For the term $|S^{(0)}(\eta_R, t) - S^{(0)*}(\eta_R, t)|$, we have

$$\begin{aligned}|S^{(0)}(\eta_R, t) - S^{(0)*}(\eta_R, t)| &= |n^{-1}\sum_{i=1}^{n} Y_i(t)\{\exp(\eta_R^{\mathrm{T}}\widehat{w}_{iR} + \rho_i) - \exp(\eta_{R0}^{\mathrm{T}}\widehat{w}_{iR})\}| \\ &= |n^{-1}\sum_{i=1}^{n} Y_i(t)\rho_i\exp(\pi_i) - (2n)^{-1}\sum_{i=1}^{n} Y_i(t)\exp(\pi_i)\exp\{(\rho_i^*)^2\}\rho_i^2|,\end{aligned}$$

where ρ_i^* is between 0 and ρ_i. Therefore, we have

$$\begin{aligned}& |S^{(0)}(\eta_R, t) - S^{(0)*}(\eta_R, t)| \\ \leqslant\; & |n^{-1}\sum_{i=1}^{n} Y_i(t)\exp(\pi_i)\nu_i| + |n^{-1}\sum_{i=1}^{n} Y_i(t)\exp(\pi_i)e_i| \\ & + C|n^{-1}\sum_{i=1}^{n} Y_i(t)\exp(\pi_i)\nu_i^2| + C|n^{-1}\sum_{i=1}^{n} Y_i(t)\exp(\pi_i)e_i^2| \\ & + Cn^{-1}\sum_{i=1}^{n} |Y_i(t)|\exp(\pi_i) \times \max_{1\leqslant i\leqslant n}[(\exp\{(\rho_i^*)^2\} - 1)\rho_i^2] \\ =\; & I_1 + I_2 + CI_3 + CI_4 + CI_5.\end{aligned}$$

We will calculate the order of each term in $\{I_k\}_{k=1}^5$ respectively.

It follows from Lemma 2 that

$$\begin{aligned}I_1 &= |n^{-1}\sum_{i=1}^{n} Y_i(t)\exp(\pi_i)\sum_{j=1}^{r_n}\int_{\mathcal{S}} X_i(s)(\widehat{\phi}_j(s) - \phi_j(s))ds\beta_j| \\ &\leqslant n^{-1}\sum_{i=1}^{n} |Y_i(t)\exp(\pi_i)| \times ||X_i||(\sum_{j=1}^{r_n}||\widehat{\phi}_j - \phi_j||^2)^{1/2}(\sum_{j=1}^{r_n}\beta_j^2)^{1/2} \\ &\leqslant n^{-1}\sum_{i=1}^{n} |Y_i(t)|\exp(\pi_i) \times ||X_i||\{8^{1/2}(\sum_{j=1}^{r_n}\delta_j^{-2})^{1/2}|||\widehat{K} - K||| \times ||\eta||\}.\end{aligned}$$

According to the definition of $Y_i(t)$, we have $|Y_i(t)|\exp(\pi_i) \times ||X_i|| \leqslant \exp(\pi_i) \times ||X_i||$ and

$$E[\{\exp(\pi_i)\}^2]E(||X_i||^2) = O(1)$$

uniformly for $t \in [0, \tau]$ and $||\eta|| \leqslant C$, which indicates that $n^{-1}\sum_{i=1}^{n}|Y_i(t)|\exp(\pi_i) \times ||X_i|| = O_p(1)$ uniformly for $t \in [0, \tau]$ and $||\eta|| \leqslant C$. Combining the fact that $\sum_{j=1}^{r_n}\delta_j^{-2} \leqslant \sum_{j=1}^{r_n}(j^{2a+2}) = O(r_n^{2a+3})$ and $|||\widehat{K} - K||| = O_p(n^{-1/2})$ by Lemma 1, we have $I_1 = O_p(n^{-1/2}r_n^{a+3/2})$ uniformly for $t \in [0, \tau]$ and $||\eta|| \leqslant C$.

We note that

$$\begin{aligned} |Y_i(t)\exp(\pi_i)\sum_{j=r_n+1}^{\infty}\xi_{ij}\beta_{j0}| &\leqslant |\exp(\pi_i)\sum_{j=r_n+1}^{\infty}\xi_{ij}\beta_{j0}|, \\ E(|\exp(\pi_i)\sum_{j=r_n+1}^{\infty}\xi_{ij}\beta_{j0}|)^2 &\leqslant E[\{\exp(\pi_i)\}^2]\sum_{j=r_n+1}^{\infty}\lambda_j\beta_{j0}^2. \end{aligned}$$

Since $\sum_{j=1}^{\infty}\lambda_j < \infty$, we have $\sum_{j=r_n+1}^{\infty}\lambda_j\beta_{j0}^2 \leqslant (\sum_{j=r_n+1}^{\infty}\lambda_j)\max_{j\geqslant r_n+1}\beta_{j0}^2 = o(r_n^{-2b})$ since $\sum_{j=r_n+1}^{\infty}\lambda_j = o(1)$ as $r_n \to \infty$. It follows from Markov's inequality that

$$I_2 \leqslant n^{-1}\sum_{i=1}^{n}|Y_i(t)\exp(\pi_i)\sum_{j=r_n+1}^{\infty}\xi_{ij}\beta_{j0}| = o_p(r_n^{-b})$$

holds uniformly for all $t \in [0, \tau]$. Since I_2 is not related to η, we know that $I_2 = o_p(r_n^{-b})$ holds uniformly for all $t \in [0, \tau]$ and $||\eta|| \leqslant C$.

Similar to I_1, we have

$$\begin{aligned} I_3 &= |n^{-1}\sum_{i=1}^{n}Y_i(t)\exp(\pi_i)\{\sum_{j=1}^{r_n}\int_{\mathcal{S}}X_i(s)(\widehat{\phi}_j(s) - \phi_j(s))ds\beta_j\}^2| \\ &\leqslant n^{-1}\sum_{i=1}^{n}|Y_i(t)| * \exp(\pi_i)||X_i||^2 * (\sum_{j=1}^{r_n}||\widehat{\phi}_j - \phi_j||^2) \times (\sum_{j=1}^{r_n}\beta_j^2) \\ &\leqslant n^{-1}\sum_{i=1}^{n}|Y_i(t)| * \exp(\pi_i)||X_i||^2\{8(\sum_{j=1}^{r_n}\delta_j^{-2})|||\widehat{K} - K|||^2 \times ||\eta||^2\} = O_p(n^{-1}r_n^{2a+3}) = o_p(n^{1/2}r_n^{a+3/2}). \end{aligned}$$

Similar to I_2, we have

$$
\begin{aligned}
E\{Y_i(t)\exp(\pi_i)(\sum_{j=r_n+1}^{\infty}\xi_{ij}\beta_{j0})^2\} &\leqslant E(\{\exp(\pi_i)\}^2)E(\sum_{j=r_n+1}^{\infty}\xi_{ij}\beta_{j0})^2 \\
&= E(\{\exp(\pi_i)\}^2)\sum_{j=r_n+1}^{\infty}\lambda_j\beta_{j0}^2 = o(r_n^{-2b}).
\end{aligned}
$$

It follows from Markov's inequality that $I_4 = o_p(r_n^{-2b}) = o_p(r_n^{-b})$ holds uniformly for all $t \in [0,\tau]$ and $||\eta|| \leqslant C$ as $b > 0$.

For I_5, notice that $\max_{1\leqslant i\leqslant n}|\rho_i| = o_p(1)$ uniformly for $||\eta|| \leqslant C$, one has $\max_{1\leqslant i\leqslant n}\{\exp(\rho_i^2) - 1\} = o_p(1)$ uniformly for $||\eta|| \leqslant C$. Meanwhile, we have $n^{-1}\sum_{i=1}^n |Y_i(t)|\exp(\pi_i) = O_p(1)$. Thus, one has $I_5 = o_p(\max_{1\leqslant i\leqslant n}|\rho_i|^2) = o_p(r_n^{2a+3}n^{-1}\log(n) + r_n^{1-2b}\log(n))$ uniformly for all $t \in [0,\tau]$ and $||\eta|| \leqslant C$. As $\sum_{j=1}^{\infty}\lambda_j < \infty$, one has $a > 1$, which indicates that $r_n^{2a+3}n^{-1}\log(n) = o(r_n^{a+3/2}n^{-1/2})$ by condition (C1). Combining the results of all $\{I_k, 1 \leqslant k \leqslant 5\}$ leads to (5.5).

We prove (5.6) as follows. We note that $||S^{(1)}(\eta_R,t) - S^{(1)*}(\eta_R,t)||$ is bounded above by

$$
\begin{aligned}
||S^{(1)}(\eta_R,t) - S^{(1)*}(\eta_R,t)|| &\leqslant ||n^{-1}\sum_{i=1}^n w_{iR}Y_i(t)\{\exp(\eta_R^{\mathrm{T}}\widehat{w}_{iR}+\rho_i) - \exp(\eta_R^{\mathrm{T}}\widehat{w}_{iR})\}|| \\
&\quad +||n^{-1}\sum_{i=1}^n(\widehat{w}_{iR}-w_{iR})Y_i(t)\{\exp(\eta_R^{\mathrm{T}}\widehat{w}_{iR}+\rho_i)\}|| \\
&\quad +||n^{-1}\sum_{i=1}^n(\widehat{w}_{iR}-w_{iR})Y_i(t)\{\exp(\eta_R^{\mathrm{T}}\widehat{w}_{iR}+\rho_i) - \exp(\eta_R^{\mathrm{T}}\widehat{w}_{iR})\}|| \\
&= I_6 + I_7 + I_8.
\end{aligned}
$$

For I_6, we have

$$
\begin{aligned}
I_6 &= ||n^{-1}\sum_{i=1}^n w_{iR}\exp(\pi_i)Y_i(t)\rho_i - (2n)^{-1}\sum_{i=1}^n w_{iR}\exp(\pi_i)Y_i(t)\exp\{(\rho_i^*)^2\}\rho_i^2|| \\
&\leqslant ||n^{-1}\sum_{i=1}^n w_{iR}\exp(\pi_i)Y_i(t)\nu_i|| + ||n^{-1}\sum_{i=1}^n w_{iR}\exp(\pi_i)Y_i(t)e_i|| \\
&\quad +C||n^{-1}\sum_{i=1}^n w_{iR}\exp(\pi_i)Y_i(t)\nu_i^2|| + C||n^{-1}\sum_{i=1}^n w_{iR}\exp(\pi_i)Y_i(t)e_i^2|| +
\end{aligned}
$$

$$
\begin{aligned}
& Cn^{-1}\sum_{i=1}^{n}||w_{iR}\exp(\pi_i)Y_i(t)||\max_{i=1,\ldots,n}\rho_i^2[\exp\{(\rho_i^*)^2\}-1] \\
= \; & I_{61}+I_{62}+CI_{63}+C\times I_{64}+C\times I_{65}.
\end{aligned}
$$

