ABSTRACT

Title of dissertation: EXTENDED ESTIMATING EQUATIONS AND EMPIRICAL LIKELIHOOD

Xuan Yao, Doctor of Philosophy, 2018

Dissertation directed by: Dr. Eric Slud
Department of Mathematics

Classic Estimating Equations (CEE) were first introduced by Godambe in [16] and have been widely used under both parametric and nonparametric settings. However, under some prominent semiparametric models, CEE cannot be used to identify certain low-dimensional parameters. We prove that under regularity conditions, for the Cox (1972) survival-time model, a CEE for the structural parameter does not exist; and under more restrictive conditions, a CEE for the structural parameter in the *Accelerated Failure Time* (AFT) model does not exist when lifetime is subject to random right censoring with unknown distribution. Motivated by this lack of coverage of CEE's for finite-dimensional parameters in semiparametric problems, we establish a method named *Extended Estimating Equation* (EEE). The EEE's relax the requirement in the CEE of which the estimating function must be a function of the *independently identically distributed* (i.i.d.) summands and instead allow the estimating function to incorporate ratio of the sums of functions depending on two of the i.i.d. arguments.

To our knowledge, the broadest class of semiparametric models that can be investigated using EEE is the φ-transformation model class that we construct, where φ is a given function of covariate, structural parameter and random error with unknown hazard rate. With different choice of φ, the model can represent the general transformation model, nonlinear location-shift model, models incorporating cumulative integrated functions of times at risk and others. Inspired by Tsiatis's work in [38], by defining martingale structure on the residual scale, we are able to prove the asymptotic linearity of the associated EEE, which leads to the asymptotic normality of the structural estimator.

Another perspective from which to view EEE is to use it as a constraint in the *Empirical Likelihood* (EL) method. We first show that under the CEE setting, regardless of the continuity of the criterion function, there exists a neighbourhood of the true structural parameter on which there always exists a probability vector that maximizes the EL. The same conclusion can be generalized to the EEE setting with continuous criterion function as well as the discontinuous criterion function with the martingale structure of the φ-transformation model or the Cox model. A point estimator for the structural parameter can be defined via maximizing the *Profile Empirical Likelihood* (pEL) associated with the EEE. We show that the pEL estimator is asymptotically normal, with asymptotic variance covariance matrix identical to that of the Z-estimator obtained by directly solving for the root of EEE.

Finally, we develop algorithms to compute and compare the Z-estimator and pEL estimator associated with the EEE, and decide the minimal sample size for the two estimators to achieve asymptotic normality under three different parametric settings. Simulation shows a more symmetric covariate usually leads to a smaller threshold sample size, and the Z-estimator and pEL estimator are close in value and variance -covariance matrices. We also conclude that the pEL function tends to be much smoother, in settings where the EEE criterion function is non-smooth, than EEE itself, by comparing the plots of the projection of each function.

EXTENDED ESTIMATING EQUATION AND EMPIRICAL LIKELIHOOD

by

Xuan Yao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Dr. Eric Slud, Chair/Advisor
Dr. Shuo Chen
Dr. Benjamin Kedem
Dr. Yan Li, Dean's Representative
Dr. Tingni Sun

Acknowledgments

I would like to express my gratitude to everyone who has been there during my graduate life. Thank you for being part of this unforgettable and fruitful experience.

I would like to thank my advisor, Dr. Eric Slud. Thank you for giving me all the challenging yet intriguing questions. He is always patient with my questions, and has pulled me out of seemingly unsolvable situations with his advice. He has taught me how to view mathematical and statistical questions not as a student, but as a researcher. It has been an honor to learn from him.

I would like to thank Dr. Benjamin Kedem, Dr. Tingni Sun, Dr. Shuo Chen and Dr. Yan Li for serving on my dissertation committee. Thank you for reviewing my work thoroughly and for all the valuable comments and suggestions. Thanks to all the other professors in the Mathematics department who have taught me.

I would like to thank my parents and my husband for been supportive consistently during the past seven years. Thanks to Brook Qiu, a member who joined my family in 2014, who has always been the brightest sunshine in my life since then, and who can smooth any difficulties away with her laughter.

I would like to thank my peer and my friend Xia Li. Thank you for running towards the finish line with me together, and pushing me with your diligence and intelligence during the hard times. Thanks to all my other friends.

Table of Contents

List of Abbreviations and Notations

AFT	*Accelerated Lifetime Model*	page 4
CEE	*Classic Estimating Equation*	page 1
CLT	*Central Limit Theorem*	page 42
EEE	*Extended Estimating Equation*	page 5
EL	*Empirical Likelihood*	page 1
GTM	*General Transformation Model*	page 4
i.i.d.	*Independently Identiccally Distributed*	page 1
LIL	*Law of Iterated Logarithm*	page 70
LLN	*Law of Large Numbers*	page 40
MLE	*Maximum Likelihood Estimator*	page 1
T_i	Lifetime	page 6
C_i	Right-Censoring Variable	page 6
$\Delta_i = I\{T_i \leq C_i\}$	Non-Censored Indicator	page 6
$V_i = \min(T_i, C_i)$	Event Time	page 6
Z_i	Covariate	page 6
$\mathcal{Z}$	Support of Z_1	page 34
ε_i	Random Error with Unknown Hazard Rate	page 34
X_i	i.i.d. Observations	page 1
β	p-Dimensional Structural Parameter	page 1
ν	Infinite Dimensional Nuisance Parameter	page 1
$\theta = (\beta, \nu)$	Parameter in Semiparametric Models	page 1
$\mathcal{H}$	Infinite Dimensional Nuisance Parameter Space	page 1
$m(x, \beta)$	Classic Estimating Function	page 2
$m_n(x, \boldsymbol{x}, \beta)$	Extended Estimating Function	page 6
$S_n(\beta)$	Summation of $m(X_i, \beta)$ or $m_n(X_i, \boldsymbol{X}, \beta)$	page 2
$\mathrm{Conv}(\beta)$	Convex hull of $m_n(X_i, \boldsymbol{X}, \beta)$	page 83
$int\,(\mathrm{Conv}(\beta))$	Interior of $\mathrm{Conv}(\beta)$	page 86

Chapter 1: Introduction

The method of *Classic Estimating Equations* (CEE), first raised by Godambe in [16], is a powerful tool for constructing estimators for the structural parameter in a semiparametric model and has been extensively discussed in statistical literature. Given a mean zero estimating function depending on data and structural parameter alone, the CEE method defines an estimator as the root of the empirical integral of the criterion function. Such an estimator is known as a *Z-Estimator*.

To give a formal definition of CEE, let us consider *independently identically distributed* (i.i.d.) observations

$$X_1 \ldots, X_n \overset{i.i.d.}{\sim} P_\theta, \qquad \theta = (\beta, \nu) \in \mathbb{R}^p \times \mathcal{H}, \tag{1.1}$$

where $\{X_i\}_{i=1}^n$ are d-dimensional random vectors with support $\mathcal{X}$, $\beta \in \mathbb{R}^p$ is a finite-dimensional structural parameter, ν is a nuisance parameter in an infinite-dimensional space $\mathcal{H}$ such as a function space. Let β_0 and ν_0 respectively denote the true structural and nuisance parameter value.

1.1 Classic Estimating Equations

In our usage, CEE is a summation of mean zero functions of single independent data elements and structural parameter only, called *estimating equations*.

Definition 1.1.1 (Classic Estimating Equation) *Let $X_1 \ldots, X_n$ be i.i.d. observations as in (1.1), and $m(x, \beta) : \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}^d$ with*

$$E_\theta \{m(X_1, \beta)\} = 0, \quad \textit{for all} \qquad \theta = (\beta, \nu), \quad \nu \in \mathcal{H} \tag{1.2}$$

where β is a generally a proper subvector of the whole unknown parameter θ. If there exists a set U_{β_0}, a neighbourhood of β_0 such that $\hat{\beta}_n$ is a unique solution to

$$S_n(\beta) = \sum_{i=1}^{n} m(X_i, \beta) = 0, \qquad \textit{for } \beta \in U_{\beta_0} \tag{1.3}$$

then (1.3) is called an estimating equation for β.

CEE provide consistent and asymptotically normal estimators under regularity conditions including

$$E_{\theta_0} \{\nabla_\beta m(X_1, \beta)\} \text{ is non singular for } \beta \in U_{\beta_0}, \tag{1.4}$$

where $U_{\beta_0} \subset \mathbb{R}^p$ is a neighbourhood of β_0. Such conditions can be found in various statistical literature, such as Theorem 5.7 and Theorem 5.23 in [41].

A well known example of estimator that is constructed by means of estimating equation is the *Maximum Likelihood Estimator* (MLE) when ν is not present. Consider a simple parametric case in which the parameter is $\theta = \beta$. Then the MLE

maximizes $\prod_{i=1}^{n} f(x;\theta)$, or equivalently, $\sum_{i=1}^{n} \ln f(X_i;\theta)$, where $f(x;\theta)$ is the density function of X_1 with respect to the Lebesgue measure. If $f(x,\theta)$ is differentiable with respect to θ for each fixed value x, then the MLE is a solution to

$$\sum_{i=1}^{n} l(\theta; X_i) = 0, \qquad \text{where} \qquad l(\theta; x) = \partial \ln f(x;\theta)/\partial\theta \tag{1.5}$$

Note that under regularity conditions such as those summarized in Section 3.2 of [42], then $l(\theta; x)$ has mean zero. Under further regularity conditions guaranteeing the other parts of Definition 1.1.1, (1.5) is a classic estimating equation with the choice of $m(x,\theta) = l(\theta; x)$.

The CEE is also related to the *Empirical Likelihood* (EL) method as shown by Owen in [31]. Using the criterion function associated with CEE as a constraint, under a non-parametric setting, Owen established the EL ratio confidence intervals for a single functional in [32]. Both the Z-estimator and Owen's theory in [32] require the dimension of the CEE to be equal to that of the structural parameter. By allowing the former to exceed the latter, Qin and Lawless generalized Owen's conclusion in [33]. Qin and Lawless also constructed a point estimator for the structural parameter by maximizing the EL, and showed that such an estimator is asymptotically normal with a sandwich-formed asymptotic variance covariance matrix.

Despite the positive features of CEE, it has limitations in some prominent semi-parametric models in survival analysis, such as the Cox model (restircted to the non-time-dependent version in this thesis) and *Accelerated Lifetime Model* (AFT).

Cox model, or proportional hazard rate model was proposed by Cox in [11]. It assumes that the conditional hazard rate function given covariates is proportional (as a function of time) to the nuisance or "baseline" hazard function, by a factor depending on a linear combination of the covariates. Cox proposed to estimate the structural regression-coefficient parameter via maximizing the partial likelihood in [12], and Anderson and Gill developed the essential martingale-based large-sample distributional properties in [2]. Efron discussed the effciency of the partial liekli-hood estimator of the Cox model in [14]. In Section 5.2 of [39], Tsiatis showed the structural estimator constructed via maximizing the partial likelihood is globally semiparametric efficient. However, the equation system associated with the partial likelihood does not fit the definition of CEE due to the appearance of the quotient of higher order summations.

When the Cox model does not fit a possibly right-censored sample of survival data, an important alternative model is AFT. AFT can be considered as a special case of a *General Transformation Model* (GTM). The latter assumes that the lifetime transformed via a known monotone function depends linearly on the covariate with unknown regression coefficient plus an independent random error with unknown hazard rate function. When the monotone transforming function is chosen to be natural logarithm, then the GTM becomes AFT. An equivalent construction of AFT is to assume that the lifetime is conditionally given the covariates proportional by a factor equal to an exponential function of a linear combination of the covariates to some unknown baseline lifetime that is independent of the covariates. AFT has

been extensively investigated by Miller [30], Buckley and James [8], Koul et al. [25], Louis [28], Wei and Gail [44], James and Smith [22], Ritov and Wellner in [35], Lai and Ying in [26], Wei et al in [45], and Ritov in [34]. Tsiatis proposed a class of linear rank statistic estimators by constructing a martingale on the residual scale in [38]. He also showed that the "estimating equation" through which the structural parameter is defined is asymptotically linear, and he established the asymptotic normality of the structural estimator using the asymptotic linearity and martingale central limit theorem. In [34], Ritov showed that the linear rank statistic estimator is efficient. However, like the equation related to partial likelihood for Cox model, the "estimating equation" in [38] for AFT also involves a quotient of higher order summations and again does not fit the definition of CEE.

As shown by Cox and Tsiatis in [13] and [38], the structural estimators constructed under Cox and AFT model assumptions are usually related solving equations that are summations of non i.i.d. distributed summands, and the non-i.i.d violates the usual assumption on CEE. In fact, we prove that under regularity conditions, a CEE does not exist for Cox model. For the AFT, under more restrictive regularity conditions and with right censored data, the CEE does not exist, either.

1.2 Extended Estimating Equations

In order to extend the regime of CEE to cover right censored semiparametric models, we define the class of *Extended Estimating Equations* (EEE) by allowing the estimating function to depend not only on single observations but on quotients

of averages with respect to one index of functions of the structural parameter and two observations from the sample. Recall that for Cox model and *Accelerated Lifetime Model* (AFT), the estimators for β are usually constructed by solving

$$\text{AFT Model: } \sum_{i=1}^{n} \Delta_i \left\{ Z_i - \frac{\sum_{j=1}^{n} Z_j I\left\{\ln(V_j) - \beta^{tr} Z_j \geq \ln(V_i) - \beta^{tr} Z_i\right\}}{\sum_{j=1}^{n} I\left\{\ln(V_j) - \beta^{tr} Z_j \geq \ln(V_i) - \beta^{tr} Z_i\right\}} \right\} \tag{1.6}$$

$$\text{Cox Model: } \sum_{i=1}^{n} \Delta_i \left\{ Z_i - \frac{\sum_{j=1}^{n} Z_j I\left\{V_j \geq V_i\right\} e^{\beta^{tr} Z_j}}{\sum_{j=1}^{n} I\left\{V_j \geq V_i\right\} e^{\beta^{tr} Z_j}} \right\}, \tag{1.7}$$

where notations of in (1.6) - (1.7) are defined on page v. Evidently, (1.6) and (1.7) do not satisfy the definition of classic estimating equations, because of the quotient of two i.i.d. summations within the curly brackets. Inspired by the formulation of the summands in (1.6) and (1.7), let

$$m_n(X_i, \boldsymbol{X}, \beta) = Q(X_i, \beta) \left\{ C(X_i) - \frac{\sum_{j=1}^{n} C(X_j) k(X_i, X_j, \beta)}{\sum_{j=1}^{n} k(X_i, X_j, \beta)} \right\}. \tag{1.8}$$

For both the AFT and Cox models, $X_i = (T_i, C_i, Z_i, \Delta_i)$, and we can choose

$$Q(X_i, \beta) = \Delta_i = I\left\{T_i \leq C_i\right\}, \qquad C(X_i) = Z_i$$

and

$$\text{for AFT Model: } k(X_i, X_j, \beta) = I\left\{\ln(V_j) - \beta^{tr} Z_j \geq \ln(V_i) - \beta^{tr} Z_i\right\}$$

$$\text{for Cox Model: } k(X_i, X_j, \beta) = I\left\{V_j \geq V_i\right\} e^{\beta^{tr} Z_j}$$

Then an extended estimating equation is defined in the following way,

Definition 1.2.1 (Extended Estimating Equations) *Let $X_1, \ldots, X_n$ be random vectors as in (1.1). Let $Q(x, \beta) : \mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}$, $k(x, y, \beta) : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}^+$ and $C(x) : \mathbb{R}^d \mapsto \mathbb{R}^p$ be measurable functions. Assume that*

$$E_{\beta,\nu}\left\{m_n(X_1, \boldsymbol{X}, \beta)\right\} = 0, \quad \textit{for all} \qquad \theta = (\beta, \nu), \quad \nu \in \mathcal{H}. \tag{1.9}$$

where β is generally a proper subvector of the whole unknown parameter θ. If there exists a unique solution $\hat{\beta}_n$ to

$$S_n(\beta) = \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta) = 0, \qquad \beta \in U_{\beta_0} \tag{1.10}$$

then (1.10) is the extended estimating equation.

In Definition 1.2.1, it is required that the estimating function $m_n(X_1, \boldsymbol{X}, \beta_0)$ has mean zero, which is the same assumption as in the classic estimating equation. For the specific example of Cox model and AFT, this assumption will be verified through a martingale property. We will discuss this in detail in Chapter 3. In order to ensure the existence of $\hat{\beta}_n$ and its consistency and asymptotic normality, more assumptions need to be made. For example, the quotient term in (1.8) cannot explode to infinity, as n goes to infinity, and in order to have asymptotic normality, there should exist a neighbourhood of the β_0 such that for all β in this neighbourhood, $\partial m_n(X_i, \boldsymbol{X}, \beta)/\partial\beta$ cannot be singular. We include these series of assumptions in Chapter 4.

In this thesis, we will also be applying CEE's and EEE's in an EL estimation framework. Rather than constructing the EL confidence regions, in this thesis, EL is primarily an approach to computing estimators although it could readily be further developed for its more common use in determining confidence regions. We will give self-contained definitions and proofs of EL constructions in Chapter 4 and 5 using empirical process theory. In Chapter 6, we discuss the asymptotic normality of the two estimators associated with the EEE, namely, the one given by directly solving the EEE, and the one by maximizing the EL, which share the identical sandwich

formed variance covariance matrices under regularity conditions.

One thing that we would like to point out is that the definition of EEE is different from the term "martingale estimating equation" in literature like [4] by Bibby et al, [29] by Merkouris, or [20] by Hwang et al, whose primary interest were to construct an estimating function that estimate structural and nuisance parameters simultaneously.

To modify the CEE estimator and confidence regions defined through the EL method to become applicable to right censored semiparametric models, Hjort, Mckeague and van Keilegom proposed a "plug-in" method in [19], i.e., using the empirical estimator of the nuisance parameter. They also showed the slower than $\sqrt{n}$ -rate of convergence, and settings with large numbers of estimating equations compared to the sample size. In [46] and [48], Zhou extended the Wilks type confidence region in [32] for right censored data by replacing the unknown survival function of right censoring variable with the Kaplan-Meier estimator. However, after replacing the unknown nuisance parameter with its empirical estimator, the estimating functions no longer satisfy the definition of CEE.

We develop a broader class of semiparametric models for which EEE definition holds, i.e., the φ-*Transformation Models* with right censored data. These models can be considered as generalizations of the GTM in the sense of characterizing the relation between the dependent variable and covariates by a known function φ, where φ depend on the covariates, structural parameter and error with unknown

hazard rate, and φ is monotone with respect to the error term. With different choices of φ, the φ-transformation models include a series of semiparametric models, including the AFT, the linear model in [34] and [38], and a nonlinear regression model in [39] that allows location and shape to change based according to covariates. The transformation function φ can also have a non-analytical form. For example, it can be defined as an integral of a given wear-out rate function of structural parameter and covariate. Such models can be found in [10] and [3]. Following Tsiatis's work in [38], by constructing a martingale structure on the residual, we show that for the φ-transformation model, the estimating equation that defines an estimator of the structural parameter alone satisfies the definition of EEE. We also prove the EEE associated with the φ-transformation model is asymptotically linear, which together with the martingale *Central Limit Theorem* (CLT) implies the asymptotic normality of the structural estimator.

Similar to CEE, we can also use EEE as a constraint in the EL method, then construct a structural estimator via maximizing the pEL. Under the CEE setting, an element in Owen's, or Qin and Lawless's work in [32] and [33] is specifying a neighbourhood of the structural parameter, in which there exists a unique probability vector that maximizes the EL with probability approaching 1. Therefore we begin, in Chapter 5, by constructing such a neighbourhood for continuous estimating functions under CEE setting, and then generalize the conclusion to discontinuous estimating functions. Finally, we show that for EEE, when the estimating function is continuous with respect to the structural parameter, then the local uniqueness

of solutions to EL maximization is also guaranteed with probability approaching 1; when the criterion functions is discontinuous, with the martingale structure as described in the φ-Transformation model or the Cox model, then the same conclusion can be drawn. Note that the martingale assumption is satisfied by all the EEE examples that we know up to know, including the φ-transformation model and the Cox model.

After proving lemmas that are parallel to Owen, and Qin and Lawless in [32] and [33] using empirical process theory and some classic examples of Donsker class and Glivenko-Cantelli Class listed in [41], we are able to establish the asymptotic normality of the structural point estimator, of which the asymptotic variance covariance matrix is identical to the sandwich-formed asymptotic variance covariance matrix of the corresponding Z-estimator.

Finally, we validate the EEE theory by simulation under the AFT model assumption with R (3.4.1). Since the criterion function corresponding to the EEE of AFT is discontinuous due to the appearance of the indicator function, the EEE may not have a root. Therefore instead, we define the Z-estimator as the value minimizes the Euclidean norm of the EEE. To calculate the maximum empirical likelihood estimator, we first construct the Lagrange multiplier in the maximization problem as a function of structural parameter, then calculate the structural parameter by maximizing the pEL.

Despite the lack of continuity of EEE under AFT model assumption, we proved that a unique solution of the EL maximization always exists with probability approaching 1. The intuition is that the pEL function, or equivalently, the summation of the negative logarithm of the pEL appears very smooth for large n. We show this conjecture by plotting the projection of the pEL in randomly generated unit directions. From the pictures, we can see that for a "moderate" sample size depending on the censoring rate and the skewness of covariates, the plots are always very smooth and almost parabolic around the location of the maximum pEL estimator. On the other hand, plots of the projections of the EEE have many jumps even around the true structural parameter value. We also develop a quantitative way to compare the continuity pattern of around the pEL estimator and around the true structural parameter, and we found the pEL acquires a very similar pattern to the true parameter value.

From simulation, we can also see that for a moderate sample size, the EEE estimator and pEL estimator are very close measured by the L^1 distance between the two. The differences between their variance covariance matrices is also very small, evidenced by small magnitude of eigenvalues. Heuristically, this is because the Lagrange multiplier associated with pEL has a very small magnitude, which makes the constraint function for pEL "almost the same" as the EEE.

Finally, we check the asymptotic normality of the pEL estimator under three different parametric settings with right censored data, including the non-normally

distributed, the normally distributed and the severely skewed distributed covariates. We found it takes a larger sample size for pEL estimator to reach normality if the corresponding covariates is not normally distributed, and the sample size needs to be even larger as the skewness of the covariate grows.

Chapter 2: Non-Existence of Classic Estimating Equations

Cox model and *Accelerated Failure Time* (AFT) model are two semiparametric models that have been extensively used in survival analysis, especially when data may be subject to various types of censoring. Usual ways of constructing estimators for the structural parameters for Cox and AFT involve solving equations as mentioned in [13], [38], and [41]. However, as far as we see from the literature, the equations through which an estimator for the structural parameter is constructed do not satisfy the definition of *Classic Estimating Equations* (CEE). Therefore a natural question is whether a CEE method exist for these semiparametric models.

Despite many advantages of the CEE we discussed on page 1, under some circumstances CEE for a subvector β of parameters may not exist. In this chapter, we prove the nonexistence of the classic EE under the Cox model and randomly right-censored AFT.

More specifically, let us consider a statistical model $\{P_\theta, \theta \in \Theta\}$, where $\theta = (\beta, \lambda)$ consists of a structural parameter $\beta \in \mathbb{R}^p$ and a nuisance parameter $\lambda \in \mathcal{H}$ with infinite dimension, where

$$\mathcal{H} \equiv \left\{ \lambda(t) : \lambda(t) > 0 \text{ a.e. in } t, \int_0^\infty \lambda(t)dt = \infty \right\}. \tag{2.1}$$

Let β_0 and λ_0 denote the true parameter. The question this chapter aims to answer in some special cases is whether there exists a CEE that depends only on data and the structural parameter, i.e., we would like to know if there exists a function $m(X, \beta) : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^p$, such that for any λ_0 and β_0

$$E_{\beta_0,\lambda_0} \{m(X, \beta_0)\} = 0, \tag{2.2}$$

and

$$\nabla_\beta E_{\beta_0,\lambda_0} \{m(X, \beta)\} = E_{\beta_0,\lambda_0} \{\nabla_\beta m(X, \beta)\} \text{ is nonsingular for } \beta \in U_{\beta_0}, \tag{2.3}$$

under assumptions that the passage of ∇_β inside $E\{\cdot\}$ is allowed, where U_{β_0} is a bounded domain in $\mathbb{R}^p$ that contains a neighbourhood of β_0. We show that an estimating function $m(X, \beta)$ satisfying (2.2) and (2.3) and the following regularity conditions does not exist for Cox model in Section 2.1, and for censored AFT in Section 2.2.

Assumptions

(A.1) Z is supported on a bounded set $\mathcal{Z} \subset \mathbb{R}^p$.

(A.2) $\nabla_\beta \int m(t, z, \beta)p(z)dz = \int \nabla_\beta m(t, z, \beta)p(z)dz$, for all $\beta \in U_{\beta_0}$ and $t \in \mathbb{R}^+$.

(A.3) For all $h \in \mathcal{H}$ and sufficiently small ε,

$$\iint m(t, z, \beta)e^{\beta^{tr} z}h(t)e^{-e^{\beta^{tr} z}(1-\varepsilon)H(t)}p(z)dzdt < \infty. \tag{2.4}$$

(A.4) $P\{T < C\} > \delta > 0$.

(A.5) $E_{\beta,\lambda}\{\|m(T, Z, \beta)\|\} < \infty$, for any $\beta \in U_{\beta_0}$ and $\lambda \in \mathcal{H}$.

Remark 2.0.1 *Given assumption (A.5), assumption (A.3) is satisfied for all $h \in \mathcal{H}$. This is because for any $h \in \mathcal{H}$ and $\varepsilon \in (0,1)$, since $1-\varepsilon > 0$, $\int (1-\varepsilon)h(t)dt = (1-\varepsilon)\int h(t)dt = \infty$, therefore $h_\varepsilon(t) = (1-\varepsilon)h(t) \in \mathcal{H}$. Then*

$$\begin{aligned} E_{\beta,h_\varepsilon}\|m(T,Z,\beta)\| &= \iint \|m(t,z,\beta)\| e^{\beta^{tr}z} h_\varepsilon(t) e^{-e^{\beta^{tr}z}H_\varepsilon(t)} p(z)dzdt \\ &= (1-\varepsilon)\iint m(t,z,\beta) e^{\beta^{tr}z} h(t) e^{-e^{\beta^{tr}z}(1-\varepsilon)H(t)} p(z)dzdt \end{aligned} \tag{2.5}$$

Under assumption (A.5), the left hand side of (2.5) is bounded, therefore so is the right hand side, which is what (A.3) states.

2.1 Non-Existence of Classic Estimating Equations of Cox Model

As proposed in [11], the Cox model assumes that the conditional hazard rate of lifetime T given covariate Z is proportional to an unknown baseline hazard rate function $\lambda(t)$, i.e.,

$$\lambda_{T|Z}(t|z) = e^{\beta^{tr}z}\lambda(t), \tag{2.6}$$

where $\beta \in \mathbb{R}^p$ is the regression coefficient to be estimated, and Z is the covariate with density function $p_Z(z)$. The Cox model (2.6) is semiparametric, with parameter

$$\theta = (\beta, \lambda) \in \Theta = U_{\beta_0} \times \mathcal{H}, \tag{2.7}$$

where U_{β_0} is an open and bounded subset of $\mathbb{R}^p$ that contains the true parameter value β_0, and $\mathcal{H}$ is defined in (2.1).

Different from the Kaplan-Meier estimator that primarily constructs an estimator for the nuisance parameter, with the incorporation of the regression-like argument $e^{\beta^{tr}z}$, the Cox model (2.6) gives a way of estimating the structural parameter under a

semiparametric assumption. Cox introduced the notion of partial likelihood in [12], and $\hat{\beta}_n$, the maximum partial likelihood estimator for β_0, is defined as the solution to

$$\sum_{i=1}^{n} \Delta_i \left\{ Z_i - \frac{\sum_{j=1}^{n} Z_j e^{\beta^{tr} Z_j} I\left\{V_j \geq V_i\right\}}{\sum_{j=1}^{n} e^{\beta^{tr} Z_j} I\left\{V_j \geq V_i\right\}} \right\} = 0 \tag{2.8}$$

Andersen and Gill derived the large-sample theoretical properties of $\hat{\beta}_n$ in [2]. It can be shown that $\hat{\beta}_n$ constructed via solving equation (2.8) is a semiparametric efficient estimator. Details of semiparametric efficiency can be found in [39]. However, equation (2.8) does not satisfy the definition of classic EE because the summands are not *independently identically distributed* (i.i.d.). In fact, in this section, we will show that under regularity conditions, a CEE does not exist for the Cox model.

Under the Cox model assumption, the survival function of T given Z is

$$S_{T|Z}(t) = \exp\left\{ -e^{\beta^{tr} z} \int_0^t \lambda(s) ds \right\},$$

and the density function of T is

$$f(t) = \int e^{\beta^{tr} z} e^{-e^{\beta^{tr} z} \Lambda(t)} \lambda(t) p_Z(z) dz, \tag{2.9}$$

where $\Lambda(t) = \int_0^t \lambda(s) ds$ for $t \geq 0$. We would like to see whether a function

$$m(t, z, \beta) : \mathbb{R}^+ \times \mathcal{Z} \times U_{\beta_0} \to \mathbb{R}^p$$

of lifetime, covariate and structural parameter alone can exist, which satisfies for all $\theta \in \Theta$ and all $\lambda \in \mathcal{H}$,

$$E_{\beta,\lambda}\left\{m(T, Z, \beta)\right\} = \iint m(t, z, \beta) e^{\beta^{tr} z - e^{\beta^{tr} z} \Lambda(t)} \lambda(t) p_Z(z) dz dt \equiv 0. \tag{2.10}$$

By the dominated convergence theorem applied to difference quotients with respect to ε, for each $h \in \mathcal{H}$, with g ranging freely over functions bounded by 1, we have the following theorem

Theorem 2.1.1 *For the Cox model given in (2.6), there does not exist a function $m(t, z, \beta)$ of data $X = (T, Z)$ supported on $\mathbb{R}^+ \times \mathcal{Z}$ and structural parameter $\beta \in U_{\beta_0}$ alone, which at the same time satisfies assumptions (A.1)-(A.5) and equations (2.2) and (2.3).*

The proof of this theorem can be found in Section 2.3. In the next section, we discuss the non-existence of CEE for the AFT model.

2.2 Non-Existence of Classic Estimating Equations of Censored AFT

AFT assumes that conditionally given covariate Z, the lifetime T is proportional to some baseline lifetime T_0, i.e.,

$$T = e^{-\beta^{tr} z} T_0, \tag{2.11}$$

where T_0 is a lifetime with unknown hazard rate function $\lambda(t)$, $\beta \in \mathbb{R}^P$ is the regression coefficient to be estimated, and Z is the covariate with density function $p_Z(z)$. Assume that Z is supported on a bounded subset $\mathcal{Z} \in \mathbb{R}^p$, hence (2.11) forms a semiparametric model with parameter

$$\theta = (\beta, \lambda) \in \Theta = U_{\beta_0} \times \mathcal{H}, \tag{2.12}$$

where U_{β_0} is an open and bounded subset of $\mathbb{R}^p$ that contains the true parameter value β_0, and

$$\mathcal{H} \equiv \left\{ \lambda(t) : \lambda(t) \geq 0 \text{ for all } t, \int_0^\infty \lambda(t)dt = \infty \right\} \tag{2.13}$$

is an infinite dimensional space. The survival function of T given Z is given by

$$S_{T|Z}(t) = \exp\left\{ -\Lambda(te^{\beta^{tr} z}) \right\}$$

and the density function of T is given by

$$f(t) = \int e^{\beta^{tr} z - \Lambda(te^{\beta^{tr} z})} \lambda(te^{\beta^{tr} z}) p_Z(z) dz \tag{2.14}$$

There are two different constructions of the regression parameter β depending on whether β has an intercept term.