By using conditions (B4) and (A5), we have $E(w_{ikR}^4)=O(1)$ uniformly for $1\leqslant k\leqslant r_n$. It follows from Lemma 5 that we have $E\{\exp(C\pi_i)\}=O(1)$. Thus, we get

$$
\begin{aligned}
E(I_{61}^2) &= n^{-2}E[\sum_{k=1}^{r_n}\{\sum_{i=1}^{n}w_{ikR}\exp(\pi_i)Y_i(t)\nu_i\}^2]\leqslant n^{-1}E[\sum_{k=1}^{r_n}\sum_{i=1}^{n}\{w_{ikR}\exp(\pi_i)Y_i(t)\nu_i\}^2] \\
&\leqslant n^{-1}\sum_{k=1}^{r_n}\sum_{i=1}^{n}\{E(w_{ikR}^4)\}^{1/2}[E\{Y_i(t))^4\}]^{1/2}\{E(\nu_i^4)\}^{1/2}[E\{\exp(4\pi_i)\}]^{1/2} \\
&\leqslant Cn^{-1}nr_n[E\{\sum_{j=1}^{r_n}\int_{\mathcal{S}}X_i(s)(\widehat{\phi}_j(s)-\phi_j(s))ds\beta_j\}^4]^{1/2} \\
&\leqslant Cr_n(\sum_{j=1}^{r_n}\delta_j^{-2})||\eta||^2\times\{E(||X_i||^4)\}^{1/2}\times(E|||\widehat{K}-K|||^4)^{1/2} \\
&= O(n^{-1}r_n^{2a+4}),
\end{aligned}
$$

which yields that $I_{61}=O_p(n^{-1/2}r_n^{a+2})$ holds uniformly for all $t\in[0,\tau]$ and $||\eta||\leqslant C$. By using similar techniques, we can show that

$$
\begin{aligned}
& I_{62}=o_p(r_n^{-b+1/2}),\quad I_{63}=o_p(n^{-1}r_n^{2a+7/2})=o_p(n^{-1/2}r_n^{a+2}), \\
& I_{64}=o_p(r_n^{-2b+1/2})=o_p(r_n^{-b+1/2}),\quad I_{65}=r_nO_p(\max_{i=1,\ldots,n}\rho_i^4)
\end{aligned}
$$

hold uniformly for $t\in[0,\tau]$ and $||\eta||\leqslant C$. By combining the above results for $\{I_{6k},1\leqslant k\leqslant 5\}$, by condition (C1)-(C3), we have $I_6=O_p(n^{-1/2}r_n^{a+2}+r_n^{-b+1/2}+r_n^{3/2-2b}\log(n))$.

Notice that $\lambda_j=\sum_{k=j}^{\infty}(\lambda_k-\lambda_{k+1})\geqslant C\sum_{k=j}^{\infty}Ck^{-a-1}=O(j^{-a})$ by Condition (B1), one has $\sum_{j=1}^{r_n}\delta_j^{-2}\lambda_j^{-1}\leqslant C\sum_{j=1}^{r_n}j^{3a+2}=O(r_n^{3a+3})$. Thus, for I_7, we have

$$
I_7 \;=\; ||n^{-1}\sum_{i=1}^{n}(\widehat{w}_{iR}-w_{iR})Y_i(t)\exp(\eta_{R0}^{\mathrm{T}}\widehat{w}_{iR}+\rho_i)||
$$

$$
\begin{aligned}
&= n^{-1}[\sum_{j=1}^{r_n}\{\sum_{i=1}^{n}(\widehat{\xi}_{ij}-\xi_{ij})\lambda_j^{-1/2}Y_i(t)\exp(\pi_i)\}^2]^{1/2} \\
&\leqslant n^{-1}[\sum_{j=1}^{r_n}\{||X_i||\times||\widehat{\phi}_j-\phi_j||\lambda_j^{-1/2}|Y_i(t)|\exp(\pi_i)\}^2]^{1/2} \\
&\leqslant n^{-1}\{\sum_{j=1}^{r_n}(\delta_j^{-1}\lambda_j^{-1/2})^2\}^{1/2}|||\widehat{K}-K|||\sum_{i=1}^{n}||X_i||\exp(\pi_i) \\
&= O_p(r_n^{3a/2+3/2}n^{-1/2})
\end{aligned}
$$

uniformly for $t\in[0,\tau]$ and $||\eta||\leqslant C$.

For I_8, we have

$$
\begin{aligned}
I_8 &= ||n^{-1}\sum_{i=1}^{n}(\widehat{w}_{iR}-w_{iR})Y_i(t)\{\exp(\eta_{R0}^{\mathrm{T}}\widehat{w}_{iR}+\rho_i)-\exp(\eta_{R0}^{\mathrm{T}}\widehat{w}_{iR})\}|| \\
&= ||n^{-1}\sum_{i=1}^{n}(\widehat{w}_{iR}-w_{iR})\exp(\pi_i)Y_i(t)\rho_i-(2n)^{-1}\sum_{i=1}^{n}(\widehat{w}_{iR}-w_{iR})\exp(\pi_i)Y_i(t)\exp\{(\rho_i^*)^2\}\rho_i^2|| \\
&\leqslant ||n^{-1}\sum_{i=1}^{n}(\widehat{w}_{iR}-w_{iR})\exp(\pi_i)Y_i(t)\nu_i||+||n^{-1}\sum_{i=1}^{n}(\widehat{w}_{iR}-w_{iR})\exp(\pi_i)Y_i(t)e_i|| \\
&\quad +C||n^{-1}\sum_{i=1}^{n}(\widehat{w}_{iR}-w_{iR})\exp(\pi_i)Y_i(t)\nu_i^2||+C||n^{-1}\sum_{i=1}^{n}(\widehat{w}_{iR}-w_{iR})\exp(\pi_i)Y_i(t)e_i^2|| \\
&\quad +Cn^{-1}\sum_{i=1}^{n}||(\widehat{w}_{iR}-w_{iR})\exp(\pi_i)Y_i(t)||\max_{i=1,\dots,n}\rho_i^2[\exp\{(\rho_i^*)^2\}-1] \\
&= I_{81}+I_{82}+CI_{83}+CI_{84}+CI_{85}.
\end{aligned}
$$

For I_{81}, we have

$$
\begin{aligned}
I_{81} &= n^{-1}[\sum_{k=1}^{r_n}\{\sum_{i=1}^{n}(\widehat{\xi}_{ik}-\xi_{ik})\lambda_k^{-1/2}\exp(\pi_i)Y_i(t)\nu_i\}^2]^{1/2} \\
&\leqslant n^{-1}[\sum_{k=1}^{r_n}\{\sum_{i=1}^{n}||X_i||*||\widehat{\phi}_k-\phi_k||\lambda_k^{-1/2}|Y_i(t)|*||X_i||\sum_{j=1}^{r_n}||\widehat{\phi}_j-\phi_j||*|\beta_j|\}^2]^{1/2} \\
&\leqslant n^{-1}C[\sum_{k=1}^{r_n}\{\sum_{i=1}^{n}||X_i||^2*\delta_k^{-1}|||\widehat{K}-K|||\lambda_k^{-1/2}\sum_{j=1}^{r_n}\delta_j^{-1}|||\widehat{K}-K|||*|\beta_j|\}^2]^{1/2} \\
&= n^{-1}C\{\sum_{k=1}^{r_n}(\delta_k^{-1}\lambda_k^{-1/2})^2\}^{1/2}\sum_{i=1}^{n}||X_i||^2|||\widehat{K}-K|||^2(\sum_{j=1}^{r_n}\delta_j^{-2})^{1/2}||\eta||
\end{aligned}
$$

$$= O_p(r_n^{5a/2+3} n^{-1})$$

uniformly for $t \in [0, \tau]$ and $||\eta|| \leqslant C$. Using similar techniques, we can show that

$$I_{82} = o_p(r_n^{-b+3a/2+3/2} n^{-1/2}), \quad I_{83} = O_p(r_n^{7a/2+9/2} n^{-3/2}) = o_p(r_n^{5a/2+3} n^{-1}),$$

$$I_{84} = o_p(r_n^{-2b+3a/2+3/2} n^{-1}) = o_p(r_n^{-b+3a/2+3/2} n^{-1/2})$$

$$I_{85} = r_n o_p(\max_{i=1,\ldots,n} \rho_i^4)$$

hold uniformly for all $t \in [0, \tau]$ and $||\eta|| \leqslant C$.

By combing the results of $\{I_{8k}, 1 \leqslant k \leqslant 5\}$, we have $I_8 = o_p(n^{-1/2} r_n^{a+2} + r_n^{-b+1/2} + r_n^{3/2-2b} \log(n))$ by conditions (C1)-(C3). As $\sum_{j=1}^{\infty} \lambda_j < \infty$, we have $a > 1$, which indicates $n^{-1/2} r_n^{a+2} = o(n^{-1/2} r_n^{3a/2+3/2})$. Finally, combining the above results of I_6, I_7 and I_8 leads to (5.6).

We prove (5.7) as follows. With some calculations, we have

$$\begin{aligned}
& ||S^{(2)}(\eta_R, t) - S^{(2)*}(\eta_R, t)||_F \\
\leqslant\ & ||n^{-1} \sum_{i=1}^{n} w_{iR}^{\otimes 2} Y_i(t) \{\exp(\eta_R^{\mathrm{T}} \widehat{w}_{iR} + \rho_i) - \exp(\eta_R^{\mathrm{T}} \widehat{w}_{iR})\}||_F \\
& + ||n^{-1} \sum_{i=1}^{n} \{(\widehat{w}_{iR})^{\otimes 2} - (w_{iR})^{\otimes 2}\} Y_i(t) \{\exp(\eta_R^{\mathrm{T}} \widehat{w}_{iR} + \rho_i)\}||_F \\
& + ||n^{-1} \sum_{i=1}^{n} \{(\widehat{w}_{iR})^{\otimes 2} - (w_{iR})^{\otimes 2}\} Y_i(t) \{\exp(\eta_R^{\mathrm{T}} \widehat{w}_{iR} + \rho_i) - \exp(\eta_R^{\mathrm{T}} \widehat{w}_{iR})\}||_F \\
=\ & I_9 + I_{10} + I_{11}.
\end{aligned}$$