Case 1 β does not have an intercept term. In this case, λ is unrestricted. We do not know any existing CEE based on β and data alone. There probably is no such estimating equation even though we have not proved this.

Case 2 β has an intercept term. Without loss of generality, assume that the expected value of T_0 is 1, that is

$$E\left\{T_0\right\} = E\left\{ e^{\beta^{tr} Z} T \;\middle|\; Z \right\} = \int t\lambda(t) e^{-\Lambda(t)} dt \equiv 1 \tag{2.15}$$

Therefore for any $p(z)dz$ integrable functions $a(z) : \mathbb{R}^p \to \mathbb{R}^p$ and $b(z) : \mathbb{R}^p \to \mathbb{R}^p$ such that $E\left\{a(Z)\right\} = E\left\{b(Z)\right\} \neq 0$,

$$m(T, Z, \beta) = Te^{\beta^{tr} Z} a(Z) - b(Z) \tag{2.16}$$

has mean zero and is an estimating function if there exists U_{β_0} such that $\nabla_\beta m(T, Z, \beta)$ is nonsingular for all $\beta \in U_{\beta_0}$. For example, $a(z)$ can be a linear function of Z, i.e.,

$$a(z) = Az, \quad \text{where } A = E\left\{b(Z)\right\} \cdot E\left\{Te^{\beta^{tr}Z}Z\right\}^{-1}$$

is a $p \times p$ matrix.

The discussion in previous paragraphs shows we have different conclusions for the non-existence of CEE when lifetime T is always observable. Next, we introduce the concept of censoring, which is a commonly seen situation in practice. When lifetimes are subject to right censoring, instead of observing T and Z, we observe

$$X = (V, \Delta, Z), \quad \text{where } V = \min(T, C),\ \Delta = I\left\{T \leq C\right\} \tag{2.17}$$

and C is the right censoring variable with hazard function Λ_C and hazard rate λ_C. Koul, Susarla and Van Ryzin proposed a classic estimating equation for β in [25] under the assumptions

(K1): C is independent of (T, Z);

(K2): The survival function of C is known and is denoted by $S_C(c)$

Actually, assumption (K2) is seldom reasonable as a modeling assumption, unless function S_C is estimated from another source or censoring is purely "administrative", i.e., occurs when the study observation period ends and the pattern of times of entry into the survival study does not depend on covariates or survival-time. When the distribution function of C is unknown, Koul, Susarla and Van Ryzin proposed to use the Kaplan-Meier estimator instead.

However, the estimating function used to construct the estimator for the structural parameter is does not satisfy the definition of CEE due to the presence of the Kaplan-Meier estimator, because the summands in it are no longer independent. Similarly, neither the general "plug-in" method discussed by Hjort, Mckeague and van Keilegom in [19], or the method proposed by Zhou in Chapter7 of [48] uses estimating functions that satisfies the definition of CEE defined in Chapter 1. In fact, in this section, we show that under regularity conditions, if the expectation of the estimating function is identical to zero when lifetime is arbitrarily right censored, for al possible model parameters, then a CEE does not exist.

Let C be the right censoring variable that is independent of the (T, Z), and let

$$m(t, z, \beta) = \begin{cases} m^1(t, z, \beta) & \text{when } \Delta = 1; \\ \\ m^0(t, z, \beta) & \text{when } \Delta = 0, \end{cases} \tag{2.18}$$

We show that when $m^0(t, z, \beta) \equiv 0$ for all $t \in \mathbb{R}^+$, $z \in \mathcal{Z}$ and $\beta \in U_{\beta_0}$, then a non-trivial CEE does not exist. Note that under most right-censoring CEE formulations, the equations through which an estimator for the structural parameter is defined are usually in the form of $\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)$, where $X_i = (V_i, Z_i, \Delta_i)$ and

$$m_n(X_i, \boldsymbol{X}, \beta) = \Delta_i \cdot m_n^*(X_i, \boldsymbol{X}, \beta), \tag{2.19}$$

and the assumption $m^0(t, z, \beta) \equiv 0$ for all $t \in \mathbb{R}^+$, $z \in \mathcal{Z}$ is satisfied by (2.19).

Estimating function $m(T, Z, \beta)$ must be mean zero, which implies for all hazard functions $\Lambda_C(c) = \int_0^c \lambda_C(s)\, ds \in \mathcal{H}$, $\lambda \in \mathcal{H}$ and $\beta \in U_{\beta_0}$.

$$
\begin{aligned}
0 \equiv & \iiint_0^c m^1(t, z, \beta) e^{\beta^{tr} z - \Lambda(te^{\beta^{tr} z}) - \Lambda_C(c)} \lambda_C(c) \lambda(te^{\beta^{tr} z}) p(z)\, dt\, dz\, dc \\
& + \iiint_0^t m^0(t, z, \beta) e^{\beta^{tr} z - \Lambda(te^{\beta^{tr} z}) - \Lambda_C(c)} \lambda_C(c) \lambda(te^{\beta^{tr} z}) p(z)\, dc\, dz\, dt \\
= & \iint m^1(t, z, \beta) \lambda(te^{\beta^{tr} z}) e^{\beta^{tr} z - \Lambda(te^{\beta^{tr} z}) - \Lambda_C(t)} p(z)\, dt\, dz \\
& + \iint m^0(c, z, \beta) \lambda_C(c) e^{-\Lambda(ce^{\beta^{tr} z}) - \Lambda_C(c)} p(z)\, dc\, dz,
\end{aligned}
\tag{2.20}
$$

Now, let us present the main theorem of this section

Theorem 2.2.1 *Under assumptions (A.1), (A.2), (A.4), and (A.5), for AFT given in (2.11) with arbitrarily right censored data (2.17),*

(a) *A CEE does not exist when assuming*

$$m^0(c, z, \beta) \equiv 0 \qquad \textit{for all } c \in \mathbb{R}^+,\ z \in \mathcal{Z} \textit{ and } \beta \in U_{\beta_0}. \tag{2.21}$$

(b) *A CEE does not exist when assuming*

$$m^1(c, z, \beta) \equiv 0 \qquad \textit{for all } c \in \mathbb{R}^+,\ z \in \mathcal{Z} \textit{ and } \beta \in U_{\beta_0}. \tag{2.22}$$

The conclusion described in (a) is the main result we would like to present. It can be shown by proving

$$\beta_0^{tr} E\left\{\nabla_\beta m^1(T, Z, \beta_0)\right\} = 0,$$

which violates the non-singularity assumption of the gradient described in (2.3).

2.3 Some Proofs

In this section, for simplicity, the ranges of integrations for u and t are equal to $(0, \infty)$ unless otherwise specified. Lemma 2.3.1 gives an important identity that will be used to derive the non-existence of CEE under Cox model assumption, and Lemma 2.3.2 shows how the non-existence conclusion can be drawn without imposing smoothness assumptions on $m_n(\cdot, z, \beta)$.

Lemma 2.3.1 *Under assumption (A.1) - (A.6), for a.e. $t \in \mathbb{R}^+$ and all $\beta \in U_{\beta_0}$,*

$$\int m(t, z, \beta) e^{\beta^{tr} z - e^{\beta^{tr} z} H(t)} p(z) dz = \int_t^\infty \left[\int m(u, z, \beta) e^{2\beta^{tr} z - e^{\beta^{tr} z} H(u)} h(u) p(z) dz \right] du \tag{2.23}$$

Proof To start with, let us consider the following construction of hazard function. Let $h(t) \in \mathcal{H}$ be a candidate baseline hazard function, and

$$H(t) = \int_0^t h(s) ds.$$

Suppose that the baseline hazard rate in (2.6) is of the form

$$\lambda(t) = h(t) e^{\varepsilon g(t)},$$

where ε is a positive constant, and $g(t) \in \mathcal{G}$ is a continuous and bounded function, i.e.,

$$\mathcal{G} \equiv \left\{ g(t) : g(t) \text{ is continuous, and } |g(t)| \leq M_g < \infty \text{ for all } t \in \mathbb{R}^+ \right\} \tag{2.24}$$

The boundedness assumption on g guarantees that $\lambda(t)$ integrates to ∞ and consequently is a hazard rate function. This is because if $|g(t)| \leq M$, for all $t \in \mathbb{R}^+$, then

for any $\varepsilon > 0$, $g(t)\varepsilon \geq -\varepsilon M$, therefore for any $t \in \mathbb{R}^+$ and $h \in \mathcal{H}$,

$$\int_0^t h(x)e^{\varepsilon g(x)}dx \geq e^{-\varepsilon M}\int_0^t h(x)dx$$

Since $\int_0^\infty h(t)dt = \infty$, we know that $\int_0^\infty h(t)e^{\varepsilon g(t)}dt = \infty$.

Let us consider

$$\lambda(t) = h(t)e^{\varepsilon g(t)}, \quad \text{where } g \in \mathcal{G}. \tag{2.25}$$

By (2.10), for all $\theta \in \Theta$ and $g \in \mathcal{G}$,

$$E_\theta\left\{m(T,Z,\beta)\right\} = \iint m(t,z,\beta)e^{\beta^{tr}z - e^{\beta^{tr}z}\int_0^t h(s)e^{\varepsilon g(s)}ds}h(t)e^{\varepsilon g(t)}p(z)dtdz \equiv 0, \tag{2.26}$$

Consider the double integral in (2.26) as a function of ε, then for any fixed $h \in \mathcal{H}$, $g \in \mathcal{G}$, $\beta \in \Theta$ and $\varepsilon \geq 0$,

$$r(\varepsilon,g,h) = \iint m(t,z,\beta)e^{\beta^{tr}z - e^{\beta^{tr}z}\int_0^t h(s)e^{\varepsilon g(s)}ds}h(t)e^{\varepsilon g(t)}p(z)dtdz = 0. \tag{2.27}$$

For fixed $g \in \mathcal{G}$ and $h \in \mathcal{H}$, $r(\varepsilon,g,h) = 0$, for any $\varepsilon \geq 0$. Therefore $\nabla_1 r(0,g,h) = 0$. By (2.26)-(2.27) applied with $\varepsilon > 0$ and $\varepsilon = 0$,

$$\begin{aligned}\nabla_1 r(0,g,h) = \lim_{\varepsilon\to 0}\iint & m(t,z,\beta)e^{\beta^{tr}z}h(t)p(z) \\ & \times \frac{1}{\varepsilon}\left\{e^{-e^{\beta^{tr}z}\int_0^t h(s)e^{\varepsilon g(s)}ds}e^{\varepsilon g(t)} - e^{-e^{\beta^{tr}z}H(t)}\right\}dtdz = 0\end{aligned} \tag{2.28}$$

To apply the Dominated Convergence Theorem and pass the limit into the double integral of (2.28), we first re-write the difference quotient in the second line as

$$e^{-e^{\beta^{tr}z}H(t)}\left\{\frac{1}{\varepsilon}\left\{e^{-e^{\beta^{tr}z}\int_0^t h(s)(e^{\varepsilon g(s)}-1)ds + \varepsilon g(t)} - e^{\varepsilon g(t)}\right\} + \frac{1}{\varepsilon}\left\{e^{\varepsilon g(t)} - 1\right\}\right\} \tag{2.29}$$

Note that for any fixed g and $t \in \mathbb{R}^+$, $(e^{\varepsilon g(t)} - 1)/\varepsilon = O(g(t))$ is uniformly bounded. As for the first term in (2.29), it is equal to

$$\frac{e^{\varepsilon g(t)}}{\varepsilon}\left\{e^{-e^{\beta^{tr}z}\varepsilon\int_0^t h(s)\frac{1}{\varepsilon}(e^{\varepsilon g(s)}-1)ds} - 1\right\} = \frac{e^{\varepsilon g(t)}}{\varepsilon}\left\{e^{-e^{\beta^{tr}z}\varepsilon\int_0^t h(s)(g(s)+o(1))ds} - 1\right\} \tag{2.30}$$

Since $g(\cdot)$ is bounded, denote $\int_0^t h(s)(g(s)+o(1))ds = \theta(t)H(t)$, where $|\theta(t)| \leq M_0 \in \mathbb{R}^+$ for all $t \in \mathbb{R}^+$. Therefore the integrand in (2.28) is bounded by

$$\|m(t,z,\beta)\|\, h(t)e^{\beta^{tr}z - e^{\beta^{tr}z}H(t)\{1+\varepsilon\theta(t)\}}p(z),$$

and under assumption (A.3), $\iint m(t,z,\beta)h(t)e^{\beta^{tr}z - e^{\beta^{tr}z}H(t)\{1+\varepsilon\theta(t)\}}p(z)dzdt < \infty$. Therefore by the Dominated Convergence Theorem, the limit in equation (2.28) can be passed into the integral in (2.28). Hence

$$\iint \frac{\partial}{\partial\varepsilon}\left\{m(t,z,\beta)e^{\beta^{tr}z}h(t)p(z)e^{-e^{\beta^{tr}z}\int_0^t h(s)e^{\varepsilon g(s)}ds}e^{\varepsilon g(t)}\right\}\Bigg|_{\varepsilon=0} dtdz = 0, \tag{2.31}$$

for all $\theta \in \Theta$ and $g \in \mathcal{G}$, which implis that

$$\iint m(t,z,\beta)e^{\beta^{tr}z - e^{\beta^{tr}z}H(t)}h(t)p(z)\left\{g(t) - e^{\beta^{tr}z}\int_0^t h(s)g(s)ds\right\}dtdz \equiv 0, \tag{2.32}$$

for all $\theta \in \Theta$ and $g \in \mathcal{G}$.

We can re-write (2.32) by the Fubini Theorem and get for all $\theta \in \Theta$ and $g \in \mathcal{G}$

$$\begin{aligned} 0 \equiv \int g(t)h(t)\Bigg\{ &\int m(t,z,\beta)e^{\beta^{tr}z - e^{\beta^{tr}z}H(t)}p(z)dz \\ &- \int_t^\infty \int m(u,z,\beta)e^{2\beta^{tr} - e^{\beta^{tr}z}H(u)}h(u)p(z)dzdu\Bigg\}dt \end{aligned} \tag{2.33}$$

Equation (2.33) implies that for a.e. $t \in \mathbb{R}^+$ and all $\beta \in U_{\beta_0}$,

$$\int m(t,z,\beta)e^{\beta^{tr}z - e^{\beta^{tr}z}H(t)}p(z)dz = \int_t^\infty \left[\int m(u,z,\beta)e^{2\beta^{tr}z - e^{\beta^{tr}z}H(u)}h(u)p(z)dz\right]du, \tag{2.34}$$

which is the assertion of the Lemma. □

Lemma 2.3.2 *Under assumptions (A.1) - (A.5),*

$$\nabla_t \int m(t,z,\beta)e^{\beta^{tr}z - e^{\beta^{tr}z}s}p(z)dz \equiv 0, \quad \textit{for all } s,t \in \mathbb{R}^+ \textit{ and } \beta \in U_{\beta_0}. \tag{2.35}$$

Proof For fixed $\beta \in U_{\beta_0}$ and $h \in \mathcal{H}$, denote

$$f(t) = \int m(t, z, \beta) e^{\beta^{tr} z - e^{\beta^{tr} z} H(t)} p(z) dz.$$

Then by (2.34) $f(t)$ is differentiable for all $t \in \mathbb{R}^+$ and $h \in \mathcal{H}$. Differentiating both sides of (2.34), for all $\beta \in U_{\beta_0}$ and $h \in \mathcal{H}$,

$$f'(t) = -\int m(t, z, \beta) e^{2\beta^{tr} z - e^{\beta^{tr} z} H(t)} h(t) p(z) dz \tag{2.36}$$

By the definition of the left hand side of (2.36)

$$\begin{aligned} f'(t) = \lim_{\delta \to 0} \frac{1}{\delta} \Bigg\{ & \int m(t+\delta, z, \beta) e^{\beta^{tr} z - e^{\beta^{tr} z} H(t+\delta)} \, p(z) \, dz \\ & - \int m(t, z, \beta) e^{\beta^{tr} z - e^{\beta^{tr} z} H(t)} \, p(z) \, dz \Bigg\} \end{aligned} \tag{2.37}$$