For I_9, we have

$$\begin{aligned}
I_9 &= ||n^{-1} \sum_{i=1}^{n} w_{iR}^{\otimes 2} \exp(\pi_i) Y_i(t) \rho_i - (2n)^{-1} \sum_{i=1}^{n} w_{iR}^{\otimes 2} \exp(\pi_i) Y_i(t) \exp\{(\rho_i^*)^2\} \rho_i^2||_F \\
&\leqslant ||n^{-1} \sum_{i=1}^{n} w_{iR}^{\otimes 2} \exp(\pi_i) Y_i(t) \nu_i||_F + ||n^{-1} \sum_{i=1}^{n} w_{iR}^{\otimes 2} \exp(\pi_i) Y_i(t) e_i||_F
\end{aligned}$$

$$
\begin{aligned}
& +C||n^{-1}\sum_{i=1}^{n} w_{iR}^{\otimes 2}\exp(\pi_i)Y_i(t)\nu_i^2||_F + C||n^{-1}\sum_{i=1}^{n} w_{iR}^{\otimes 2}\exp(\pi_i)Y_i(t)e_i^2||_F \\
& +Cn^{-1}\sum_{i=1}^{n}||w_{iR}^{\otimes 2}\exp(\pi_i)Y_i(t)|| \max_{i=1,\ldots,n}\rho_i^2[\exp\{(\rho_i^*)^2\}-1] \\
= \; & I_{91} + I_{92} + C\times I_{93} + C\times I_{94} + C\times I_{95}.
\end{aligned}
$$

Similar as I_{61}, we get

$$
\begin{aligned}
E(I_{91}^2) &= n^{-2}E[\sum_{k=1}^{r_n}\sum_{k'=1}^{r_n}\{\sum_{i=1}^{n} w_{ikR}w_{ik'R}\exp(\pi_i)Y_i(t)\nu_i\}^2] \\
&\leqslant n^{-1}E[\sum_{k=1}^{r_n}\sum_{k'=1}^{r_n}\sum_{i=1}^{n}\{w_{ikR}w_{ik'R}\exp(\pi_i)Y_i(t)\nu_i\}^2] \\
&\leqslant n^{-1}\sum_{k=1}^{r_n}\sum_{k'=1}^{r_n}\sum_{i=1}^{n}\{E(w_{ikR}^4)\}^{1/2}\{E(w_{ik'R}^4)\}^{1/2}[E\{Y_i(t))^4\}]^{1/2}[E\{\nu_i)^4\}]^{1/2}[E\{\exp(4\pi_i)\}]^{1/2}) \\
&\leqslant Cn^{-1}nr_n[E\{\sum_{j=1}^{r_n}\int_{\mathcal{S}} X_i(s)(\widehat{\phi}_j(s)-\phi_j(s))ds\beta_j\}^4]^{1/2} \\
&\leqslant Cr_n^2(\sum_{j=1}^{r_n}\delta_j^{-2})||\eta||^2\times\{E(||X_i|^4)\}^{1/2}\times(E||\widehat{K}-K|||^4)^{1/2} = O(n^{-1}r_n^{2a+5}),
\end{aligned}
$$

which yields that $I_{91} = O_p(n^{-1/2}r_n^{a+5/2})$ uniformly for $t\in[0,\tau]$ and $||\eta||\leqslant C$. Using similar techniques as we control , we obtain

$$
I_{92} = o_p(r_n^{-b+1}), \quad I_{93} = o_p(n^{-1}r_n^{2a+4}) = o_p(n^{-1/2}r_n^{a+5/2}),
$$

$$
I_{94} = o_p(r_n^{-2b+1}) = o_p(r_n^{-b+1}), \quad I_{95} = o_p(r_n^{2a+4}n^{-1}\log(n) + r_n^{3-2b}\log(n))
$$

hold uniformly for all $t\in[0,\tau]$ and $||\eta||\leqslant C$.

For I_{10}, we can write

$$
\begin{aligned}
I_{10} &= ||n^{-1}\sum_{i=1}^{n}\{(\widehat{w}_{iR})^{\otimes 2}-(w_{iR})^{\otimes 2}\}Y_i(t)\{\exp(\eta_{R0}^{\mathrm{T}}\widehat{w}_{iR}+\rho_i)\}||_F \\
&= n^{-1}[\sum_{j=1}^{r_n}\sum_{k=1}^{r_n}\{\sum_{i=1}^{n}(\widehat{\xi}_{ij}\widehat{\xi}_{ik}-\xi_{ij}\xi_{ik})\lambda_j^{-1/2}\lambda_k^{-1/2}Y_i(t)\exp(\pi_i)\}^2]^{1/2}
\end{aligned}
$$

$$
\begin{aligned}
&\leqslant n^{-1}(\sum_{j=1}^{r_n}\sum_{k=1}^{r_n}[\sum_{i=1}^{n}\{(\widehat{\xi}_{ij}-\xi_{ij})\xi_{ik}+(\widehat{\xi}_{ik}-\xi_{ik})\xi_{ij}+(\widehat{\xi}_{ij}-\xi_{ij})(\widehat{\xi}_{ik}-\xi_{ik})\}\\
&\quad \lambda_j^{-1/2}\lambda_k^{-1/2}Y_i(t)\exp(\pi_i)]^2)^{1/2}\\
&\leqslant n^{-1}[3\sum_{j=1}^{r_n}\sum_{k=1}^{r_n}\{\sum_{i=1}^{n}(\widehat{\xi}_{ij}-\xi_{ij})\xi_{ik}\lambda_j^{-1/2}\lambda_k^{-1/2}Y_i(t)\exp(\pi_i)\}^2\\
&\quad +3\sum_{j=1}^{r_n}\sum_{k=1}^{r_n}\{\sum_{i=1}^{n}(\widehat{\xi}_{ik}-\xi_{ik})\xi_{ij}\lambda_j^{-1/2}\lambda_k^{-1/2}Y_i(t)\exp(\pi_i)\}^2\\
&\quad +3\sum_{j=1}^{r_n}\sum_{k=1}^{r_n}\{\sum_{i=1}^{n}(\widehat{\xi}_{ij}-\xi_{ij})(\widehat{\xi}_{ik}-\xi_{ik})\lambda_j^{-1/2}\lambda_k^{-1/2}Y_i(t)\exp(\pi_i)\}^2]^{1/2}\\
&= n^{-1}(3I_{10,1}+3I_{10,2}+3I_{10,3})^{1/2}.
\end{aligned}
$$

For $I_{10,1}$, we can write

$$
\begin{aligned}
I_{10,1} &\leqslant \sum_{j=1}^{r_n}\sum_{k=1}^{r_n}\{\sum_{i=1}^{n}||X_i||*||\widehat{\phi}_j-\phi_j||\delta_j^{-1}\xi_{ik}\lambda_j^{-1/2}\lambda_k^{-1/2}Y_i(t)\exp(\pi_i)\}^2\\
&= \sum_{j=1}^{r_n}(\delta_j^{-1}\lambda_j^{-1/2})^2|||\widehat{K}-K|||^2\sum_{k=1}^{r_n}\{\sum_{i=1}^{n}||X_i||\xi_{ik}\lambda_k^{-1/2}Y_i(t)\exp(\pi_i)\}^2.
\end{aligned}
$$

It is easy to see that $E[\sum_{k=1}^{r_n}\{\sum_{i=1}^{n}||X_i||\xi_{ik}\lambda_k^{-1/2}Y_i(t)\exp(\int_{\mathcal{S}}X_i(s)\beta_0(s)ds+\sum_{k=1}^{p}Z_{ik}\gamma_k)\}^2]=O(n^2r_n)$, we can easily see $I_{10,1}=O_p(r_n^{3a+4}n)$ uniformly for $t\in[0,\tau]$ and $||\eta||\leqslant C$. Similarly, $I_{10,2}=O_p(r_n^{3a+4}n)$ and $I_{10,3}=O_p(r_n^{6a+6})$ uniformly for $t\in[0,\tau]$ and $||\eta||\leqslant C$. Thus, one has $I_{10}=O_p(r_n^{3a/2+2}n^{-1/2})$ uniformly for $t\in[0,\tau]$ and $||\eta||\leqslant C$ by condition (C1).

For I_{11}, one has

$$
\begin{aligned}
I_{11} &= ||n^{-1}\sum_{i=1}^{n}\{(\widehat{w}_{iR})^{\otimes 2}-(w_{iR})^{\otimes 2}\}Y_i(t)\{\exp(\eta_{R0}^{\mathrm{T}}\widehat{w}_{iR}+\rho_i)-\exp(\eta_{R0}^{\mathrm{T}}\widehat{w}_{iR})\}||_F\\
&= ||n^{-1}\sum_{i=1}^{n}\{(\widehat{w}_{iR})^{\otimes 2}-(w_{iR})^{\otimes 2}\}\exp(\pi_i)Y_i(t)\rho_i\\
&\quad -(2n)^{-1}\sum_{i=1}^{n}(\widehat{w}_{iR}-w_{iR})\exp(\pi_i)Y_i(t)\exp\{(\rho_i^*)^2\}\rho_i^2||_F\\
&\leqslant ||n^{-1}\sum_{i=1}^{n}\{(\widehat{w}_{iR})^{\otimes 2}-(w_{iR})^{\otimes 2}\}\exp(\pi_i)Y_i(t)\nu_i||_F
\end{aligned}
$$

$$
\begin{aligned}
&+||n^{-1}\sum_{i=1}^{n}\{(\widehat{w}_{iR})^{\otimes 2}-(w_{iR})^{\otimes 2}\}\exp(\pi_i)Y_i(t)e_i||_F \\
&+C||n^{-1}\sum_{i=1}^{n}\{(\widehat{w}_{iR})^{\otimes 2}-(w_{iR})^{\otimes 2}\}\exp(\pi_i)Y_i(t)\nu_i^2||_F \\
&+C||n^{-1}\sum_{i=1}^{n}\{(\widehat{w}_{iR})^{\otimes 2}-(w_{iR})^{\otimes 2}\}\exp(\pi_i)Y_i(t)e_i^2||_F \\
&+C(n^{-1}\sum_{i=1}^{n}||(\widehat{w}_{iR}^{\otimes 2}-w_{iR}^{\otimes 2})\exp(\pi_i)Y_i(t)||\max_{i=1,\ldots,n}\rho_i^2|\exp\{(\rho_i^*)^2\}-1|) \\
= \quad & I_{11,1}+I_{11,2}+CI_{11,3}+CI_{11,4}+CI_{11,5}
\end{aligned}
$$

By some simple algebra as the previous derivations, we have

$$
\begin{aligned}
&I_{11,1}=O_p(r_n^{5a/2+7/2}n^{-1}), \quad I_{11,2}=o_p(r_n^{-b+3a/2+2}n^{-1/2}), \\
&I_{11,3}=O_p(r_n^{7a/2+5}n^{-3/2})=o_p(r_n^{5a/2+7/2}n^{-1}), \\
&I_{11,4}=o_p(r_n^{-2b+3a/2+2}n^{-1})=o_p(Ir_n^{-b+3a/2+2}n^{-1/2}) \\
&I_{11,5}=o_p(r_n^{7a/2+6}n^{-3/2}\log(n)+r_n^{4-2b+3a/2}\log(n)n^{-1/2}).
\end{aligned}
$$

uniformly for $t\in[0,\tau]$ and $||\eta||\leqslant C$.