Rewrite the difference quotient in (2.37) as

$$\begin{aligned} & \frac{1}{\delta} \int m(t, z, \beta) \left\{ e^{\beta^{tr} z - e^{\beta^{tr} z} H(t+\delta)} - e^{\beta^{tr} z - e^{\beta^{tr} z} H(t)} \right\} p(z) \, dz \\ & \qquad + \frac{1}{\delta} \int e^{\beta^{tr} z - e^{\beta^{tr} z} H(t+\delta)} \left\{ m(t+\delta, z, \beta) - m(t, z, \beta) \right\} p(z) \, dz \end{aligned} \tag{2.38}$$

Next, we discuss the two lines in (2.38). The uniform boundedness of h leads to $\sup_t |H(t+\delta) - H(t)| \leq M\varepsilon < \infty$ for $|\delta| < \varepsilon < \infty$. So by the Dominated Convergence Theorem, the first term of (2.38) converges to

$$\begin{aligned} & \int m(t, z, \beta) \nabla_t \left\{ e^{\beta^{tr} z - e^{\beta^{tr} z} H(t)} \right\} p(z) \, dz \\ = & - \int m(t, z, \beta) e^{2\beta^{tr} z - e^{\beta^{tr} z} H(t)} h(t) \, p(z) \, dz. \end{aligned} \tag{2.39}$$

From (2.36), (2.38), and (2.39), we know that as $\delta \to 0$, the second line of (2.38) converges to zero for all $\beta \in U_{\beta_0}$, i.e.,

$$\lim_{\delta \to 0} \frac{1}{\delta} \int \left\{ m(t+\delta, z, \beta) - m(t, z, \beta) \right\} e^{\beta^{tr} z - e^{\beta^{tr} z} H(t+\delta)} \, p(z) \, dz \equiv 0 \tag{2.40}$$

Write the left hand side of (2.40) as the sum of

$$\frac{1}{\delta}\int \{m(t+\delta,z,\beta)-m(t,z,\beta)\}\, e^{\beta^{tr}z-e^{\beta^{tr}z}H(t)}\, p(z)\, dz \tag{2.40.a}$$

and

$$\frac{1}{\delta}\int \{m(t+\delta,z,\beta)-m(t,z,\beta)\}\left\{e^{\beta^{tr}z-e^{\beta^{tr}z}H(t+\delta)}-e^{\beta^{tr}z-e^{\beta^{tr}z}H(t)}\right\} p(z)\, dz \tag{2.40.b}$$

As $\delta \to 0$, (2.40.a) converges to $\nabla_t \left\{\int m(t,z,\beta)e^{\beta^{tr}z-e^{\beta^{tr}z}s}p(z)dz\right\}\Big|_{s=H(t)}$, which is the left hand side of (2.35) when $H(t)$ is regarded as a free variable $s \in \mathbb{R}^+$. Therefore it suffices to prove that as $\delta \to 0$, (2.40.b) converges to 0 for all $\beta \in U_{\beta_0}$ and $h \in \mathcal{H}^*$, where $\mathcal{H}^*$ is also a dense class of functions in L^1 defined as

$$\begin{aligned}\mathcal{H}^* = \Big\{ & h : h \in \mathcal{H};\ \text{for each } h, \text{ there exists } \varepsilon > 0 \text{ such that} \\ & h(t) = 0 \text{ for all } t \in (0,\varepsilon); \\ & h \text{ bounded above and below, for all } t \geq \varepsilon; \\ & h \text{ bounded above, for all } t \geq \varepsilon \Big\}\end{aligned} \tag{2.41}$$

Then consider $t > \delta$ and write (2.40.b) as

$$\begin{aligned}&\int e^{\beta^{tr}z}p(z)\left\{m(t+\delta,z,\beta)e^{-e^{\beta^{tr}z}H(t+\delta)}-m(t,z,\beta)e^{-e^{\beta^{tr}z}H(t)}\right\} dz \\ -&\int e^{\beta^{tr}z}p(z)m(t,z,\beta)\left\{e^{-e^{\beta^{tr}z}H(t+\delta)}-e^{-e^{\beta^{tr}z}H(t)}\right\} dz \\ -&\int e^{\beta^{tr}z}p(z)e^{-e^{\beta^{tr}z}H(t)}\{m(t+\delta,z,\beta)-m(t,z,\beta)\}\, dz\end{aligned} \tag{2.42}$$

Then using the identity equation (2.34) proved in Lemma 2.3.1 and the additivity property, $-\int_{t+\delta}^{\infty}\ell(s)\,ds = \int_t^{t+\delta}\ell(s)\,ds - \int_t^{\infty}\ell(s)\,ds$ for any integrable function $\ell(\cdot)$,

we re-write (2.42) as the sum of

$$C_1 = \int_t^{t+\delta} \int m(u,z,\beta) e^{2\beta' z} \Big[h(u-\delta) \exp(-e^{\beta' z} H(u-\delta)) \\ - h(u) \exp(-e^{\beta' z} H(u)) \Big] p(z) dz du \tag{2.43}$$

and

$$C_2 = \int_t^{\infty} \int m(u,z,\beta) e^{2\beta' z} \Big[2h(u) \exp(-e^{\beta' z} H(u)) \tag{2.44}$$
$$- h(u-\delta) \exp(-e^{\beta' z} H(u-\delta)) - h(u+\delta) \exp(-e^{\beta' z} H(u+\delta)) \Big] p(z) dz du$$

First, by the Mean Value Theorem, the square-bracketed integrand in C_1 is $O(\delta)$, bounded by

$$M^2 \delta \left[|h'(u-\theta\delta)| \, e^{-e^{\beta' z} H(u-\theta\delta)} + h^2(u-\theta\delta) e^{-e^{\beta' z} H(u-\theta\delta)} \right] \tag{2.45}$$

where M^2 is a uniform upper bound for $e^{\beta^{tr} z}$, $h \in \mathcal{H}^*$, and $\theta \in (0,1)$. Let k be a function and C_1^* be a constant such that for all $\delta \in (0,\varepsilon)$ the lower bound of $C_1^* k(u)$ is given by

$$\begin{cases} \text{a constant } M_1 \in \mathbb{R} & \text{when } u < \varepsilon \\ \sup_{\theta \in [-1,1]} \left(h'(u-\theta\delta), h^2(u-\theta\delta) \right) & \text{when } u > \varepsilon. \end{cases} \tag{2.46}$$

Let $K(u) \geq H(u-\varepsilon)$ for all $u \geq \varepsilon$. Then (2.45) is bounded by

$$M^2 \, \delta \, C_1^* \, k(u) \, e^{-e^{\beta' z} K(u)}, \quad \text{for all } u > \varepsilon, \, |\theta| \leq 1, \, 0 < \delta < \varepsilon,$$

So the fact that the outer integral in C_1 is taken over a shrinking interval $(t, t+\delta)$ allows us to say that the integral of the integrand given by δ times $dz\,du$ integrable function

$$\|m(u,z,\beta)\| k(u) e^{2\beta^{tr} z - e^{\beta^{tr} z} K(u)} p(z) = o(\delta).$$

Now we move on to C_2. Using the Mean Value Theorem again, the square-bracketed integrand in C_2 is bounded by

$$\delta^2 \frac{\partial^2}{\partial x^2} \left(h(x) \exp(-e^{\beta' z} H(x)\right) \Big|_{x=u+\theta^*\delta} \tag{2.47}$$

where $h \in \mathcal{H}^*$ and $|\theta^*| \leq 1$ but cannot be controlled further. Let k be a function and C_2^* be a constant such that $C_2^* k(u)$ is lower bounded by

$$\begin{cases} \text{a constant } M_2 \in \mathbb{R} & \text{when } u < \varepsilon \\ \sup_{\theta \in [-1,1]} \left(h''(u+\theta\delta), 2h(u+\theta\delta)h'(u+\theta\delta), h^2(u+\theta\delta)h'(u+\theta\delta)\right) & \text{when } u > \varepsilon. \end{cases} \tag{2.48}$$

for all $0 < \delta < \varepsilon$. Let $K(u) \geq H(u-\varepsilon)$ for all $u \geq \varepsilon$. Then (2.47) is controlled by

$$\delta^2 C_2^* k(u) e^{-e^{\beta^{tr} z} K(u)}, \quad \text{for all } u > \varepsilon,\ |\theta^*| \leq 1,\ 0 < \delta < \varepsilon.$$

Thus , even though the range of integration is now not small, C_2 can be shown (with $h \in \mathcal{H}^*$) to be $O(\delta^2)$. □

Lemma 2.3.3 (Weierstrass) *Suppose f is a continuous real-valued function defined on real interval $[a, b]$. For every $\varepsilon > 0$, there exists a polynomial $p_f(x)$ such that for all x in $[a, b]$, $|f(x) - p_f(x)| < \varepsilon$, or equivalently,* $\sup_{x \in [a,b]} |f(x) - p_f(x)| < \varepsilon$.

2.3.1 Proof of Theorem 2.1.1

Multiply both sides of (2.35) by $a(t) \in \mathcal{A}$, where

$$\begin{aligned} \mathcal{A} \equiv \{a(t) : \mathbb{R} \to \mathbb{R}; a \text{ is compactly supported ,} \\ \text{and continuously differentiable on } \mathbb{R}^+\}, \end{aligned} \tag{2.49}$$

then integrate with respect to t, yielding for all $a \in \mathcal{A}$, $\beta \in U_{\beta_0}$ and $s \in \mathbb{R}^+$,

$$\int a(t) \left\{ \frac{\partial}{\partial t} \int m(t, z, \beta) e^{\beta^{tr} z - e^{\beta^{tr} z} s} p(z) dz \right\} dt \equiv 0. \tag{2.50}$$

Integrate (2.50) by parts, yielding for all $a \in \mathcal{A}$, $\beta \in U_{\beta_0}$ and $s \in \mathbb{R}^+$,

$$\int a'(t) \int m(t, z, \beta) e^{\beta^{tr} z - e^{\beta^{tr} z} s} p(z) dz dt \equiv 0. \tag{2.51}$$

Now, multiply both sides of (2.34) by $a'(t)$, then integrate with respect to t. Together with equation (2.51), we conclude that for all $a \in \mathcal{A}$, $\beta \in U_{\beta_0}$ and $h \in \mathcal{H}$,

$$0 \equiv \int a'(t) \left\{ \int_t^\infty \int m(u, z, \beta) e^{2\beta^{tr} z - e^{\beta^{tr} z} H(u)} h(u) p(z) dz du \right\} dt \tag{2.52}$$

Integrating (2.52) by parts, then we know that for all $a \in \mathcal{A}$, $\beta \in U_{\beta_0}$ and $h \in \mathcal{H}$,

$$0 \equiv \int_0^\infty a(t) \int m(t, z, \beta) e^{2\beta^{tr} z - e^{\beta^{tr} z} H(t)} h(t) p(z) dz dt \tag{2.53}$$

Therefore for all $a \in \mathcal{A}$, $\beta \in U_{\beta_0}$ and s, $t \in \mathbb{R}^+$,

$$0 \equiv \int m(t, z, \beta) e^{2\beta^{tr} z - e^{\beta^{tr} z} s} p(z) dz, \tag{2.54}$$

where we replaced $H(t)$ by $s \in \mathbb{R}^+$ since H is free to be any function in $\mathcal{H}$. Integrating both sides of (2.54) with respect to s from x to ∞ implies

$$0 \equiv \int m(t, z, \beta) e^{\beta^{tr} z - e^{\beta^{tr} z} x} p(z) dz \tag{2.55}$$

After integrating again with respect to x on s, ∞ or (separately) by differentiating equation (2.55) under the integral sign arbitrarily many times, then the same formula (2.55) holds with the first term $e^{\beta^{tr} z}$ term replaced by $e^{k\beta^{tr} z}$ for any non-negative integer k.

Let $b(\cdot)$ be any element of $\mathcal{B}$, the continuous functions on the real line and apply the Weierstrass' theorem stated in Lemma 2.3.3 to approximate b by polynomials uniformly on the compact set of possible $\beta^{tr}z$ values, yielding

$$0 \equiv \int m(t,z,\beta)b(\beta^{tr}z)p(z)dz, \quad \text{for all } t \in \mathbb{R}^+, \beta \in U_{\beta_0}, b \in \mathcal{B}. \tag{2.56}$$

Now, let us demonstrate the singularity of the matrix $E_{\beta_0,\lambda_0}\{\nabla_\beta m(t,z,\beta_0)\}$, which is a violation of (2.3) with $\beta = \beta_0$ and thus proves the non-existence of the CEE. Differentiate both sides of (2.55) with respect to β^{tr}, i.e., the Jacobian of function $m(t,z,\cdot)$, yielding for all $t, s \in \mathbb{R}^+$ and $\beta \in U_{\beta_0}$,

$$0 \equiv \int \nabla_\beta m(t,z,\beta)e^{\beta^{tr}z - e^{\beta^{tr}z}s}p(z)dz - \int m(t,z,\beta)c_s(\beta^{tr}z)\, z\, p(z)\, dz, \tag{2.57}$$

where $c_s(\beta^{tr}z)z = \nabla_\beta(e^{\beta^{tr}z - e^{\beta^{tr}z}s})$. Multiply both sides of (2.57) by β^{tr} and use equation (2.57) by choosing $c_s(\beta^{tr}z)\beta^{tr}z \in \mathcal{B}$ and then setting $s = H(t)$. Then for all $t \in \mathbb{R}^+$, $h \in \mathcal{H}$ and $\beta \in U_{\beta_0}$

$$0 \equiv \beta^{tr} \int \nabla_\beta m(t,z,\beta)e^{\beta^{tr}z - e^{\beta^{tr}z}H(t)}p(z)dz. \tag{2.58}$$

Equation (2.58) holds for a specific choice of $\beta = \beta_0$, which contradicts the non-sigularity of matrix $\int \nabla_\beta m(t,z,\beta_0)e^{\beta_0^{tr}z - e^{\beta_0^{tr}z}s}p(z)dz$ as described in (2.3). $\square$

2.3.2 Proof of Theorem 2.2.1

Note that (2.20) is true for any hazard rate function λ_C, therefore it is satisfied when C puts its mass at a point c, i.e., for a constant $c \in \mathbb{R}^+$, $\Lambda_C(s) = \Lambda^c(s)$, where

$$\Lambda^c(s) = \begin{cases} 0 & \text{when } s < c, \\ \infty & \text{when } s > c. \end{cases} \tag{2.59}$$

Then (2.20) implies that for any $c \in \mathbb{R}^+, \lambda \in \mathcal{H}$ and β,

$$\begin{aligned} 0 \equiv & \iint_0^c m^1(t,z,\beta)\lambda(te^{\beta^{tr}z})e^{\beta^{tr}z-\Lambda(te^{\beta^{tr}z})}p(z)dtdz \\ & \qquad\qquad + \int m^0(c,z,\beta)e^{-\Lambda(ce^{\beta^{tr}z})}p(z)dz \\ = & \iint_0^{ce^{\beta^{tr}z}} m^1(se^{-\beta^{tr}z},z,\beta)\lambda(s)e^{-\Lambda(s)}p(z)dsdz \\ & \qquad\qquad + \int e^{-\Lambda(ce^{\beta^{tr}z})}m^0(c,z,\beta)p(z)dz, \end{aligned} \tag{2.60}$$

Since (2.60) is satisfied by any $\lambda \in \mathcal{H}$, let $\Lambda(s) = \Lambda^t(s)$, where $\Lambda^t(s)$ is defined as in (2.59). Then (2.60) implies

$$\begin{aligned} 0 \equiv & \int I\left\{t < ce^{\beta^{tr}z}\right\} m^1(te^{-\beta^{tr}z},z,\beta)p(z)dz \\ & + \int I\left\{t > ce^{\beta^{tr}z}\right\} m^0(c,z,\beta)p(z)dz, \qquad \text{for all } c \in \mathbb{R}^+, t \in \mathbb{R}^+, \beta \end{aligned} \tag{2.61}$$

Next, we prove statement (a). Then (b) can be shown with the same strategy.