By combining all the results and Conditions (C1)-(C3), we have

$$
||S^{(2)}(\eta_R,t)-S^{(2)*}(\eta_R,t)||_F=O_p(r_n\alpha_n).
$$

Proof of Lemma 7:

With some calculations, we have

$$
||\partial_{\eta_R}Q(\eta_{R0})-\partial_{\eta_R}l(\eta_{R0})||\leqslant||\sum_{i=1}^{n}\int_0^{\tau}(\widehat{w}_{iR}-w_{iR})dN_i(t)||+||\int_0^{\tau}(\frac{S^{(1)*}}{S^{(0)*}}-\frac{S^{(1)}}{S^{(0)}})d\overline{N}(t)||.
$$

Since $E(\int_0^\tau ||X_i|| dN_i(t)) < \infty$, we have

$$\begin{aligned} ||\sum_{i=1}^n \int_0^\tau (\widehat{w}_{iR} - w_{iR}) dN_i(t)|| &= (\sum_{j=1}^{r_n} [\sum_{i=1}^n \int_0^\tau \{(\widehat{\xi}_{ij} - \xi_{ij})\lambda_j^{-1/2}\} dN_i(t)]^2)^{1/2} \\ &\leqslant \{(\sum_{i=1}^n \int_0^\tau ||X_i|| dN_i(t))^2 \sum_{j=1}^{r_n} (||\widehat{\phi}_j - \phi_j||\lambda_j^{-1/2})^2\}^{1/2} \\ &\leqslant \{Cn^2 |||\widehat{K} - K|||^2 \sum_{j=1}^{r_n} (\delta_j^{-1}\lambda_j^{-1/2})^2\}^{1/2} \\ &= O_p(n^{1/2} r_n^{3a/2+3/2}) = O_p(n\alpha_n) \end{aligned}$$

uniformly for $t \in [0, \tau]$ and $||\eta|| \leqslant C$.

Notice $\overline{N}(t) = \sum_{i=1}^n N_i(t)$, it follows from Lemma 5 that

$$||\int_0^\tau \left(\frac{S^{(1)*}}{S^{(0)*}} - \frac{S^{(1)}}{S^{(0)}} \right) d\overline{N}(t)|| = O_p(n\alpha_n)$$

uniformly for $t \in [0, \tau]$ and $||\eta|| \leqslant C$. Combing the above results leads to (5.8).

For the second part of the lemma, $||\partial^2_{\eta_R} Q(\eta_R) - \partial^2_{\eta_R} l(\eta_R)||_F$ can be written as

$$||\int_0^\tau \left\{ \frac{S^{(2)*}}{S^{(0)*}} - \left(\frac{S^{(1)*}}{S^{(0)*}} \right)^{\otimes 2} \right\} d\overline{N}(t) - \int_0^\tau \left\{ \frac{S^{(2)}}{S^{(0)}} - \left(\frac{S^{(1)}}{S^{(0)}} \right)^{\otimes 2} \right\} d\overline{N}(t)||_F.$$

It follows from Lemma 5 that $||\partial^2_{\eta_R} Q(\eta_R) - \partial^2_{\eta_R} l(\eta_R)||_F = O_p(nr_n\alpha_n)$ uniformly for $t \in [0, \tau]$ and $||\eta|| \leqslant C$.

Proof of Lemma 8:

Since

$$S_T(0,\gamma) - S(0,\gamma) = \sum_{i=1}^n \int_0^\tau (\widehat{w}_i - w_i) dN_i(t) - \int_0^\tau \frac{\sum_{i=1}^n (\widehat{w}_i - w_i) Y_i(t) \exp(Z_i^{\mathrm{T}}\gamma)}{\sum_{i=1}^n Y_i(t) \exp(Z_i^{\mathrm{T}}\gamma)} d\overline{N}(t),$$

we have

$$||S_T(0,\gamma) - S(0,\gamma)|| \leqslant ||\sum_{i=1}^n \int_0^\tau (\widehat{w}_i - w_i) dN_i(t)|| + ||\int_0^\tau \frac{n^{-1}\sum_{i=1}^n (\widehat{w}_i - w_i) Y_i(t) \exp(Z_i^{\mathrm{T}}\gamma)}{n^{-1}\sum_{i=1}^n Y_i(t) \exp(Z_i^{\mathrm{T}}\gamma)} d\overline{N}(t)||.$$

For the first term, we have

$$\begin{aligned}||\sum_{i=1}^{n}\int_0^{\tau}(\widehat{w}_i-w_i)dN_i(t)||^2 &= \sum_{j=1}^{r_n}[\sum_{i=1}^{n}\int_0^{\tau}\{\int_{\mathcal{S}}X_i(s)(\widehat{\phi}_j(s)-\phi_j(s))ds\}dN_i(t)]^2\\ &= \sum_{j=1}^{r_n}[\int_{\mathcal{S}}\sum_{i=1}^{n}(\int_0^{\tau}dN_i(t)X_i(s))(\widehat{\phi}_j(s)-\phi_j(s))ds]^2\\ &\leqslant \sum_{j=1}^{r_n}[\int_{\mathcal{S}}W_n(s)(\widehat{\phi}_j(s)-\phi_j(s))ds]^2\leqslant||W_n||^2\times\sum_{j=1}^{r_n}||\widehat{\phi}_j-\phi_j||^2,\end{aligned}$$

where $W_n(s)=\sum_{i=1}^{n}(\int_0^{\tau}dN_i(t)X_i(s))$. Since $X_i(s)$ is a centered process, by Condition (B6), one know that $W_n(s)$ converges weakly to a centered gaussian process with some covariance function $\Sigma(s,t)$. Thus, one has $\sup_{s\in\mathcal{S}}|W_n(s)|=O_p(n^{1/2})$ and $||W_n||=O_p(n^{1/2})$. For the second term, by Lemma 1 and 2, one has $\sum_{j=1}^{r_n}||\widehat{\phi}_j-\phi_j||^2\leqslant\sum_{j=1}^{r_n}\delta_j^{-2}|||\widehat{K}-K|||^2=O_p(r_n^{2a+3}n^{-1})$. Thus, one has $||\sum_{i=1}^{n}\int_0^{\tau}(\widehat{w}_i-w_i)dN_i(t)||=O_p(r_n^{a+3/2})=o_p(n^{1/2})$.

For the second term, we have

$$\begin{aligned}&||n^{-1}\sum_{i=1}^{n}(\widehat{w}_i-w_i)Y_i(t)\exp(Z_i^{\mathrm{T}}\gamma)||^2\\ =\;&||n^{-1}\sum_{i=1}^{n}\int_{\mathcal{S}}X_i(s)(\widehat{\phi}_j(s)-\phi_j(s))dsY_i(t)\exp(Z_i^{\mathrm{T}}\gamma)||^2\\ =\;&n^{-2}\sum_{j=1}^{r_n}[\sum_{i=1}^{n}\int_{\mathcal{S}}X_i(s)Y_i(t)\exp(Z_i^{\mathrm{T}}\gamma)(\widehat{\phi}_j(s)-\phi_j(s))ds]^2\\ \leqslant\;&n^{-2}\{\sum_{i=1}^{n}\int_{\mathcal{S}}|X_i(s)|^2\exp(Z_i^{\mathrm{T}}\gamma)]^2ds\}\sum_{j=1}^{r_n}||\widehat{\phi}_j-\phi_j||^2=O_p(n^{-1}r_n^{a+3/2}).\end{aligned}$$

Similarly, we have $\sup_{t\in[0,\tau]}|n^{-1}\sum_{i=1}^{n}Y_i(t)\exp(Z_i^{\mathrm{T}}\gamma)|=O_p(1)$. Finally, we have

$$||\int_0^{\tau}\frac{n^{-1}\sum_{i=1}^{n}(\widehat{w}_i-w_i)Y_i(t)\exp(Z_i^{\mathrm{T}}\gamma)}{n^{-1}\sum_{i=1}^{n}Y_i(t)\exp(Z_i^{\mathrm{T}}\gamma)}d\overline{N}(t)||=O_p(r_n^{a+3/2})=o_p(n^{1/2}).$$

Thus, $\sup_{||\gamma||\leqslant C}||S_T(0,\gamma)-S(0,\gamma)||=o_p(n^{1/2})$. Similarly, we can show that $\sup_{||\gamma||\leqslant C}||I_T(0,\gamma)-I(0,\gamma)||_F=o_p(n^{1/2})$, which completes the proof.

Proof of Lemma 9:

This Lemma is just a special case of Corollary 1 of Peng and Schick (2014) by setting $V_n = I_{r_n}$ and $\mu_n = 0$.

7. Proofs of Theorems

Proof of Theorem 1.

Recall that $\alpha_n = O_p(r_n^{-b+1/2} + r_n^{3a/2+3/2} n^{-1/2} + r_n^{3/2-2b} \log(n))$, it is sufficient to show that for any given $\epsilon > 0$, there exists a large constant C such that

$$P\{\sup_{||u||=C} Q(\eta_{R0} + \alpha_n u) < Q(\eta_{R0})\} \geqslant 1 - \epsilon. \tag{7.1}$$

This implies that there exists a local maximizer such that $||\widehat{\eta}_R - \eta_{R0}|| = O_p(\alpha_n)$.

It follows from Taylor's expansion that

$$\begin{aligned} & Q(\eta_{R0} + \alpha_n u) - Q(\eta_{R0}) \\ = \; & \alpha_n u^{\mathrm{T}} \partial_{\eta_R} Q(\eta_{R0}) + \frac{\alpha_n^2}{2} u^{\mathrm{T}} \partial^2_{\eta_R} Q(\eta_R^*) u \\ = \; & \alpha_n u^{\mathrm{T}} \partial_{\eta_R} l(\eta_{R0}) + \alpha_n u^{\mathrm{T}} \{\partial_{\eta_R} Q(\eta_{R0}) - \partial_{\eta_R} l(\eta_{R0})\} + \frac{\alpha_n^2}{2} u^{\mathrm{T}} \partial^2_{\eta_R} Q(\eta_R^*) u \\ = \; & J_1 + J_2 + J_3, \end{aligned}$$

where $\partial^2_{\eta_R} = \frac{\partial^2}{\partial \eta_R \partial \eta_R^{\mathrm{T}}}$ and η_R^* lies between η_{R0} and $\eta_{R0} + \alpha_n u$.