Proof of (a) In this part, we assume $m^0(c,z,\beta) \equiv 0$ for all $c \in \mathbb{R}^+$, $z \in \mathcal{Z}$ and $\beta \in U_{\beta_0}$. Hence (2.61) becomes

$$0 \equiv \int I\left\{t < ce^{\beta^{tr}z}\right\} m^1(te^{-\beta^{tr}z},z,\beta)p(z)dz, \qquad \text{for all } t \in \mathbb{R}^+, \beta \tag{2.62}$$

Since (2.62) holds for all $t, c \in \mathbb{R}^+$, therefore consider a class of functions $\mathcal{A}$ as follows,

$$\mathcal{A} \equiv \left\{a(t) : a(t) \text{ is essentially bounded }, t \in \mathbb{R}^+\right\}$$

then for any $a(t) \in \mathcal{A}$, β, and $c \in \mathbb{R}^+$, integrating $a(t)$ multiplied by (2.62) against Lebesgue measure on $(0,\infty)$ gives for all $c \in \mathbb{R}^+$, $a \in \mathcal{A}$, $\beta \in U_{\beta_0}$ and $z \in \mathcal{Z}$,

$$0 \equiv \iint_0^{ce^{\beta^{tr}z}} a(t)m^1(te^{-\beta^{tr}z},z,\beta)p(z)dtdz. \tag{2.63}$$

Differentiate (2.63) with respect to c, yielding for all $a \in \mathcal{A}$, $\beta \in U_{\beta_0}$, $z \in \mathcal{Z}$ and a.e. in c,

$$0 \equiv \int e^{\beta^{tr} z} a(ce^{\beta^{tr} z}) m^1(c, z, \beta) p(z) dz. \tag{2.64}$$

By specific series of choices of $a(ce^{\beta^{tr} z}) = e^{\beta^{tr} z \cdot k} I\left\{ce^{\beta^{tr} z} \geq \tau\right\}$ for fixed $\tau, c \in \mathbb{R}^+$ and $k \in \mathbb{Z}^+$, we know that (2.64) implies for all $k \in \mathbb{Z}^+$ and $\tau \in \mathbb{R}^+$,

$$0 = \int e^{\beta^{tr} z \cdot k} I\left\{ce^{\beta^{tr} z} \geq \tau\right\} m^1(c, z, \beta) p(z) dz \tag{2.65}$$

Let $\tau \to 0$, for all $a \in \mathcal{A}$, $\beta \in U_{\beta_0}$, $k \in \mathbb{Z}^+$, $z \in \mathcal{Z}$ and $c \in \mathbb{R}^+$

$$0 = \int e^{\beta^{tr} z \cdot k} m^1(c, z, \beta) p(z) dz \tag{2.66}$$

By Weierstrass' theorem described in Lemma 2.3.3, for $b \in \mathcal{B}$, $\beta \in U_{\beta_0}$, $k \in \mathbb{Z}^+$, $z \in \mathcal{Z}$ and $c \in \mathbb{R}^+$,

$$0 = \int b(\beta^{tr} z) m^1(c, z, \beta) p(z) dz, \tag{2.67}$$

where $\mathcal{B}$ is the family of differentiable functions with bounded support.

To show the non existence of CEE via deriving the singularity of

$$E_{\beta_0, \lambda_0}\left\{\nabla_\beta m^1(T, Z, \beta_0)\right\}, \tag{2.68}$$

let us differentiate both sides of (2.67) with respect to β, then for $b \in \mathcal{B}$, $\beta \in U_{\beta_0}$, $k \in \mathbb{Z}^+$, $z \in \mathcal{Z}$ and $c \in \mathbb{R}^+$,

$$\int \beta^{tr} \nabla_\beta m^1(c, z, \beta) b(\beta^{tr} z) p(z) dz = -\int \beta^{tr} z b'(\beta^{tr} z) m^1(c, z, \beta) p(z) dz = 0 \tag{2.69}$$

Choose $b(\beta^{tr} z) = e^{\beta^{tr} z} h(e^{\beta^{tr} z}) e^{-H(e^{\beta^{tr} z} c)}$ and integrate against dc. Then (2.69) implies

$$\beta_0^{tr} E_{\beta_0, \lambda_0}\left\{\nabla_\beta m^1(T, Z, \beta_0)\right\} \equiv 0. \tag{2.70}$$

This contradiction of non-singularity completes the proof of (a).

Proof of (b) In this part, we assume that $m^1(t, z, \beta) \equiv 0$ for all $t \in \mathbb{R}^+$, $z \in \mathcal{Z}$ and $\beta \in U_{\beta_0}$. Hence (2.61) becomes

$$0 \equiv \int I\left\{t > ce^{\beta^{tr} z}\right\} m^0(c, z, \beta) p(z) dz, \quad \text{for all } c \in \mathbb{R}^+, z \in \mathcal{Z}, \beta \in U_{\beta_0} \tag{2.71}$$

In fact, we can show the contradiction of singularity described in (2.70) following the same steps in proving (a) on page 31. □

Chapter 3: φ-Transformation Model

3.1 Introduction

In this section, we discuss the φ-Transformation model, which serves as the most general worked-out example for the *Extended Estimating Equations* (EEE). Let T_i be the lifetime for the ith individual for $i = 1, \ldots, n$, and consider the model

$$T_i = \varphi(\varepsilon, \beta_0, Z_i), \quad i = 1, \ldots, n \tag{3.1}$$

where $Z_i = (Z_{i1}, \ldots Z_{ip})^{tr}$ is a covariate, ε_i is the error that is *independently identically distributed* (i.i.d.) with common differentiable distribution $F(x)$, and hazard rate function $\lambda(x)$, $\beta_0 \in \mathbb{R}^p$ is the unknown coefficient to be estimated, and $\varphi(x, y, z) : \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^+$ is a given function, where $\varphi(\cdot, y, z)$ is strictly increasing and continuously differentiable, and $\varphi(x, \cdot, z)$ is differentiable. We also assume that the support of T and the support of ε do not depend on parameter (β, F).

In practice, sometimes we are not able to observe the complete lifetime T_i. Let us use C_i to denote the right censoring variable, assumed to be identically distributed with a common distribution function denoted by H. Assume that conditionally given Z_i, T_i and C_i are independent. When T_i are subject to right censoring, the

data collected are random vectors

$$(V_i, \Delta_i, Z_i), \quad i = 1, \ldots, n \tag{3.2}$$

where $V_i = \min(T_i, C_i)$, and $\Delta_i = I\{T_i \leq C_i\}$ is equal to 1 when T_i does not exceed the right censoring variable C_i, and 0 otherwise.

The φ-transformation model relates to many other important models in statistics, and we mention the following examples. First, when $\varphi(x, y, z)$ depends on the structural parameter y only through $y^{tr}z$ and is linear in $y^{tr}z$ and x, i.e.,

$$T_i = \beta_0^{tr} Z_i + \varepsilon_i, \tag{3.3}$$

then the φ-transformation model becomes the usual linear model. Nothing needs to be changed when the lifetime T_i is replaced by $h(T_i)$, where $h(\cdot)$ is a known monotone function. When $h(T_i) = \ln T_i$, (3.3) becomes the *Accelerated Failure Time* (AFT) model. Therefore the AFT model is an example of the φ-transformation model. A review of the linear model and the AFT can be found in page 8 of Chapter 1.

Another important class of models that is related to the φ-transformation model is the nonlinear regression model with additive independent errors ([39], Chapter 5), namely,

$$T_i = \mu(Z_i, \beta_0) + \varepsilon_i, \tag{3.4}$$

where $\mu(\cdot, \beta_0)$ is given, T_i is continuous, and ε_i is independent of Z_i. This model assumes that there is a basic underlying distribution for the lifetime, but the location shifts according to covariate Z. In (3.4), $\varphi(\varepsilon, Z, \beta) = \mu(Z, \beta) + \varepsilon$, and one way to

generalize it is to allow a shape change in T_i, i.e.,

$$h(T_i) = b_0(\beta_1^{tr} Z_i) + \varepsilon b_1(\beta_2^{tr} Z_i),$$

where $h\mathbb{R}^+ \mapsto \mathbb{R}$ is a known monotone function, b_0 and b_1 are both known functions, and b_1 is strictly positive. This way, both the location and the shape change according to covariates.

It is not necessary that the transformation function $\varphi(x, y, z)$ has an analytical form. A possible choice is to let $\varphi(x, y, z)$ be an integral of a known positive rate function, i.e., let $b(s, \beta, z) : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+$ be a given function, then

$$h(T_i) = \varphi(\varepsilon, Z_i, \beta_0) = \int_{-\infty}^{\varepsilon} b(s, \beta_0, Z_i) ds. \tag{3.5}$$

The time T_i is known as "operational time", and model (3.5) may have different interpretation in practice depending on the choice of $b(s, \beta_0, z)$. For example, $b(s, \beta_0, z)$ can describe the wearing-out rate of a device. This rate can depend on a structural parameter β_0, and a covariate Z_i that differs from device to device. This type of model is also discussed by Nikulin in [3].

This chapter is organized as follows. In Section 3.2, we establish an EEE for β_0 using martingale theory. In Section 3.3, we prove the consistency and asymptotic normality of the estimator of β_0. Technical lemmas not given in detail in Section 3.3 are postponed to Section 3.4.

3.2 Extended Estimating Equations and Martingales

To construct the EEE that yields an estimator of β_0 for the φ-transformation model (3.1), we follow Tsiatis [38] by building a martingale on the residual scale. Let $N_i^T(v)$ be the counting process for the ith individual for lifetime T_i, and let $Y_i^T(v)$ be the at-risk indicator for lifetime T_i, i.e.,

$$N_i^T(v) = \Delta_i \cdot I\{V_i \leq v\}; \quad Y_i^T(v) = I\{V_i \geq v\} \tag{3.6}$$

Since $\varphi(\cdot, y, z)$ is strictly increasing, we can define its inverse function,

$$\varphi^{-1}(t, y, z) : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R} \quad \text{such that} \quad \varphi(u, y, z) = t \;\Leftrightarrow \varphi^{-1}(t, y, z) = u. \tag{3.7}$$

Let ζ_i be the residual when T_i is censored, i.e., $\zeta_i = \varphi^{-1}(C_i, \beta_0, Z_i)$. Then we can write the counting process and at-risk indicator for residual ε_i as follows,

$$N_i^\varepsilon(u) = I\{\min(\varepsilon_i, \zeta_i) \leq u, \varepsilon_i \leq \zeta_i\} \tag{3.8}$$

Since $\varphi^{-1}(u, y, z)$ is also strictly increasing,

$$\begin{aligned} I\{\min(\varepsilon_i, \zeta_i) \leq u\} &= I\{\min(\varphi(\varepsilon_i, \beta_0, Z_i), \varphi(\zeta_i, \beta_0, Z_i)) \leq \varphi(u, \beta_0, Z_i)\} \\ &= I\{\min(T_i, C_i) \leq \varphi(u, \beta_0, Z_i)\}, \end{aligned} \tag{3.9}$$

and

$$I\{\varepsilon_i \leq \zeta_i\} = I\{\varphi(\varepsilon_i, \beta_0, Z_i) \leq \varphi(\zeta_i, \beta_0, Z_i)\} = I\{T_i \leq C_i\} = \Delta_i, \tag{3.10}$$

From (3.6)-(3.10), we know that

$$N_i^\varepsilon(u) = I\{\min(T_i, C_i) \leq \varphi(u, \beta_0, Z_i), \Delta_i = 1\} = N_i^T(\varphi(u, \beta_0, Z_i)) \tag{3.11}$$

As for the at-risk indicator $Y_i^{\varepsilon}(u)$,

$$\begin{aligned} Y_i^{\varepsilon}(u) &= I\left\{\min(\varepsilon_i, \zeta_i) \geq u\right\} \\ &= I\left\{\min(\varphi(\varepsilon_i, \beta_0, Z_i), \varphi(\zeta_i, \beta_0, Z_i)) \geq \varphi(u, \beta_0, Z_i)\right\} \\ &= I\left\{\min(T_i, C_i) \geq \varphi(u, \beta_0, Z_i)\right\} = Y_i^T\left(\varphi(u, \beta_0, Z_i)\right). \end{aligned} \tag{3.12}$$

Since the counting process and at-risk indicator for ε_i and T_i have the relationship described in (3.11) and (3.12), from now on we will mainly use $N_i^T(\cdot)$ and $Y_i^T(\cdot)$ to construct estimators, and define $N_i(\cdot) = N_i^T(\cdot)$, $Y_i(\cdot) = Y_i^T(\cdot)$ for the sake of simplicity. Assume that V_i's are nondegenerate, then there exists a constant T^* such that for some $\xi > 0$,

$$P\left\{V_i \geq T^* + \xi\right\} \geq \psi > 0, \quad \text{for all } i. \tag{3.13}$$

Let us introduce additional notations to be used throughout this chapter. For a function $g(x_1, x_2, x_3)$, let $\nabla_i g(x_1, x_2, x_3) = \partial g(x_1, x_2, x_3)/\partial x_i$. Define

$$J(u, \beta_1, \beta_2, z) \equiv \varphi^{-1}(\varphi(u, \beta_1, z), \beta_2, z) \tag{3.14}$$

Note that by the chain rule, as shown in Lemma 3.4.1,

$$\gamma(u, \beta_0, z) = \nabla_2 J(u, \beta, \beta_0, z)|_{\beta=\beta_0} = \frac{\nabla_2 \varphi(u, \beta_0, z)}{\nabla_1 \varphi(u, \beta_0, z)}. \tag{3.15}$$

In addition, assume the following regularity conditions:

(A.1) The density function of ε, $f(x) = dF(x)/dx$, exists and is bounded by K_1 on $(-\infty, T^* + \xi]$, where T^*, ξ are as in (3.13). The hazard rate $\lambda(x)$ is twice differentiable.

(A.2) The density function for C_i exists, and $h(x) = -dH(x)/dx \leq K_2$, for all $x \leq T^* + \xi$.

(A.3) There exists $\theta(u, \beta_0, Z)$ with $E|\theta(u, \beta_0, Z)| < \infty$ such that

$$\begin{aligned} &|\lambda(J(u, \beta, \beta_0, Z_i)) - \lambda(u) - \gamma(u, \beta_0, Z_i)^{tr}(\beta - \beta_0)\lambda'(u)| \\ &\qquad \leq \|\beta - \beta_0\|^2 \theta(u, \beta_0, Z_i), \text{ for } \beta \in U_{\beta_0}(n^{-1/2}), \text{ and } u \in \mathbb{R}, \end{aligned} \tag{3.16}$$

almost surely in Z_i, where $U_{\beta_0}(n^{-1/2})$ is a $n^{-1/2}$ neighbourhood of β_0.

(A.4) There exists a constant c such that $P\{\|Z_i\| < c\} = 1$.

(A.5) Let $\mu_{\beta_0}(u, \beta) = E\{Z_1 I\{V_1 > \varphi(u, \beta, Z_1)\}\}/E\{I\{V_1 > \varphi(u, \beta, Z_1)\}\}$. Then

$$\sup_{\beta \in U_{\beta_0}(n^{-1/2}), u \leq T^* + \xi} \|\bar{Z}(u, \beta) - \mu_{\beta_0}(u, \beta)\| \xrightarrow{P} 0, \text{ as } n \to \infty,$$

where

$$\bar{Z}(u, \beta) = \frac{\sum_{j=1}^n Z_j Y_j(\varphi(u, \beta, Z_j))}{\sum_{j=1}^n Y_j(\varphi(u, \beta, Z_j))}. \tag{3.17}$$

(A.6) Let $A_{\beta_0}(u, \beta) \equiv E\left\{I(V_1 > \varphi(u, \beta, Z_1))(Z_1 - \bar{Z}(u, \beta))\gamma(u, \beta_0, Z_1)^{tr}\right\}$, then

$$\sup_{\beta \in U_{\beta_0}(n^{-1/2}), u \leq T^* + \xi} \left| \frac{1}{n} \sum_{i=1}^n Y_i(\varphi(u, \beta, Z_i)) \left\{Z_i - \bar{Z}(u, \beta)\right\} \gamma(u, \beta_0, Z_i)^{tr} - A_{\beta_0}(u, \beta) \right|$$

approaches zero in probability, and for all $\beta \in U_{\beta_0}(n^{-1/2})$

$$\int_{-\infty}^{T^*} \lambda'(u) A_{\beta_0}(u, \beta) du \text{ is non singular.}$$

(A.7) For $\beta \in U_{\beta_0}(n^{-1/2})$, $E\{\Lambda(J(T^*, \beta_0, \beta, Z))\} = M < \infty$.

(A.8) $\varphi(\cdot, y, z)$, $\varphi(x, \cdot, z)$ and $\varphi^{-1}(\cdot, y, z)$ are all Lipschitz continuous; $\varphi(\cdot, y, z)$ is continuously differentiable, and $\nabla_2 \varphi(x, \cdot, z)$ is Lipschitz continuous.

(A.9) $\varphi(\cdot, y, z)$ is strictly increasing.

(A.10) $E\left\|\int_{-\infty}^{T^*} \gamma(u, \beta_0, Z_i)\lambda'(u)du\right\| < \infty$.

3.2.1 Remarks and Sufficient Conditions for Assumptions

First, let us discuss assumption (A.3). Note that the left hand side of (3.16) is the Taylor expansion of $\lambda(J(u, \cdot, \beta_0, Z_i))$ at $\beta = \beta_0$, the remainder of which is $(\beta - \beta_0)^{tr}\partial^2\lambda(J(u, \beta, \beta_0, Z_i))/\partial\beta^2|_{\beta=\beta^*}(\beta - \beta_0)$, for β^*. Also recall that $J(u, \beta_1, \beta_2, z) = \varphi^{-1}(\varphi(u, \beta_1, z), \beta_2)$. Therefore one way to guarantee (3.16) in assumption (A.3) is to assume that

(i) $\partial^2\varphi(u, \beta, z)/\partial\beta^2$ and $\partial^2\varphi^{-1}(u, \beta, z)/\partial\beta^2$ exists

(ii) $\lambda'(u)$ is Lipschitz continuous continuous.

As for the boundedness of $E|\theta(u, \beta_0, Z_i)|$, it is guaranteed if the derivatives in (i) are continuous in z.

In assumptions (A.5) and (A.6), the pointwise convergence for fixed β and u can be obtained using the *Law of Large Numbers* (LLN) . We can show that this convergence uniform in $\beta \in U_{\beta_0}(n^{-1/2})$ and $u \leq T^* + \xi$ using empirical process theory. For example, we can show that

$$\mathcal{F} \equiv f(v, z; u, \beta) : I\left\{v \geq \varphi(u, \beta, z)(z - \bar{Z}(u, \beta))\right\}$$

is a Glivenko-Cantelli class, which can be established under the assumption that function $\varphi(x, y, z)$ is Lipschitz continuous in x and y using the Example 19.11 and Example 19.20 in [41].

The boundedness of the expected value mentioned in (A.7) and (A.10) is guaranteed by the smoothness of $\Lambda(\cdot)$ and $J(u, \beta_0, \cdot, z)$, and the bounded support of Z. Under assumption (A.8), $J(u, \beta_1, \cdot, z)$ is Lipschitz. Since $J(u, \beta_0, \beta_0, z) = 1$, we know that $|J(u, \beta_0, \beta, z)| \leq 1 + c\|\beta - \beta_0\|$, hence (A.7) is satisfied. (A.10) is guaranteed if we assume that $\gamma(u, \beta_0, z)$ is bounded for any u and z.

Define

$$S_n(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{T^*} dN_i\left(\varphi(u, \beta, Z_i)\right) \left\{Z_i - \bar{Z}(u, \beta)\right\}, \tag{3.18}$$

where $\bar{Z}(u, \beta)$ is defined in (3.17). From now on, let us use P denote the probability measure under the true nuisance and structural parameter. Then we construct the martingale in the following proposition.

Proposition 3.2.1 *$M_i\left\{\varphi(u, \beta, Z_i)\right\}$ is a martingale with respect to the filtration*

$$F_n(u, \beta) = \sigma\left(Z_i, I\left\{V_i \leq \varphi(s, \beta, Z_i)\right\}, \Delta_i I\left\{V_i \leq \varphi(s, \beta, Z_i)\right\}, \quad i = 1, \ldots, n, s \leq u\right), \tag{3.19}$$

under P, where

$$\begin{aligned} M_i\left(\varphi(u, \beta, Z_i)\right) =& N_i\left(\varphi(u, \beta, Z_i)\right) \\ &- \int_{-\infty}^{J(u,\beta,\beta_0,Z_i)} \lambda(x) Y_i\left(\varphi(x, \beta_0, Z_i)\right)\left(\rho(x, \beta_0, \beta, Z_i)\right)^{-1} dx, \end{aligned} \tag{3.20}$$

and

$$\rho(u, \beta, \beta_0, Z_i) = \nabla_1 J(u, \beta, \beta_0, Z_i) = \frac{\nabla_1 \varphi(u, \beta, Z_i)}{\nabla_1\left(\varphi(J(u, \beta, \beta_0, Z_i), \beta_0, Z_i)\right)} \tag{3.21}$$

Proposition 3.2.1 is a direct result of the compensated martingale associated with the counting process $N_i(\varphi(u, \beta, Z_i))$. The details of calculation can be found in

Section 3.4. As a special case of Proposition 3.2.1, when $\beta = \beta_0$,

$$dM_i\left(\varphi(u, \beta_0, Z_i)\right) = dN_i\left(\varphi(u, \beta_0, Z_i)\right) - \lambda(u) Y_i\left(\varphi(u, \beta_0, Z_i)\right) du \tag{3.22}$$

is a martingale differential with respect to measure P and filtration $F_n(u, \beta_0)$ defined in (3.19). Since $S_n(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{T^*} dM_i\left(\varphi(u, \beta, Z_i)\right)\left\{Z_i - \bar{Z}(u, \beta)\right\}$, it follows that

$$E_{\beta_0}\left\{S_n(\beta_0)\right\} = 0. \tag{3.23}$$

The martingale *Central Limit Theorem* (CLT) implies that $n^{-1/2} S_n(\beta_0)$ is asymptotically normal with mean zero and variance $\sigma^2(\beta_0)$, where

$$\sigma^2(\beta_0) = \int_0^{T^*} (Z_1 - \bar{Z}(u, \beta_0))^{\otimes 2} \lambda(u) P\left\{V_1 \geq \varphi(u, \beta_0, Z_1)\right\} du. \tag{3.24}$$

Now, (3.18) is in the form of extended estimating equations defined in the previous chapter. We can re-write (3.18) as

$$S_n(\beta) = \sum_{i=1}^{n} \Delta_i \left\{ Z_i - \frac{\sum_{j=1}^{n} Z_j I\left\{\varphi^{-1}(V_j, \beta, Z_j) \geq \varphi^{-1}(V_i, \beta, Z_i)\right\}}{\sum_{j=1}^{n} I\left\{\varphi^{-1}(V_j, \beta, Z_j) \geq \varphi^{-1}(V_i, \beta, Z_i)\right\}} \right\}. \tag{3.25}$$

With the choice of $X_i = (V_i, Z_i, \Delta_i)$,

$$Q(X_i, \beta) = \Delta_i, \quad C(X_i) = Z_i, \quad k(X_i, X_j, \beta) = I\left\{\varphi^{-1}(V_j, \beta, Z_j) \geq \varphi^{-1}(V_i, \beta, Z_i)\right\},$$

equation (3.18) is exactly in the form of $\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)$ mentioned in (1.10), where

$$\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta) = \sum_{i=1}^{n} Q(X_i, \beta) \left\{ C(X_i) - \frac{\sum_{j=1}^{n} C(X_j) k(X_i, X_j, \beta)}{\sum_{j=1}^{n} k(X_i, X_j, \beta)} \right\}$$

with 0 mean under the true parameter value shown in (3.23).

Corollary 3.2.1 *Under the φ-transformation model assumption, $S_n(\beta) = 0$ is an extended estimating equation, where $S_n(\beta)$ is defined in (3.25).*

3.3 Estimator for Structural Parameter

In the previous section, we constructed an extended estimating equation $S_n(\beta)$ in equation (3.18). The goal of this section is to establish the $\sqrt{n}$-consistency of estimator $\hat{\beta}_n$, the estimator for β_0 estimated via extended estimating equation $S_n(\beta)$. In Theorem 3.3.3, we prove that $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is asymptotically normal with mean zero.

Note that $S_n(\beta)$ is a step function, which brings the following two problems. First, there may not exist a root for equation $S_n(\beta) = 0$. Second, the usual Taylor expansion method does not apply to $S_n(\beta)$. Consequently, we adopt the definition of $\hat{\beta}_n$ given by Jurečková in Section 4 of [24]. Namely, for $S_n(\beta) = (S_{n,1}(\beta), \ldots, S_{n,p}(\beta))^{tr}$,

$$\hat{\beta}_n = \arg\min_{\beta \in U_{\beta_0}} \sum_{j=1}^{p} \left\{ S_{n,j}(\beta) \right\}^2, \tag{3.26}$$

where U_{β_0} is a neighbourhood of β_0 where there is a unique solution to the minimization problem (3.26). In the case when the minimization problem (3.26) has more than one solution, estimator $\hat{\beta}_n$ is defined as the one with a smaller lexicographic norm, i.e., if both vectors $\hat{b}_i = (\hat{b}_{i,1}, \ldots, \hat{b}_{i,p})$ satisfy (3.26) for $i = 1, 2$, and $\hat{b}_{1,j} = \hat{b}_{2,j}$ for $j = 1, \ldots, k$, $k < p$, and if $\hat{b}_{1,k+1} < \hat{b}_{2,k+1}$, then $\hat{\beta}_n = \hat{b}_1$.

Next, we show that $S_n(\beta)$ is asymptotically linear in a neighbourhood of β_0. Let

$$g(\beta) = \int_{-\infty}^{T^*} \lambda'(u) A_{\beta_0}(u, \beta) du, \tag{3.27}$$

where $A_{\beta_0}(u,\beta)$ is the function defined in assumption (A.6) in the previous section on page 39. Define a linear function of β as follows,

$$\tilde{S}_n(\beta) = S_n(\beta_0) + ng(\beta_0)(\beta - \beta_0). \tag{3.28}$$

Let β_n^* be the root of $\tilde{S}_n(\beta) = 0$. Consider $\beta \in U_{\beta_0}(n^{-1/2})$, a $n^{-1/2}$-neighbourhood of β_0. If we can show that $S_n(\beta)$ is 'asymptotically equivalent" to $\tilde{S}_n(\beta)$, i.e., the l^∞ norm of $S_n(\beta) - \tilde{S}_n(\beta)$ converges to zero in probability, then $\hat{\beta}_n$ is also "asymptotically equivalent" to β_n^*, namely, they are both asymptotically normal with mean zero and the identical asymptotic variance. From (3.28), we know that if $g(\beta_0)$ is nonsingular,

$$\sqrt{n}(\beta_n^* - \beta_0) = n^{-1}\left\{g(\beta_0)\right\}^{-1} S_n(\beta_0), \qquad \text{for } \beta \in U_{\beta_0}(n^{-1/2})$$

is asymptotically normal with mean zero and variance $\left\{g(\beta_0)\right\}^{-1}\sigma(\beta_0)\left\{g(\beta_0)\right\}^{-1}$. If we can show that $\sqrt{n}(\hat{\beta}_n - \beta_n^*) \xrightarrow{P} 0$, then this would imply $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is asymptotically distributed the same as $\sqrt{n}(\beta_n^* - \beta_0)$. Argued by Jurečková in [23] and [24], it would suffice to show that

$$\sup_{\beta \in U_{\beta_0}(n^{-1/2})} n^{-1/2}|S_n(\beta) - \tilde{S}_n(\beta)| \xrightarrow{P} 0. \tag{3.29}$$

We will show (3.29) in two steps. In Theorem 3.3.1, we will show the pointwise convergence of (3.29), then we get the uniformity in β in Theorem 3.3.2.

Now, let us start with the first step, i.e., the pointwise convergence by writing $S_n(\beta)$ as the summation of

$$S_{n1}(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{T^*} dM_i\left\{\varphi(u,\beta,Z_i)\right\}\left\{Z_i - \bar{Z}(u,\beta)\right\} \tag{3.30}$$

and

$$S_{n2}(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{T^*} \left\{ \lambda(J(u,\beta,\beta_0,Z_i)\rho(u,\beta,\beta_0,Z_i) - \lambda(u) \right\} \tag{3.31}$$

$$\times Y_i \left\{ \varphi(u,\beta,Z_i) \right\} \left\{ Z_i - \bar{Z}(u,\beta) \right\} du.$$

We show that $n^{-1/2}|S_{n1}(\beta) - S_n(\beta_0)| \xrightarrow{P} 0$ for fixed value $\beta \in U_{\beta_0}(n^{-1/2})$ using Lenglart's Inequality and the predictable variation process in Lemma 3.4.2 and Lemma 3.4.3, and that $n^{-1/2}|S_{n2} - ng(\beta_0)(\beta - \beta_0)| \xrightarrow{P} 0$ in Lemma 3.4.4 and 3.4.5 using the Lipshitz assumption mentioned in (A.8). A careful statement and proof of Lemma 3.4.2-3.4.5 can be found on page 50-57 of Section 3.4. Now, we are ready to discuss the pointwise asymptotic linearity of $S_n(\beta)$ for any fixed $\beta \in U_{\beta_0}(n^{-1/2})$ as follows,

Theorem 3.3.1 *Under assumptions (A.1), (A.3), and (A.5)-(A.8) for any fixed β_n that belongs to $U_{\beta_0}\left(n^{-1/2}\right)$,*

$$\frac{1}{\sqrt{n}}|S_n(\beta_n) - \tilde{S}_n(\beta_n)| \xrightarrow{P} 0. \tag{3.32}$$

Proof: Note that we can write $S_n(\beta) = S_{n1}(\beta) + S_{n2}(\beta)$, and that $\tilde{S}_n(\beta) = S_n(\beta_0) + ng(\beta_0)(\beta - \beta_0)$. Therefore the proof is complete since

$$\frac{1}{\sqrt{n}}|S_{n1}(\beta_n) - S_n(\beta_0)| \xrightarrow{P} 0, \tag{3.33}$$

$$\frac{1}{\sqrt{n}}|S_{n2}(\beta_n) - ng(\beta_0)(\beta_n - \beta_0)| = \frac{1}{\sqrt{n}} \cdot n \cdot \|\beta_n - \beta_0\| o_p(1) \xrightarrow{P} 0, \tag{3.34}$$

where we get (3.33) and (3.34) by Lemma 3.4.3 and 3.4.5, respectively. □

In Theorem 3.3.1, we have shown the pointwise linearity of $S_n(\beta)$. The next theorem guarantees that such linearity is uniform for β that belongs to a small

neighbourhood of β_0.

Theorem 3.3.2 *Under the assumptions (A.1)-(A.8), for any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\lim_{n\to\infty} P\left\{\sup_{0\le\|\beta^*-\beta_n\|\le\delta n^{-1/2}} n^{-1/2}\|S_n(\beta^*) - S_n(\beta_n)\| \ge \varepsilon\right\} = 0, \tag{3.35}$$

for any $|d| < C$.