It follows from (C2) that $||\partial_{\eta_R} l(\eta_R)|| = O_p(r_n^{1/2} n^{1/2}) = o_p(n\alpha_n)$ and

$$|J_1| \leqslant \alpha_n ||\partial_{\eta_R} l(\eta_{R0})|| \times ||u|| = o_p(n\alpha_n^2)||u||.$$

It follows from Lemma 7 that

$$|J_2| \leqslant \alpha_n ||\partial_{\eta_R} Q(\eta_{R0}) - \partial_{\eta_R} l(\eta_{R0})|| * ||u|| = O_p(n\alpha_n^2 ||u||).$$

For J_3, we have

$$\partial^2_{\eta_R} Q(\eta^*_R) \quad = \quad \{\partial^2_{\eta_R} Q(\eta^*_R) - \partial^2_{\eta_R} l(\eta^*_R)\} + \{\partial^2_{\eta_R} l(\eta^*_R) - \partial^2_{\eta_R} l(\eta_{R0})\}$$

For the first term, as $||\eta^*_R - \eta_{R0}|| = O(\alpha_n)$, by Lemma 3, Lemma 7 and conditions (C1)-(C3), $||\partial^2_{\eta_R} Q(\eta^*_R) - \partial^2_{\eta_R} l(\eta^*_R)|| = o_p(n)$. For the second term, since $||\eta^*_R - \eta_{R0}|| = O(\alpha_n) = o(1)$, we have $\{\partial^2_{\eta_R} l(\eta^*_R) - \partial^2_{\eta_R} l(\eta_{R0})\} = o_p(n)$.

Similar to Cai, Fan, Li, and Zhou (2005), it follows from the Chebyshev inequality that

$$P\{||n^{-1}\partial^2_{\eta_R} Q(\eta_{R0}) + \Sigma(\eta_{R0})||_F \geqslant r_n^{-1}\epsilon\} \leqslant \frac{r_n^4}{n\epsilon^2} = o(1)$$

as $r_n^4/n \to 0$ by condition (C1) combining the fact that $a > 1$. Thus, we have

$$||n^{-1}\partial^2_{\eta_R} Q(\eta_{R0}) + \Sigma(\eta_{R0})||_F = o_p(r_n^{-1}).$$

Hence, we have

$$J_3 = -\frac{n\alpha_n^2}{2} u'\Sigma(\eta_{R0})u(1 + o_p(1)).$$

By condition (A4), J_3 uniformly dominates both J_1 and J_2, which leads to (7.1). Thus, it is easy to see that $||\widehat{\eta}_R - \eta_{R0}|| = O_p(\alpha_n)$.

We prove the convergence rate of $||\widehat{\beta}(s) - \beta_0(s)||_{L_2}$ as follows:

$$\begin{aligned}
||\widehat{\beta}(s) - \beta_0(s)||_{L_2} &= ||\sum_{j=1}^{r_n} \widehat{\beta}_{j0}\widehat{\phi}_j(t) - \sum_{j=1}^{\infty} \beta_{j0}\phi_j(t)|| \\
&\leqslant ||\sum_{j=1}^{r_n} \widehat{\beta}_{j0}\widehat{\phi}_j(t) - \sum_{j=1}^{r_n} \beta_{j0}\widehat{\phi}_j(t)|| + ||\sum_{j=1}^{r_n} \beta_{j0}\widehat{\phi}_j(t) - \sum_{j=1}^{r_n} \beta_{j0}\phi_j(t)|| \\
&\quad + ||\sum_{j=r_n+1}^{\infty} \beta_{j0}\phi_j(t)|| \\
&\leqslant \sum_{j=1}^{r_n} \lambda_j^{-1}{}^{1/2} ||\widehat{\eta}_R - \eta_{R0}|| + \sum_{j=1}^{r_n} \beta_{j0}\delta_j^{-1}|||\widehat{K} - K||| + (\sum_{j=r_n+1}^{\infty} \beta_{j0}^2)^{1/2} \\
&= O_p(r_n^{a/2+1/2}\alpha_n) + O_p(r_n^{a-b+2}n^{-1/2}) + O_p(r_n^{-b+1/2})
\end{aligned}$$

It follows from (C1) and (C2) that $r_n^{a/2+1/2}\alpha_n = o(1)$. Since $b > 0$ and $r_n^{a+2}n^{-1/2} = o(1)$ by condition (C1), $r_n^{a/2-b+1} = o(1)$. Furthermore, the third term is $o_p(1)$ as $b > a/2+1 > 1/2$. Thus, we have $||\widehat{\beta}(s) - \beta_0(s)||_{L_2} = o_p(1)$.

Proof of Theorem 2.

We need to introduce some notation. Under H_0, we have $\beta_0 = 0$. Let γ_0 be the true value of γ and $\eta_0 = (0^{\mathrm{T}}, \gamma_0^{\mathrm{T}})^{\mathrm{T}}$. Define

$$S_{TE}(\eta) = \sum_{i=1}^{n}\int_0^{\tau} w_i dN_i(t) - \int_0^{\tau}\frac{E\{w_i Y_i(t)\exp(w_i^{\mathrm{T}}\eta)\}}{E\{Y_i(t)\exp(w_i^{\mathrm{T}}\eta)\}}d\overline{N}(t) = \sum_{i=1}^{n} S_{TE,i}(\eta),$$

where $S_{TE,i}(\eta) = \int_0^{\tau} w_i dN_i(t) - \int_0^{\tau}\frac{E\{w_i Y_i(t)\exp(w_i^{\mathrm{T}}\eta)\}}{E\{Y_i(t)\exp(w_i^{\mathrm{T}}\eta)\}}dN_i(t)$.

The first step is to prove that as $n \to \infty$, we have

$$[n^{-1/2}S_{TE}(\eta_0)^{\mathrm{T}}\{E(I_T(\eta_0))\}^{-1}n^{-1/2}S_{TE}(\eta_0) - r_n]/(2r_n)^{1/2} \to^d N(0,1). \qquad (7.2)$$

By using the standard martingale theory Kalbfleisch and Prentice (2002), we know that $S_{TE,i}(\eta_0)$ are i.i.d random variables with $E(S_{TE,i}(\eta_0)) = 0$ and $E\{S_{TE,i}(\eta_0)S_{TE,i}(\eta_0)^{\mathrm{T}}\} = E\{I_T(\eta_0)\}$. Let $\psi_i = \{E(I_T(\eta_0))\}^{-1/2}S_{TE,i}(\eta_0)$. Thus, ψ_i's are i.i.d random variables with mean 0 and covariance I_{r_n}. It follows from (A5) and (B5) that $E\{(S_{TE,i}(\eta_0)^{\mathrm{T}}S_{TE,i}(\eta))^2\} = O(r_n) = o(nr_n)$. Therefore, it follows from Lemma 9 that (7.2) is valid.

The second step is to prove that as $n \to \infty$, we have

$$[n^{-1/2}S_T(\eta_0)^{\mathrm{T}}\{E(I_T(\eta_0))\}^{-1}n^{-1/2}S_T(\eta_0) - r_n]/(2r_n)^{1/2} \to^d N(0,1). \qquad (7.3)$$

Since $\{Y_i(t) : t \in [0, \tau]\}$ is a Donsker class, we have

$$n^{-1}\sum_{i=1}^{n} w_i Y_i(t)\exp(w_i^{\mathrm{T}}\eta_0) - E\{w_i Y_i(t)\exp(w_i^{\mathrm{T}}\eta_0)\} = O_p(n^{-1/2}),$$
$$n^{-1}\sum_{i=1}^{n} Y_i(t)\exp(w_i^{\mathrm{T}}\eta_0) - E\{Y_i(t)\exp(w_i^{\mathrm{T}}\eta_0)\} = O_p(n^{-1/2}),$$

Thus, we have $S_T(\eta_0) - S_{TE}(\eta_0) = O_p(n^{1/2})$, which leads to (7.3).

Define $\eta_N = (0^{\mathrm{T}}, \gamma_N^{\mathrm{T}})^{\mathrm{T}}$, and notice that when $\beta = 0$,

$$Q_T(\eta_N) = Q(\eta_N) = \sum_{i=1}^{n}\int_0^{\tau} Z_i^{\mathrm{T}}\gamma dN_i(t) - \int_0^{\tau}\log\{\sum_{i=1}^{n} Y_i(t)\exp(Z_i^{\mathrm{T}}\gamma)\}d\overline{N}(t)$$

Also define $\widehat{\eta}_N = (0^{\mathrm{T}}, \widehat{\gamma}_N^{\mathrm{T}})^{\mathrm{T}}$ as the maximum partial likelihood estimate of $Q(\eta_N)$ under $\beta = 0$, we know that $\widehat{\eta}_N$ is also the maximum partial likelihood estimate of $Q_T(\eta_N)$ under $\beta = 0$.

The third step is to prove that

$$\|n^{-1}I_T(0, \widehat{\gamma}_N) - \{E(I_T(\eta_0))\}\|_F = O_p(r_n n^{-1/2}). \tag{7.4}$$

Following arguments in Theorem 8.3.2 in Fleming and Harrington (1991), we can show that $\widehat{\gamma}_N$ is a root-n consistent estimator of γ_0. We prove (7.4) by using Theorem 8.2.1(2) in Fleming and Harrington (1991). As $r_n = o(n^{1/2})$, one has $\|n^{-1}I_T(0, \widehat{\gamma}_N) - \{E(I_T(\eta_0))\}\|_F = o_p(1)$.

The fourth step is to show

$$[n^{-1/2}S_T^{\mathrm{T}}(0, \widehat{\gamma}_N)\{n^{-1}I_T(0, \widehat{\gamma}_N)\}^{-1}n^{-1/2}S_T(0, \widehat{\gamma}_N) - r_n]/(2r_n)^{1/2} \to^d N(0, 1). \tag{7.5}$$

Denote $S_T(\eta) = (\{\partial Q_T\beta(\eta)\}^{\mathrm{T}}, \{\partial Q_\gamma(\eta)\}^{\mathrm{T}})^{\mathrm{T}} = (S_{T1}^{\mathrm{T}}(\eta), S_{T2}^{\mathrm{T}}(\eta))^{\mathrm{T}}$. Since $\widehat{\gamma}_N$ is the maximum partial likelihood estimates under $\beta = 0$, we have $S_{T2}^{\mathrm{T}}(0, \widehat{\gamma}_N) = 0$. For $S_{T1}^{\mathrm{T}}(0, \widehat{\gamma}_N)$, we have $S_{T1}(0, \widehat{\gamma}_N) = S_1(0, \gamma_0) + \partial S_{1,\gamma}^{\mathrm{T}}(0, \gamma^*)(\widehat{\gamma}_N - \gamma_0)$, where γ^* is between $\widehat{\gamma}_N$ and γ_0. Since $\partial S_{T1,\gamma}^{\mathrm{T}}(0, \widehat{\gamma}_N) = \partial^2 Q_{T,\beta,\gamma}(0, \widehat{\gamma}_N) = 0$, it follows from Theorem 8.2.1(2) in Fleming and Harrington (1991) that

$\|n^{-1}\partial S_{T1,\gamma}^{\mathrm{T}}(0,\gamma^*)\|_F = O_p(r_n n^{-1/2})$. Since $n^{1/2}(\widehat{\gamma}_N - \gamma_0) = O_p(1)$, we have $n^{-1/2}||\{\partial S_{T1,\gamma}^{\mathrm{T}}(0,\widehat{\gamma}_N) - \partial S_{T1,\gamma}^{\mathrm{T}}(\eta_0)\}|| = O_p(r_n n^{-1/2})$. Finally, (7.5) follows from $r_n n^{-1/2} = o(1)$ and Slutsky's theorem.

Finally, it follows from Lemma 8 and $O_p(1)/(2r_n)^{1/2} = o_p(1)$ that we have

$$[n^{-1/2}S(0,\widehat{\gamma}_N)^{\mathrm{T}}\{n^{-1}I(0,\widehat{\gamma}_N)\}^{-1}n^{-1/2}S(0,\widehat{\gamma}_N) - r_n]/(2r_n)^{1/2} \to^d N(0,1).$$

8. Proofs of Corollaries

We consider the situation when p_n is diverging with sample size n. For most of the aforementioned lemmas and conditions, we only need to change p into p_n without changing their proofs. However, we need to change Lemma 5 into Lemma 10 as follows. We need a slightly stronger condition (A5*) than (A5) to prove Lemma 10.

LEMMA 10: *Let* $\pi_i = \eta_R^{\mathrm{T}}\widehat{w}_{iR} + \rho_i = \sum_{k=1}^{p_n} z_{ik}\gamma_k + \sum_{j=1}^{r_n}\xi_{ijR}\beta_{jR} + \sum_{j=r_n+1}^{\infty}\xi_{ij}\beta_{j0}$. *For any fixed positive integer* d, *under Conditions (A5*) and (B5), we have* $E[\{\exp(\pi_i)\}^{2d}] = O(1)$ *uniformly for* $||\eta|| \leqslant C$.

Proof of Lemma 10:

For any fixed positive integer d, we have

$$\begin{aligned}
E[\{\exp(\pi_i)\}^{2d}] &= E(\exp\{2d(\sum_{k=1}^{p_n} z_{ik}\gamma_k + \sum_{j=1}^{r_n}\xi_{ijR}\beta_{jR} + \sum_{j=r_n+1}^{\infty}\xi_{ij}\beta_{j0})\} \\
&\leqslant E[\prod_{k=1}^{p_n}\exp(2dz_{ik}\gamma_k) \times \exp\{(2d\sum_{j=1}^{r_n}\xi_{ij}^2)^{1/2}(\sum_{j=1}^{r_n}\beta_j^2)^{1/2}\} \\
&\quad \times \exp\{2d(\sum_{j=r_n+1}^{\infty}\xi_{ij}^2)^{1/2}(\sum_{j=r_n+1}^{\infty}\beta_{j0}^2)^{1/2}\}] \\
&\leqslant \prod_{k=1}^{p_n}\{E(\exp(2dz_{ik}\gamma_k))\}^{1/2} \times \exp(2d||X_i|| \times ||\eta||) \times \exp(2d||X_i|| \times ||\beta_0||)
\end{aligned}$$

By Condition (A5*),

$$\prod_{k=1}^{p_n}\{E(\exp(2dz_{ik}\gamma_k))\}^{1/2} \leqslant \prod_{k=1}^{p_n}\exp(2M^2d^2\gamma_k^2) = \exp\{2M^2d^2(\sum_{k=1}^{p_n}\gamma_k^2)\} \leqslant \exp(2M^2d^2||\eta||^2)$$

Combining with condition (B5), we have $E[\{\exp(\pi_i)\}^{2d}] = O(1)$ uniformly for $||\eta|| \leqslant C$.

After finishing proving Lemma 10, we are able to prove the following corollaries.

Proof of Corollary 1:

We can prove Corollary by repeating the *Proof of Theorem* 1 except that we use Lemma 10 and change p into p_n.

Proof of Corollary 2:

To prove Corollary 2, we make some additional changes when p_n is diverging. We first introduce some notation. Under H_0, we have $\beta_0 = 0$. Let γ_0 be the true value of γ and $\eta_0 = (0^{\mathrm{T}}, \gamma_0^{\mathrm{T}})^{\mathrm{T}}$. Define

$$S_{TE}(\eta) = \sum_{i=1}^{n}\int_0^\tau w_i dN_i(t) - \int_0^\tau \frac{E\{w_iY_i(t)\exp(w_i^{\mathrm{T}}\eta)\}}{E\{Y_i(t)\exp(w_i^{\mathrm{T}}\eta)\}}d\overline{N}(t) = \sum_{i=1}^{n} S_{TE,i}(\eta),$$

where $S_{TE,i}(\eta) = \int_0^\tau w_i dN_i(t) - \int_0^\tau \frac{E\{w_iY_i(t)\exp(w_i^{\mathrm{T}}\eta)\}}{E\{Y_i(t)\exp(w_i^{\mathrm{T}}\eta)\}}dN_i(t)$.

The first step is to prove that as $n \to \infty$, we have

$$[n^{-1/2}S_{TE}(\eta_0)^{\mathrm{T}}\{E(I_T(\eta_0))\}^{-1}n^{-1/2}S_{TE}(\eta_0) - r_n]/(2r_n)^{1/2} \to^d N(0,1). \tag{8.1}$$

By using the standard martingale theory (Kalbfleisch and Prentice, 2002), we know that $S_{TE,i}(\eta_0)$ are i.i.d random variables with $E(S_{TE,i}(\eta_0)) = 0$ and $E\{S_{TE,i}(\eta_0)S_{TE,i}(\eta_0)^{\mathrm{T}}\} = E\{I_T(\eta_0)\}$. Let $\psi_i = \{E(I_T(\eta_0))\}^{-1/2}S_{TE,i}(\eta_0)$. Thus, ψ_i's are i.i.d random variables with mean 0 and covariance I_{r_n}. It follows from (A5*) and (B5) that $E\{(S_{TE,i}(\eta_0)^{\mathrm{T}}S_{TE,i}(\eta))^2\} = O(r_n) = o(nr_n)$. Therefore, it follows from Lemma 9 that (8.1) is valid.

The second step is to prove that as $n \to \infty$, we have

$$[n^{-1/2} S_T(\eta_0)^{\mathrm{T}} \{E(I_T(\eta_0))\}^{-1} n^{-1/2} S_T(\eta_0) - r_n]/(2r_n)^{1/2} \to^d N(0,1). \tag{8.2}$$

Since $\{Y_i(t) : t \in [0, \tau]\}$ is a Donsker class, we have

$$n^{-1} \sum_{i=1}^{n} w_i Y_i(t) \exp(w_i^{\mathrm{T}} \eta_0) - E\{w_i Y_i(t) \exp(w_i^{\mathrm{T}} \eta_0)\} = O_p(n^{-1/2}),$$
$$n^{-1} \sum_{i=1}^{n} Y_i(t) \exp(w_i^{\mathrm{T}} \eta_0) - E\{Y_i(t) \exp(w_i^{\mathrm{T}} \eta_0)\} = O_p(n^{-1/2}),$$

Thus, we have $S_T(\eta_0) - S_{TE}(\eta_0) = O_p(n^{1/2})$, which leads to (8.2).

Define $\eta_N = (0^{\mathrm{T}}, \gamma_N^{\mathrm{T}})^{\mathrm{T}}$, and notice that when $\beta = 0$,

$$Q_T(\eta_N) = Q(\eta_N) = \sum_{i=1}^{n} \int_0^{\tau} Z_i^{\mathrm{T}} \gamma dN_i(t) - \int_0^{\tau} \log\{\sum_{i=1}^{n} Y_i(t) \exp(Z_i^{\mathrm{T}} \gamma)\} d\overline{N}(t)$$

Also define $\widehat{\eta}_N = (0^{\mathrm{T}}, \widehat{\gamma}_N^{\mathrm{T}})^{\mathrm{T}}$ as the maximum partial likelihood estimate of $Q(\eta_N)$ under $\beta = 0$, we know that $\widehat{\eta}_N$ is also the maximum partial likelihood estimate of $Q_T(\eta_N)$ under $\beta = 0$.

The third step is to prove that

$$\|n^{-1} I_T(0, \widehat{\gamma}_N) - \{E(I_T(\eta_0))\}\|_F = O_p(r_n^{3/2} n^{-1/2}). \tag{8.3}$$

Following arguments in Theorem 8.3.2 in Fleming and Harrington (1991), we can show that $||\widehat{\gamma}_N - \gamma_0|| = O_p(p_n/n^{1/2}) = O_p(r_n/n^{1/2})$. We prove (8.3) by using Theorem 8.2.1(2) in Fleming and Harrington (1991). As $r_n = o(n^{1/3})$, one has $\|n^{-1} I_T(0, \widehat{\gamma}_N) - \{E(I_T(\eta_0))\}\|_F = o_p(1)$.

The fourth step is to show

$$[n^{-1/2} S_T^{\mathrm{T}}(0, \widehat{\gamma}_N) \{n^{-1} I_T(0, \widehat{\gamma}_N)\}^{-1} n^{-1/2} S_T(0, \widehat{\gamma}_N) - r_n]/(2r_n)^{1/2} \to^d N(0,1). \tag{8.4}$$

Denote $S_T(\eta) = (\{\partial Q_T \beta(\eta)\}^{\mathrm{T}}, \{\partial Q_\gamma(\eta)\}^{\mathrm{T}})^{\mathrm{T}} = (S_{T1}^{\mathrm{T}}(\eta), S_{T2}^{\mathrm{T}}(\eta))^{\mathrm{T}}$. Since $\widehat{\gamma}_N$ is the maximum partial likelihood estimates under $\beta = 0$, we have $S_{T2}^{\mathrm{T}}(0, \widehat{\gamma}_N) = 0$. For $S_{T1}^{\mathrm{T}}(0, \widehat{\gamma}_N)$, we have $S_{T1}(0, \widehat{\gamma}_N) = S_1(0, \gamma_0) + \partial S_{1,\gamma}^{\mathrm{T}}(0, \gamma^*)(\widehat{\gamma}_N - \gamma_0)$, where γ^* is between $\widehat{\gamma}_N$ and γ_0. Since $\partial S_{T1,\gamma}^{\mathrm{T}}(0, \widehat{\gamma}_N) =$

$\partial^2 Q_{T,\beta,\gamma}(0,\widehat{\gamma}_N) = 0$, it follows from Theorem 8.2.1(2) in Fleming and Harrington (1991) that $\|n^{-1}\partial S^{\mathrm{T}}_{T1,\gamma}(0,\gamma^*)\|_F = O_p(r_n n^{-1/2})$. Since $r_n^{-1/2} n^{1/2}(\widehat{\gamma}_N - \gamma_0) = O_p(1)$, we have $n^{-1/2}||\{\partial S^{\mathrm{T}}_{T1,\gamma}(0,\widehat{\gamma}_N) - \partial S^{\mathrm{T}}_{T1,\gamma}(\eta_0)\}|| = O_p(r_n^{3/2} n^{-1/2})$. Finally, (8.4) follows from $r_n^{3/2} n^{-1/2} = o(1)$ and Slutsky's theorem.