The proof of the theorem uses the same technique as Tsiatis in [38]. The idea is to show that for a choice of sufficiently fine partitions of interval $[-C, C]$, function $S_n(\beta)$ does not fluctuate too much within the sub-intervals. Details of the proof are presented in Section 3.4. Now, we are ready to conclude the $\sqrt{n}$-consistency of $\hat{\beta}_n$ mentioned at the very beginning of this section

Theorem 3.3.3 *Let $\hat{\beta}_n$ the solution to*

$$\hat{\beta}_n = \arg\min_{\beta\in U_{\beta_0}} \sum_{j=1}^{p} \{S_{n,j}(\beta)\}^2, \tag{3.36}$$

where U_{β_0} is a neighbourhood of β_0 such that $\hat{\beta}_n$ is unique. Then under (A.1)-(A.8),

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, \{g(\beta_0)\}^{-1}\sigma^2(\beta_0)\{g(\beta_0)\}^{-1}).$$

Proof: Let β_n^* be the solution to $\tilde{S}_n(\beta) = 0$, where

$$\tilde{S}_n(\beta) = S_n(\beta_0) + ng(\beta_0)(\beta - \beta_0). \tag{3.37}$$

If $\sigma^2(\beta_0)$ denotes the asymptotic variance of $n^{-1/2}S_n(\beta_0)$ mentioned under (3.23), then

$$\sqrt{n}(\beta_n^* - \beta_0) = n^{-1/2}\{g(\beta_0)\}^{-1} S_n(\beta_0) \tag{3.38}$$

is asymptotically normal with mean zero and variance $\{g(\beta_0)\}^{-1}\sigma^2(\beta_0)\{g(\beta_0)\}^{-1}$. On the other hand, by Theorem 3.3.2,

$$\sup_{\beta\in U_{\beta_0}(n^{-1/2})} n^{-1/2}|S_n(\beta)-\tilde{S}_n(\beta)| \xrightarrow{P} 0, \tag{3.39}$$

which as proved by Jurečková in [23] implies $\sqrt{n}(\hat{\beta}_n-\beta_n^*) \xrightarrow{P} 0$. As a result, $\sqrt{n}(\hat{\beta}_n-\beta_0)$ follows the same asymptotic distribution as $\sqrt{n}(\beta_n^*-\beta_0)$, and we complete the proof. □

3.4 Some Proofs

Lemma 3.4.1 *Let $J(u,\beta_1,\beta_2,z)=\varphi^{-1}(\varphi(u,\beta_1,z),\beta_2,z)$, then*

(J1) $\nabla_1 J(u,\beta_1,\beta_2,z) = \rho(u,\beta_1,\beta_2,z)$, *and* $\nabla_2 J(u,\beta_1,\beta_2,z) = \gamma(u,\beta_1,\beta_2,z))$, *where*

$$\begin{aligned}\rho(u,\beta_1,\beta_2,z) &= \frac{\nabla_1\varphi(u,\beta_1,z)}{\nabla_1\varphi(J(u,\beta_1,\beta_2,z),\beta_2,z)},\\ \gamma(u,\beta_1,\beta_2,z) &= \frac{\nabla_2\varphi(u,\beta_1,z)}{\nabla_1\varphi(J(u,\beta_1,\beta_2,z),\beta_2,z)}.\end{aligned} \tag{3.40}$$

(J2) Let $x=J(u,\beta_1,\beta_2,z)$, then $u=J(u,\beta_2,\beta_1,z)$, and

$$\rho(x,\beta_1,\beta_2,z)=\frac{1}{\rho(u,\beta_2,\beta_1,z)} \tag{3.41}$$

(J3) $J(u,\cdot,\beta_2,z)$ is Lipschitz.

(J4) There exists a constant c such that $|\rho(u,\beta_1,\beta_2,z)-1|\le c\|\beta_1-\beta_2\|$.

(J1) By chain rule,

$$\begin{aligned}\nabla_1 J(u,\beta_1,\beta_2,Z_i) &= \nabla_1\varphi^{-1}(\varphi(u,\beta_1,z),\beta_2,z)\cdot\nabla_1\varphi(u,\beta_1,z)\\ \nabla_2 J(u,\beta_1,\beta_2,Z_i) &= \nabla_1\varphi^{-1}(\varphi(u,\beta_1,z),\beta_2,z)\cdot\nabla_2\varphi(u,\beta_1,z)\end{aligned} \tag{3.42}$$

Let $w = \varphi(u, \beta_1, z)$. Since $\varphi(\varphi^{-1}(w, \beta_2, z), \beta_2, z) = w$, differentiate both sides with respect to w, yielding $\nabla_1\varphi(\varphi^{-1}(w, \beta_2, z), \beta_2, z) \cdot \nabla_1\varphi^{-1}(w, \beta_2, z) = 1$. Therefore

$$\begin{aligned}\nabla_1\varphi^{-1}(\varphi(u, \beta_1, z), \beta_2, z) &= \nabla_1\varphi^{-1}(w, \beta_2, z) \\ &= \frac{1}{\nabla_1\varphi(\varphi^{-1}(w, \beta_2, z), \beta_2, z)} \\ &= \frac{1}{\nabla_1\varphi(J(u, \beta_1, \beta_2, z), \beta_2, z)}.\end{aligned} \tag{3.43}$$

Then (3.40) is obtained by plugging (3.43) into (3.42).

(J2) If $x = J(u, \beta_1, \beta_2, z) = \varphi^{-1}(\varphi(u, \beta_1, z), \beta_2, z)$, then $u = \varphi^{-1}(\varphi(x, \beta_2, z), \beta_1, z) = J(x, \beta_2, \beta_1, z)$. As for (3.41), it is true since

$$\begin{aligned}\rho(u, \beta_1, \beta_2, z) &= \frac{\nabla_1\varphi(u, \beta_1, z)}{\nabla_1\varphi\left(J(u, \beta_1, \beta_2, z), \beta_2, z\right)} \\ &= \frac{\nabla_1\varphi(J(x, \beta_2, \beta_1, z), \beta_1, z)}{\nabla_1\varphi(x, \beta_2, z)} = \frac{1}{\rho(x, \beta_2, \beta_1, z)}.\end{aligned}$$

(J3) Let b_1 and b_2 be two distinct points in U_{β_0}. Under assumption (A.8), $\varphi^{-1}(\cdot, \beta_2, z)$ and $\varphi(u, \cdot, z)$ are both Lipschitz. Therefore there exist constants c_1 and c_2 such that

$$\begin{aligned}|J(u, b_1, \beta_2, z) - J(u, b_2, \beta_2, z)| &= |\varphi^{-1}\left(\varphi(u, b_1, z), \beta_2, z\right) - \varphi^{-1}\left(\varphi(u, b_2, z), \beta_2, z\right)| \\ &\le c_1|\varphi(u, b_1, z) - \varphi(u, b_2, z)| \le c_1c_2\|b_1 - b_2\|\end{aligned}$$

(J4) By the definition of $\rho(u, \beta_1, \beta_2, z)$ in (3.21)

$$\begin{aligned}|\rho(u, \beta_1, \beta_2, Z_i) - 1| &= \left|\frac{\nabla_1\varphi(u, \beta_1, Z_i)}{\nabla_1\varphi\left(J(u, \beta_1, \beta_2, Z_i), \beta_2, Z_i\right)} - 1\right| \\ &= \frac{|\nabla_1\varphi(u, \beta_1, Z_i) - \nabla_1\varphi\left(J(u, \beta_1, \beta_2, Z_i), \beta_2, Z_i\right)|}{|\nabla_1\varphi\left(J(u, \beta_1, \beta_2, Z_i), \beta_2, Z_i\right)|}\end{aligned} \tag{3.44}$$

Since $\varphi(\cdot, x, y)$ is assumed to be strictly increasing, the denominator of (3.44) is strictly greater than a positive constant m. Use D denote the numerator of

(3.44), then (3.44) is bounded by D/m. Since both $\nabla_1\varphi(\cdot, y, z)$ and $\nabla_2\varphi(x, \cdot, z)$ are Lipschitz continuous under assumption (A.8),

$$\begin{aligned} D \leq & \left|\nabla_1\varphi(u, \beta_1, Z_i) - \nabla_1\varphi\left(J(u, \beta_1, \beta_2, Z_i), \beta_1, Z_i\right)\right| \\ & + \left|\nabla_1\varphi\left(J(u, \beta_1, \beta_2, Z_i), \beta_1, Z_i\right) - \nabla_1\varphi\left(J(u, \beta_1, \beta_2, Z_i), \beta_2, Z_i\right)\right| \\ \leq & \, c_1\|u - J(u, \beta_1, \beta_2, Z_i)\| + c_2\|\beta_1 - \beta_2\| \end{aligned} \tag{3.45}$$

Note that we can write $u = J(u, \beta_2, \beta_2, Z_i)$, therefore by (J3), the first term in (3.45) is bounded by $c_1c_3\|\beta_n - \beta_0\|$. Then the conclusion holds with the choice of $c = c_1c_2 + c_2$.

□

3.4.1 Proof of Proposition 3.2.1

Proof: Recall that we would like to prove $dM_i^T(\varphi(u, \beta, Z_i))$ is a martingale with filtration $F_n(u, \beta)$ in (3.19), where

$$\begin{aligned} dM_i^T\left(\varphi(u, \beta, Z_i)\right) = & dN_i^T\left(\varphi(u, \beta, Z_i)\right) \\ & - \lambda\left(J(u, \beta, \beta_0, Z_i)\right) Y_i^T\left(\varphi(u, \beta, Z_i)\right) \rho(u, \beta, \beta_0, Z_i)du, \end{aligned} \tag{3.46}$$

and $N_i^T(u)$ and $Y_i^T(u)$ are defined in (3.6), $\lambda(u)$ is the hazard rate function for ε, and $\rho(u, \beta, \beta_0, Z_i)$ is defined in Lemma 3.4.1. Change the variable in (3.46) by setting

$$u = J(v, \beta_0, \beta, Z_i) = \varphi^{-1}\left(\varphi(v, \beta_0, Z_i), \beta, Z_i\right) \tag{3.47}$$

then $\varphi(u, \beta, Z_i) = \varphi(v, \beta_0, Z_i)$, and

$$N_i^T(\varphi(u, \beta, Z_i)) = N_i^T(\varphi(v, \beta_0, Z_i)) = N_i^\varepsilon(v), \tag{3.48}$$

where we get the second equation in (3.48) since $N_i^\varepsilon(v) = N_i^T(\varphi(v, \beta_0, Z_i))$ as shown in (3.11). By the compensated counting process martingale established in [21], $M_i^\varepsilon(v)$ is a martingale, where

$$M_i^\varepsilon(v) = N_i^\varepsilon(v) - \lambda(v) Y_i^\varepsilon(v) dv \tag{3.49}$$

By (3.47), $v = J(u, \beta, \beta_0, Z_i) = \varphi^{-1}(\varphi(u, \beta, Z_i), \beta_0, Z_i)$, which implies that

$$dv = \nabla_1 J(u, \beta, \beta_0, Z_i) du = \rho(u, \beta, \beta_0, Z_i) du \tag{3.50}$$

where the last equation in (3.50) guaranteed by (J1) of Lemma 3.4.1. Then the conclusion is a result of (3.48), (3.49) and (3.50). □

3.4.2 Proof of Lemma 3.4.2-3.4.5

In Lemma 3.4.2 and 3.4.3, we will show that for any fixed β in a $n^{-1/2}$ neighbourhood of β_0, the L^2 distance between (3.30) and $S_n(\beta_0)$ converges to 0 in probability whereas Lemma 3.4.4 and 3.4.5 will show the same conclusion for (3.31) and the term $ng(\beta_0)(\beta - \beta_0)$.

Lemma 3.4.2 *Let β_n be a sequence of nonrandom vectors converging to β_0, then under the assumptions (A.5) and (A.7),*

$$\begin{aligned} \frac{1}{\sqrt{n}} \Bigg\{ \sum_{i=1}^{n} \int_{\infty}^{T^*} dM_i \left(\varphi(u, \beta_n, Z_i)\right\} \left\{Z_i - \bar{Z}(u, \beta_n)\right) \\ - \int_{\infty}^{T^*} dM_i \left(\varphi(u, \beta_n, Z_i)\right) \left\{Z_i - \mu_{\beta_0}(u, \beta_n)\right\} \Bigg\} \end{aligned} \tag{3.51}$$

converges to 0 in probability.

Proof: Let us first consider the scalar case when $\beta \in \mathbb{R}$. The expression in (3.51) is equal to $R(T^*)$, where by Proposition 3.2.1,

$$R(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{\infty}^{u} dM_i \left(\varphi(x, \beta_n, Z_i)\right) \left\{\bar{Z}(x, \beta_n) - \mu_{\beta_0}(x, \beta_n)\right\}.$$

is a martingale under P with respect to

$$F_n(u, \beta_n) = \sigma \left\{Z_i, I\left[V_i \leq \varphi(u, \beta_n, Z_i)\right], \Delta_i I\left[V_i \leq \varphi(u, \beta_n, Z_i)\right], i = 1, \ldots n\right\}$$

Therefore using the vector form of Lenglart's Inequality (see Appendix I, I.2 of [2]),

$$\begin{aligned} P\left\{|R(T^*)| > \varepsilon\right\} \leq \frac{\delta}{\varepsilon^2} + P\Bigg\{ & \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{T^*} (\bar{Z}(u, \beta_n) - \mu_{\beta_0}(u, \beta_n))^2 \\ & \times \lambda\left(J(u, \beta_n, \beta_0, Z_i)\right) \rho(u, \beta_n, \beta_0, Z_i) \\ & \times Y_i\left(\varphi(u, \beta_n, Z_i)\right) du > \delta \Bigg\} \end{aligned} \tag{3.52}$$

By assumption (A.5), we can find $N(\varepsilon, K)$ such that for any $n > N(\varepsilon, K)$,

$$P\left\{\sup_{u \leq T^*} |\bar{Z}(u, \beta_n) - \mu_{\beta_0}(u, \beta_n)| > K\right\} < \varepsilon,$$

hence with probability exceeding $1 - \varepsilon$, the integral in (3.52) is bounded by

$$K^2 \int_{-\infty}^{T^*} \lambda\left(J(u, \beta_n, , \beta_0, Z_i)\right) \rho(u, \beta_n, \beta_0, Z_i) Y_i\left(\varphi(u, \beta_n, Z_i)\right) du \tag{3.53}$$

Let $x = J(u, \beta_n, \beta_0, Z_i)$, then by property (J1) and (J2) in Lemma 3.4.1, $u = J(x, \beta_0, \beta_n, Z_i)$, $du = \rho(x, \beta_0, \beta_n, Z_i)dx$, and

$$\rho(u, \beta_n, \beta_0, Z_i) = \frac{1}{\rho(x, \beta_0, \beta_n, Z_i)}$$

Therefore (3.53) is bounded by

$$K^2 \int_{-\infty}^{J(T^*, \beta_0, \beta_n, Z_i)} \lambda(x) dx = K^2 \Lambda\left(J(T^*, \beta_0, \beta_n, Z_i)\right),$$

Consequently, the average of the integral in (3.52) over n is bounded by

$$K^2 n^{-1} \sum_{i=1}^{n} \Lambda\left(J(T^*, \beta_0, \beta_n, Z_i)\right)$$

with probability greater than $1-\varepsilon$. By assumption (A.7), if we choose $K \leq (\delta/M)^{1/2}$ and $\delta = \varepsilon^3$, then the probability in (1.48) is smaller than ϵ for $n > N(\varepsilon, K)$.