Finally, it follows from Lemma 8 and $O_p(1)/(2r_n)^{1/2} = o_p(1)$ that we have

$$[n^{-1/2} S(0,\widehat{\gamma}_N)^{\mathrm{T}}\{n^{-1} I(0,\widehat{\gamma}_N)\}^{-1} n^{-1/2} S(0,\widehat{\gamma}_N) - r_n]/(2r_n)^{1/2} \rightarrow^d N(0,1).$$

References

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.

Cai, J., Fan, J., Li, R., and Zhou, H. (2005). Variable selection for multivariate failure time data. *Biometrika* **92,** 303–316.

Fan, J. and Li, R. (2002). Variable selection for cox's proportional hazards model and frailty model. *The Annals of Statistics* **30,** 74–99.

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York.

Gellar, J. E., Colantuoni, E., Needham, D. M., and Crainiceanu, C. M. (2015). Cox regression models with functional covariates for survival data. *Statistical Modelling* **15,** 256–278.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35,** 70–91.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society, Series B* **68,** 109–126.

Hall, P. and Hosseini-Nasab, M. (2009). Theory for high-order bounds in functional principal components analysis. *Math. Proc. Cambridge Philos. Soc.* **146,** 225–256.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics* **34,** 1493–1517.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition.

Lei, J. (2014). Adaptive global testing for functional linear models. *Journal of the American Statistical Association* **109,** 624–634.

Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3d surface construction algorithm. *Computer Graphics* **21,** 163–169.

Murphy, S. A. and Van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association* **95,** 449–465.

Patenaude, B., Smith, S. M., Kennedy, D. N., and Jenkinson, M. (2011). A bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage* pages 907–922.

Peng, H. and Schick, A. (2014). Asymptotic normality of quadratic forms with random vectors of increasing dimension. *ftp://ftp.math.binghamton.edu/pub/anton/cltqf.pdf* .

Shi, J., Thompson, P. M., Gutman, B., and Wang, Y. (2013). Surface fluid registration of conformal representation: application to detect disease burden and genetic influence on hippocampus. *Neuroimage* **78,** 111–134.

Shi, J., Wang, Y., Ceschin, R., An, X., Lao, Y., Vanderbilt, D., Nelson, M., Thompson, P. M., Panigrahy, A., and Lepore, N. (2013). A multivariate surface-based analysis of the putamen in premature newborns: Regional differences within the ventral striatum. *PLoS ONE* **8,** e66736.

Wang, Y., Zhang, J., Gutman, B., Chan, T. F., Becker, J. T., Aizenstein, H. J., Lopez, O. L., Tamburo, R. J., Toga, A. W., and Thompson, P. M. (2010). Multivariate tensor-based morphometry on surfaces: Application to mapping ventricular abnormalities in HIV/AIDS. *NeuroImage* **49,** 2141 – 2157.

Zhang, J. and Chen, J. (2007). Statistical inference for functional data. *The Annals of Statistics* **35,** 1052–1079.

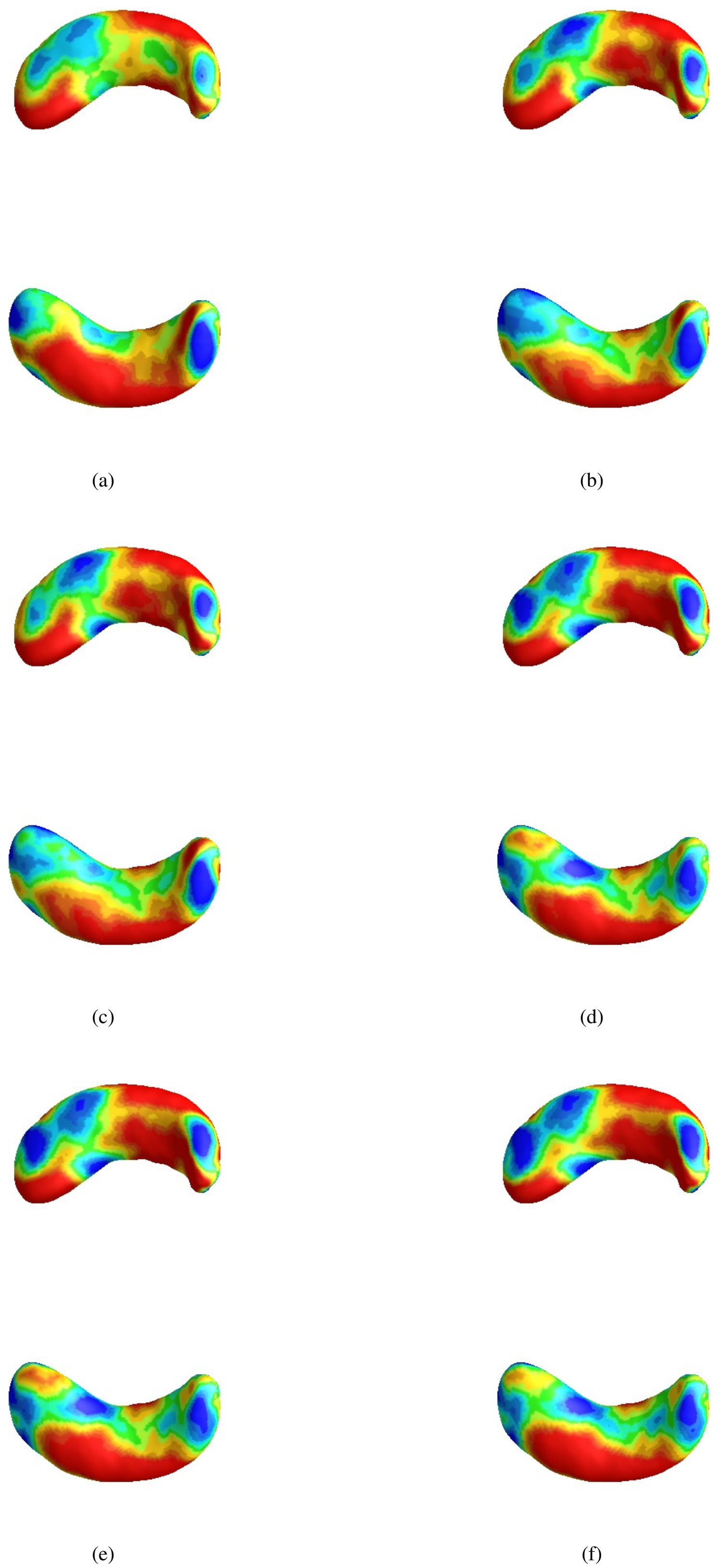

Figure S1: ADNI data analysis results: panels (a)-(f) contain the estimated coefficient functions $\beta(s)$ when $r_n = 17, 18, 19, 21, 22,$ and 23, respectively.

Table S1. Simulation results for the means of the estimates of RMSE_β, RMSE_γ, $\widehat{r}_n$ and the concordance index with their standard errors in the parentheses when $n = 200$ and censoring rate is 0.3. We vary r_n from 1 to 10, and use AIC to select r_n, and we also compare with Gellar et al. (2015)'s method.

r_n	RMSE_β	RMSE_γ	Concordance Index	$\widehat{r}_n$
1	0.22 (0.007)	0.29(0.021)	0.738(0.0004)	1
2	0.16 (0.008)	0.3(0.022)	0.742(0.0005)	2
3	0.13 (0.008)	0.3(0.021)	0.744(0.0005)	3
4	0.11 (0.008)	0.3(0.023)	0.745(0.0004)	4
5	0.17 (0.011)	0.3(0.022)	0.745(0.0004)	5
6	0.30 (0.023)	0.31(0.023)	0.743(0.0005)	6
7	0.56 (0.046)	0.32(0.024)	0.742(0.0005)	7
8	0.85 (0.062)	0.34(0.023)	0.741(0.0006)	8
9	1.34 (0.089)	0.34(0.024)	0.74(0.0006)	9
10	1.89 (0.121)	0.34(0.024)	0.738(0.0007)	10
AIC	0.28(0.055)	0.3(0.023)	0.744(0.0005)	3.35(0.16)
Gellar	0.27(0.055)	0.31(0.022)	0.744(0.0005)	NA
Qu	3.70 (0.05)	4.35(0.08)	NA	NA

Table S2. Simulation results for the means of the estimates of RMSE_β, RMSE_γ, $\widehat{r}_n$ and the concordance index with their standard errors in the parentheses when $n = 200$ and censoring rate is 0.5. We vary r_n from 1 to 10, and use AIC to select r_n, and we also compare with Gellar et al. (2015)'s method.

r_n	RMSE_β	RMSE_γ	Concordance Index	$\widehat{r}_n$
1	0.22(0.007)	0.39(0.029)	0.744(0.0005)	1
2	0.17(0.008)	0.4(0.028)	0.747(0.0005)	2
3	0.16(0.01)	0.4(0.028)	0.748(0.0005)	3
4	0.14(0.011)	0.41(0.029)	0.749(0.0005)	4
5	0.26(0.021)	0.42(0.031)	0.748(0.0005)	5
6	0.44(0.036)	0.44(0.033)	0.747(0.0006)	6
7	0.78(0.069)	0.45(0.035)	0.745(0.0006)	7
8	1.23(0.084)	0.49(0.038)	0.743(0.0007)	8
9	1.97(0.115)	0.49(0.039)	0.742(0.0008)	9
10	2.86(0.165)	0.5(0.041)	0.74(0.0008)	10
AIC	0.42(0.079)	0.42(0.032)	0.747(0.0007)	3.24(0.17)
Gellar	0.36(0.059)	0.41(0.03)	0.748(0.0006)	NA
Qu	3.74 (0.06)	4.55(0.10)	NA	NA

Table S3. Simulation results for the means of the estimates of RMSE$_\beta$, RMSE$_\gamma$, $\widehat{r}_n$ and the concordance index with their standard errors in the parentheses when $n = 500$ and censoring rate is 0.1. We vary r_n from 1 to 10, and use AIC to select r_n, and we also compare with Gellar et al. (2015)'s method.

r_n	RMSE$_\beta$	RMSE$_\gamma$	Concordance Index	$\widehat{r}_n$
1	0.19(0.004)	0.1(0.009)	0.734(0.0002)	1
2	0.12(0.005)	0.1(0.009)	0.738(0.0003)	2
3	0.07(0.004)	0.1(0.009)	0.741(0.0002)	3
4	0.04(0.002)	0.1(0.009)	0.742(0.0001)	4
5	0.06(0.003)	0.1(0.009)	0.742(0.0001)	5
6	0.09(0.007)	0.1(0.009)	0.742(0.0001)	6
7	0.17(0.012)	0.1(0.009)	0.741(0.0001)	7
8	0.26(0.019)	0.11(0.009)	0.741(0.0001)	8
9	0.39(0.031)	0.11(0.01)	0.74(0.0002)	9
10	0.58(0.04)	0.11(0.01)	0.74(0.0002)	10
AIC	0.11(0.018)	0.1(0.009)	0.742(0.0002)	4.1(0.14)
Gellar	0.15(0.02)	0.1(0.009)	0.742(0.0001)	NA
Qu	3.58 (0.02)	4.05(0.04)	NA	NA

Table S4. Simulation results for the means of the estimates of RMSE_β, RMSE_γ, $\widehat{r}_n$ and the concordance index with their standard errors in the parentheses when $n = 500$ and censoring rate is 0.3. We vary r_n from 1 to 10, and use AIC to select r_n, and we also compare with Gellar et al. (2015)'s method.

r_n	RMSE_β	RMSE_γ	Concordance Index	$\widehat{r}_n$
1	0.19(0.004)	0.13(0.01)	0.742(0.0002)	1
2	0.12(0.005)	0.13(0.01)	0.745(0.0003)	2
3	0.07(0.004)	0.13(0.011)	0.748(0.0002)	3
4	0.04(0.002)	0.13(0.011)	0.75(0.0001)	4
5	0.07(0.004)	0.13(0.011)	0.749(0.0001)	5
6	0.12(0.008)	0.13(0.011)	0.749(0.0002)	6
7	0.22(0.015)	0.13(0.011)	0.748(0.0002)	7
8	0.34(0.021)	0.13(0.011)	0.748(0.0002)	8
9	0.5(0.035)	0.14(0.012)	0.747(0.0003)	9
10	0.72(0.044)	0.14(0.012)	0.747(0.0003)	10
AIC	0.13(0.019)	0.13(0.010)	0.749(0.0002)	4.1(0.14)
Gellar	0.12(0.009)	0.13(0.01)	0.749(0.0002)	NA
Qu	3.58 (0.03)	4.14(0.04)	NA	NA

Table S5. Simulation results for the means of the estimates of RMSE_β, RMSE_γ, $\widehat{r}_n$ and the concordance index with their standard errors in the parentheses when $n = 500$ and censoring rate is 0.5. We vary r_n from 1 to 10, and use AIC to select r_n, and we also compare with Gellar et al. (2015)'s method.

r_n	RMSE_β	RMSE_γ	Concordance Index	$\widehat{r}_n$
1	0.19(0.004)	0.17(0.013)	0.748(0.0003)	1
2	0.13(0.005)	0.17(0.014)	0.752(0.0003)	2
3	0.08(0.005)	0.17(0.015)	0.754(0.0003)	3
4	0.06(0.003)	0.18(0.016)	0.755(0.0002)	4
5	0.1(0.006)	0.18(0.016)	0.755(0.0002)	5
6	0.17(0.012)	0.18(0.015)	0.754(0.0002)	6
7	0.31(0.022)	0.18(0.015)	0.753(0.0002)	7
8	0.45(0.031)	0.19(0.015)	0.752(0.0003)	8
9	0.68(0.048)	0.19(0.016)	0.752(0.0003)	9
10	1.02(0.06)	0.19(0.016)	0.751(0.0003)	10
AIC	0.2(0.032)	0.18(0.015)	0.754(0.0003)	3.84(0.16)
Gellar	0.17(0.022)	0.18(0.015)	0.754(0.0002)	NA
Qu	3.61 (0.03)	4.23(0.05)	NA	NA

Table S6. Simulation results for the means of the estimates of RMSE_β, RMSE_γ, $\widehat{r}_n$ and the concordance index with their standard errors in the parentheses when $n = 1,000$ and censoring rate is 0.1. We vary r_n from 1 to 10, and use AIC to select r_n, and we also compare with Gellar et al. (2015)'s method.

r_n	RMSE_β	RMSE_γ	Concordance Index	$\widehat{r}_n$
1	0.19(0.003)	0.05(0.004)	0.735(0.0002)	1
2	0.11(0.005)	0.05(0.004)	0.739(0.0002)	2
3	0.05(0.003)	0.05(0.004)	0.742(0.0002)	3
4	0.03(0.001)	0.05(0.004)	0.743(0.0001)	4
5	0.03(0.002)	0.05(0.004)	0.743(0.0001)	5
6	0.05(0.003)	0.05(0.004)	0.743(0.0001)	6
7	0.08(0.005)	0.05(0.004)	0.743(0.0001)	7
8	0.11(0.007)	0.05(0.004)	0.743(0.0001)	8
9	0.17(0.01)	0.05(0.004)	0.743(0.0001)	9
10	0.24(0.013)	0.05(0.004)	0.742(0.0001)	10
AIC	0.05(0.006)	0.05(0.004)	0.743(0.0001)	4.28(0.14)
Gellar	0.06(0.007)	0.05(0.004)	0.743(0.0001)	NA
Qu	3.62 (0.02)	3.95(0.03)	NA	NA

Table S7. Simulation results for the means of the estimates of RMSE_β, RMSE_γ, $\widehat{r}_n$ and the concordance index with their standard errors in the parentheses when $n = 1,000$ and censoring rate is 0.3. We vary r_n from 1 to 10, and use AIC to select r_n, and we also compare with Gellar et al. (2015)'s method.

r_n	RMSE_β	RMSE_γ	Concordance Index	$\widehat{r}_n$
1	0.19(0.003)	0.07(0.006)	0.742(0.0002)	1
2	0.11(0.005)	0.07(0.005)	0.747(0.0002)	2
3	0.06(0.003)	0.06(0.005)	0.749(0.0002)	3
4	0.03(0.001)	0.06(0.005)	0.751(0.0001)	4
5	0.04(0.002)	0.06(0.005)	0.751(0.0001)	5
6	0.06(0.003)	0.06(0.005)	0.751(0.0001)	6
7	0.1(0.006)	0.07(0.005)	0.75(0.0001)	7
8	0.15(0.009)	0.07(0.005)	0.75(0.0001)	8
9	0.23(0.013)	0.07(0.005)	0.75(0.0001)	9
10	0.32(0.018)	0.07(0.005)	0.75(0.0001)	10
AIC	0.07(0.009)	0.07(0.005)	0.751(0.0001)	4.23(0.15)
Gellar	0.09(0.011)	0.06(0.005)	0.751(0.0001)	NA
Qu	3.64 (0.02)	3.98(0.03)	NA	NA

Table S8. Simulation results for the means of the estimates of RMSE_β, RMSE_γ, $\widehat{r}_n$ and the concordance index with their standard errors in the parentheses when $n = 1,000$ and censoring rate is 0.5. We vary r_n from 1 to 10, and use AIC to select r_n, and we also compare with Gellar et al. (2015)'s method.

r_n	RMSE_β	RMSE_γ	Concordance Index	$\widehat{r}_n$
1	0.19(0.003)	0.1(0.008)	0.749(0.0002)	1
2	0.11(0.005)	0.09(0.007)	0.753(0.0002)	2
3	0.06(0.003)	0.09(0.007)	0.755(0.0002)	3
4	0.03(0.001)	0.09(0.007)	0.757(0.0001)	4
5	0.05(0.004)	0.09(0.007)	0.757(0.0001)	5
6	0.08(0.005)	0.09(0.007)	0.756(0.0001)	6
7	0.14(0.009)	0.09(0.008)	0.756(0.0001)	7
8	0.2(0.012)	0.09(0.008)	0.756(0.0001)	8
9	0.32(0.018)	0.09(0.008)	0.755(0.0002)	9
10	0.43(0.023)	0.1(0.008)	0.755(0.0002)	10
AIC	0.09(0.012)	0.09(0.007)	0.756(0.0001)	4.18(0.14)
Gellar	0.1(0.013)	0.09(0.007)	0.756(0.0001)	NA
Qu	3.67 (0.02)	4.05(0.04)	NA	NA

Table S9. Simulation results on the power of our test when $n = 200$, censoring rate is 0.1 and $C_1 = 0.1 * j$ for $j = 0, 1, \cdots, 10$. We vary PV(r_n) from 0.70 to 0.95 to select r_n. The Type I error rates corresponding to $C_1 = 0$ were calculated from $5,000$ simulated data sets and the Type II error rates were calculated from 500 simulated data sets.

PV(r_n)	$C_1 = 0$	$C_1 = 0.1$	$C_1 = 0.2$	$C_1 = 0.3$	$C_1 = 0.4$	$C_1 = 0.5$	$C_1 = 0.6$	$C_1 = 0.7$	$C_1 = 0.8$	$C_1 = 0.9$	$C_1 = 1$
0.70	0.060	0.080	0.202	0.384	0.624	0.808	0.912	0.976	1.000	0.998	1.000
0.75	0.059	0.082	0.174	0.346	0.622	0.798	0.924	0.984	1.000	1.000	1.000
0.80	0.059	0.082	0.176	0.344	0.620	0.800	0.924	0.984	1.000	1.000	1.000
0.85	0.059	0.082	0.184	0.340	0.616	0.802	0.926	0.982	0.998	1.000	1.000
0.90	0.060	0.088	0.154	0.332	0.548	0.766	0.890	0.954	0.996	0.998	1.000
0.95	0.066	0.082	0.160	0.282	0.490	0.734	0.862	0.942	0.990	0.998	1.000

Table S10. Simulation results on the power of our test when $n = 200$, censoring rate is $0..1$, and $C_2 = 0.1 * j$ for $j = 0, 1, \cdots, 10$. We vary PV(r_n) from 0.70 to 0.95 to select r_n. The Type I error rates corresponding to $C_2 = 0$ were calculated from $5,000$ simulated data sets and the Type II error rates were calculated from 500 simulated data sets.

r_n	$C_2 = 0$	$C_2 = 0.1$	$C_2 = 0.2$	$C_2 = 0.3$	$C_2 = 0.4$	$C_2 = 0.5$	$C_2 = 0.6$	$C_2 = 0.7$	$C_2 = 0.8$	$C_2 = 0.9$	$C_2 = 1$
0.70	0.060	0.066	0.138	0.246	0.434	0.602	0.778	0.882	0.978	0.990	0.992
0.75	0.059	0.070	0.120	0.218	0.392	0.564	0.738	0.870	0.968	0.988	0.994
0.80	0.059	0.070	0.120	0.214	0.390	0.564	0.736	0.870	0.968	0.988	0.994
0.85	0.059	0.068	0.128	0.210	0.382	0.562	0.736	0.870	0.960	0.986	0.994
0.90	0.060	0.070	0.102	0.206	0.334	0.520	0.670	0.826	0.938	0.978	0.988
0.95	0.066	0.076	0.102	0.198	0.302	0.482	0.620	0.786	0.920	0.964	0.978