Now, consider the vector-valued parameter case, i.e., $\beta \in \mathbb{R}^p$. Let $t = (t_1, \ldots, t_p)^{tr}$ be a unit vector, let

$$R(T^*) = \sum_{i=1}^{n} \eta_i, \qquad \text{where } \eta_i = (\eta_{i1}, \ldots, \eta_{ip})^{tr} \tag{3.54}$$

then

$$\left|t^{tr} R(T^*)\right| = \left|\sum_{i=1}^{n} t^{tr} \eta_i\right| = \left|\sum_{i=1}^{n} \sum_{j=1}^{p} t_j \eta_{ij}\right| = \left|\sum_{j=1}^{p} t_j \left(\sum_{i=1}^{n} \eta_{ij}\right)\right| \leq \sum_{j=1}^{p} |t_j| \left|\sum_{i=1}^{n} \eta_{ij}\right| \tag{3.55}$$

Since $\|t\| = 1$ implies $t_k \leq 1$ for $k = 1, \ldots, p$, therefore

$$|t^{tr} R(T^*)| \leq \sum_{j=1}^{p} \left|\sum_{i=1}^{n} \eta_{ij}\right| \tag{3.56}$$

Therefore

$$P\left\{\sup_{t \in \mathbb{R}^p, \|t\|=1} |t^{tr} R(T^*)| \geq \varepsilon\right\} \leq P\left\{\sum_{j=1}^{p} \left|\sum_{i=1}^{n} \eta_{ij}\right|\right\} \leq p \max_j P\left\{\left|\sum_{i=1}^{n} \eta_{ij}\right| \geq \frac{\varepsilon}{p}\right\} \tag{3.57}$$

The probability in (3.57) converges to zero by applying the univariate Lenglart's inequality to $P\left\{\left|\sum_{i=1}^{n} \eta_{ij}\right| \geq \varepsilon/p\right\}$. Therefore the conclusion is true for $\beta \in \mathbb{R}^p$. $\square$

Lemma 3.4.3 *Let β_n be a sequence of nonrandom vectors converging to β_0, then under assumptions (A.4), (A.5), (A.7) and (A.8),*

$$\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{n}\int_{-\infty}^{T^*} dM_i\left(\varphi(u,\beta_n,Z_i)\right)\left\{Z_i-\bar{Z}(u,\beta_n)\right\}-S_n(\beta_0)\right\}\xrightarrow{P}0. \tag{3.58}$$

Proof: (3.58) can be written as the summation of the following three terms,

$$\begin{aligned}\frac{1}{\sqrt{n}}\Bigg[&\sum_{i=1}^{n}\int_{-\infty}^{T^*} dM_i\left(\varphi(u,\beta_n,Z_i)\right)\left\{Z_i-\bar{Z}(u,\beta_n)\right\}\\ &-\sum_{i=1}^{n}\int_{-\infty}^{T^*} dM_i\left(\varphi(u,\beta_n,Z_i)\right\}\left(Z_i-\mu_{\beta_0}(u,\beta_n)\right\}\Bigg],\end{aligned} \tag{3.59}$$

$$\begin{aligned}\frac{1}{\sqrt{n}}\Bigg[&\sum_{i=1}^{n}\int_{-\infty}^{T^*} dM_i\left(\varphi(u,\beta_n,Z_i)\right)\left\{Z_i-\mu_{\beta_0}(u,\beta_n)\right\}\\ &-\sum_{i=1}^{n}\int_{-\infty}^{T^*} dM_i\left(\varphi(u,\beta_0,Z_i)\right)\left\{Z_i-\mu_{\beta_0}(u,\beta_0)\right\}\Bigg],\end{aligned} \tag{3.60}$$

and

$$\frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\int_{-\infty}^{T^*} dM_i\left(\varphi(u,\beta_0,Z_i)\right)\left\{Z_i-\mu_{\beta_0}(u,\beta_0)\right\}-S_n(\beta_0)\right]. \tag{3.61}$$

By Lemma 3.4.2, (3.59) and (3.61) converges to zero in probability. We focus on the asymptotic behavior of (3.60). Let $u=J(x,\beta_0,\beta_n,Z_i)$, then the first integral in equation (3.60) is

$$\int_{-\infty}^{J(T^*,\beta_n,\beta_0,Z_i)} dM_i\left(\varphi(x,\beta_0,Z_i)\right)\left\{Z_i-\mu_{\beta_0}\left(J(x,\beta_0,\beta_n,Z_i),\beta_n\right)\right\}. \tag{3.62}$$

Write (3.60) is equal to the summation of A, B and C, where

$$A=-\frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\int_{-\infty}^{T^*} dM_i\left(\varphi(x,\beta_0,Z_i)\right)\left\{\mu_{\beta_0}\left(J(x,\beta_0,\beta_n,Z_i),\beta_n\right)-\mu_{\beta_0}(x,\beta_0)\right\}\right], \tag{3.63}$$

$$\begin{aligned}B=\frac{1}{\sqrt{n}}\Bigg[&\sum_{i=1}^{n}\int_{T^*}^{J(T^*,\beta_n,\beta_0,Z_i)} dM_i\left(\varphi(x,\beta_0,Z_i)\right)\\ &\times I\left\{T^*>J(T^*,\beta_0,\beta_n,Z_i)\right\}\left\{Z_i-\mu_{\beta_0}\left(J(x,\beta_0,\beta_n,Z_i),\beta_n\right)\right\}\Bigg]\end{aligned} \tag{3.64}$$

and

$$C = \frac{1}{\sqrt{n}} \Bigg[\sum_{i=1}^{n} \int_{J(T^*,\beta_n,\beta_0,Z_i)}^{T^*} dM_i \left(\varphi(x,\beta_0,Z_i)\right) \\ \times I\left\{T^* < J(T^*,\beta_0,\beta_n,Z_i)\right\} \left\{Z_i - \mu_{\beta_0}\left(J(x,\beta_0,\beta_n,Z_i),\beta_n\right)\right\} \Bigg]. \tag{3.65}$$

Since A, B and C integrated up to u are all $F_n(u,\beta_0)$ martingales, condition on Z_i,

$$Var(A) = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{T^*} \left\{\mu_{\beta_0}\left(J(x,\beta_0,\beta_n,Z_i),\beta_n\right) - \mu_{\beta_0}(u,\beta_0)\right\}^{\otimes 2} \\ \times \lambda(x) P\left\{V_i \geq \varphi(x,\beta_0,Z_i)\right\} dx$$

$$Var(B) = \frac{1}{n} \sum_{i \in \kappa_1}^{n} \int_{T^*}^{J(T^*,\beta_n,\beta_0,Z_i)} \left\{Z_i - \mu_{\beta_0}\left(J(x,\beta_0,\beta_n,Z_i),\beta_n\right)\right\}^{\otimes 2} \\ \times \lambda(x) P\left\{V_i \geq \varphi(x,\beta_0,Z_i)\right\} dx$$

and

$$Var(C) = \frac{1}{n} \sum_{i \in \kappa_2}^{n} \int_{J(T^*,\beta_n,\beta_0,Z_i)}^{T^*} \left\{Z_i - \mu_{\beta_0}\left(J(x,\beta_0,\beta_n,Z_i),\beta_n\right)\right\}^{\otimes 2} \\ \times \lambda(x) P\left\{V_i \geq \varphi(x,\beta_0,Z_i)\right\} dx$$

where we calculate the variances using the predictable variation process

$$\langle dM_i, dM_i \rangle = \lambda(u) Y_i\left(\varphi(u,\beta_0,Z_i)\right).$$

By definition of $\mu_{\beta_0}(x,\beta)$ in assumption (A.5) and the boundedness of Z_i assumed in assumption (A.4), we know that $\mu_{\beta_0}(x,\beta)$ is bounded by 1 for all x and β, which implies

$$\|Var(A)\| \leq \frac{4}{n} \sum_{i=1}^{n} \int_{-\infty}^{T^*} \lambda(x) S(x) dx \leq 4.$$

Therefore by the continuity of $\mu_{\beta_0}(u,\beta)$, $\varphi(x,y)$ and the dominated convergence theorem, $\|Var(A)\|$ converges to 0. therefore (3.63) converges to 0 in probability. As for (3.64), note that

$$Var(B) = \frac{1}{n}\sum_{i\in\kappa_1}\int_{T^*}^{J(T^*,\beta_n,\beta_0,Z_i)} \{Z_i - \mu_{\beta_0}\left(J(x,\beta_0,\beta_n,Z_i),\beta_n\right)\}^{\otimes 2}$$

$$\times\, \lambda(x)P\left\{V_i \geq \varphi(x,\beta_0,Z_i)\right\}dx$$

Since $V_i = \min(T_i, C_i)$

$$P\left\{V_i \geq \varphi(x,\beta_0,Z_i)\right\} \leq P\left\{T_i \geq \varphi(x,\beta_0,Z_i)\right\} = P\left\{\varphi^{-1}(T_i,\beta_0,Z_i) \geq x\right\}$$

$$= P\left\{\varepsilon_i \geq x\right\} = S(x),$$

and that $\lambda(x)S(x) = f(x)$, we know that

$$Var(B) \leq \frac{1}{n}\sum_{\kappa_1}^{n}\int_{T^*}^{J(T^*,\beta_n,\beta_0,Z_i)} \lambda(x)S(x)dx. \tag{3.66}$$

Since $\lambda(x)S(x) = f(x)$ is bounded by 1, (3.66) implies that

$$Var(B) \leq 4K_1 n^{-1}\sum_{i=1}^{n}\left\{J(T^*,\beta_n,\beta_0,Z_i) - T^*\right\} \tag{3.67}$$

Note that $T^* = J(T^*,\beta_0,\beta_0,Z_i)$. Since $J(u,\cdot,\beta_2,z)$ is Lipschitz as shown in (J3) of Lemma 3.4.1, there exists a constant c such that the terms in the summand of the right hand side of (3.67) is bounded by $c\|\beta_n - \beta_0\|$, which implies that $Var(B)$ goes to 0 in probability. Using the same approach, we conclude that $Var(C)$ also approaches 0 in probability. Hence the proof is complete. □

Lemma 3.4.4 *Let S_{n2} be defined as in (3.31), and S_{n3} be*

$$S_{n3} = \sum_{i=1}^{n}\int_{-\infty}^{T^*}\Big\{\lambda(J(u,\beta_n,\beta_0,Z_i)) - \lambda(u)\Big\} \tag{3.68}$$

$$\times\, Y_i\left(\varphi(u,\beta_n,Z_i)\right)\left\{Z_i - \bar{Z}(u,\beta_n)\right\}du,$$

then under assumption (A.1) and (A.8),

$$n^{-1} \cdot |S_{n2}(\beta) - S_{n3}(\beta)| \xrightarrow{P} 0. \tag{3.69}$$

Proof: By (3.31) and (3.68),

$$\begin{aligned} n^{-1}|S_{n2}(\beta) - S_{n3}(\beta)| &\leq n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{T^*} \lambda(J(u, \beta_n, \beta_0, Z_i)) \\ &\quad \times |\rho(u, \beta_n, \beta_0, Z_i) - 1| Y_i\left(\varphi(u, \beta_n, Z_i)\right) \|Z_i - \bar{Z}(u, \beta_n)\| du \\ &\leq 2n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{T^*} \lambda(J(u, \beta_n, \beta_0, Z_i)) \\ &\quad \times |\rho(u, \beta_n, \beta_0, Z_i) - 1| Y_i\left(\varphi(u, \beta_n, Z_i)\right) du \end{aligned} \tag{3.70}$$

By (J4), $|\rho(u, \beta_n, \beta_0, Z_i) - 1| \leq c\|\beta_n - \beta_0\|$, therefore by (3.70), we can see that

$$\begin{aligned} n^{-1}|S_{n2}(\beta) - S_{n3}(\beta)| \leq 2c\|\beta_n - \beta_0\| n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{T^*} \lambda\left(J(u, \beta_n, \beta_0, Z_i)\right) \\ \times Y_i\left(\varphi(u, \beta_n, Z_i)\right) du \end{aligned} \tag{3.71}$$

Since

$$\begin{aligned} E\left\{Y_i\left(\varphi(u, \beta_n, Z_i)\right)\right\} &\leq P\left\{T_i \geq \varphi(u, \beta_n, Z_i)\right\} \\ &= P\left\{\varepsilon \geq J(u, \beta_n, \beta_0, Z_i)\right\} \\ &= S\left(J(u, \beta_n, \beta_0, Z_i)\right), \end{aligned} \tag{3.72}$$

we know that

$$\begin{aligned} n^{-1}|S_{n2}(\beta) - S_{n3}(\beta)| &\leq 2c\|\beta_n - \beta_0\| n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{T^*} \lambda\left(J(u, \beta_n, \beta_0, Z_i)\right) \\ &\quad \times S\left(J(u, \beta_n, \beta_0, Z_i)\right) du \\ &= 2c\|\beta_n - \beta_0\| n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{T^*} f\left(J(u, \beta_n, \beta_0, Z_i)\right) du \\ &= 2c\|\beta_n - \beta_0\| n^{-1} \sum_{i=1}^{n} F\left(J(T^*, \beta_n, \beta_0, Z_i)\right) \end{aligned} \tag{3.73}$$

By assumption (A.1), $F\left(J(T^*, \beta_n, \beta_0, Z_i)\right)$ is bounded by $\{c_1 c_2 \|\beta_n - \beta_0\| + 1\}$, therefore by (3.73), $n^{-1}|S_{n2}(\beta) - S_{n3}(\beta)|$ approaches 0 in probability. □

Lemma 3.4.5 *Let* $g(\beta_0) = \int_{-\infty}^{T^*} \lambda'(u) A_{\beta_0}(u, \beta) du$, *and* $S_{n2}(\beta)$ *as shown in (3.31), then under assumptions (A.1), (A.3), (A.6) and (A.8),*

$$n^{-1} S_{n2}(\beta) = g(\beta_0)(\beta - \beta_0) + o_P(\|\beta_n - \beta_0\|) \tag{3.74}$$

Proof: By Lemma 3.4.4, it suffices to show

$$n^{-1} S_{n3} = g(\beta_0)(\beta - \beta_0) + o_P(\|\beta - \beta_0\|). \tag{3.75}$$

Note that the left hand side of (3.75) is the summation of

$$n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{T^*} \lambda'(u) Y_i\left(\varphi(u, \beta_0, Z_i)\right) \left\{Z_i - \bar{Z}(u, \beta_n)\right\} \gamma(u, \beta_0, Z_i)^{tr} du (\beta_n - \beta_0) \tag{3.76}$$

and

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{T^*} & Y_i\left(\varphi(u, \beta_0, Z_i)\right) \left\{Z_i - \bar{Z}(u, \beta_n)\right\} \\ & \times \left\{\lambda\left(J(u, \beta_n, \beta_0, Z_i)\right) - \lambda(u) - \lambda'(u) \gamma(u, \beta_0, Z_i)^{tr} (\beta_n - \beta_0)\right\} du \end{aligned} \tag{3.77}$$

where $\gamma(u, \beta_0, Z_i)$ is defined in (3.15). By assumption (A.6), (3.76) converges to

$$\int_{-\infty}^{T^*} \lambda'(u) A_{\beta_0}(u, \beta_0) du (\beta_n - \beta_0) = g(\beta_0)(\beta_n - \beta_0). \tag{3.78}$$

As for (3.77), it is bounded by

$$2n^{-1} \|\beta_n - \beta_0\|^2 \sum_{i=1}^{n} \int_{-\infty}^{T^*} |\theta(u, \beta_n, Z_i)| du \tag{3.79}$$

Since the average over n in (3.79) is bounded in probability by assumption (A.3), we finish the proof by (3.78) and (3.79). □

3.4.3 Proof of Theorem 3.3.2

First, let us consider a uni-variate $\beta \in \mathbb{R}$. We form a mesh with space approaching zero from $-C$ to C using a finite number points $d_0, \ldots, d_m$. By Theorem 3.3.1, for $\beta_{n,i} = \beta_0 + d_i/\sqrt{n}$,

$$\max_{i \le m} \left\{ n^{-1/2} \| S_n(\beta_{n,i}) - \tilde{S}_n(\beta_{n,i}) \| \right\} \xrightarrow{P} 0. \tag{3.80}$$

In order to prove (3.35), we must show that $n^{-1/2} S_n(\beta)$ does not fluctuate too much from $\beta_{n,i}$ to $\beta_{n,i+1}$ for $i = 1, \ldots, m$ for any choice of partition points $d_0, \ldots, d_m$. More specifically, for any $\varepsilon > 0$, there exists a positive δ such that for $\beta_n = \beta_0 + \delta n^{-1/2}$,

$$\lim_{n \to \infty} P \left\{ \sup_{\beta_n \le \beta^* \le \beta_n + \delta n^{-1/2}} n^{-1/2} \| S_n(\beta^*) - S_n(\beta_n) \| \ge \varepsilon \right\} = 0, \tag{3.81}$$

for any $|d| < C$.

As for the case when $\beta \in \mathbb{R}^p$, we consider a p-dimensional mesh by allowing β to change its coordinate one at a time. More specifically, for a fixed β^*, define

$$\check{\beta}_{n,j} = (\beta^{*(1)}, \ldots, \beta^{*(j)}, \beta_n^{(j+1)}, \ldots, \beta_n^{(p)})^{tr},$$

then

$$S_n(\beta^*) - S_n(\beta_n) = \sum_{j=0}^{p-1} \left\{ S_n(\check{\beta}_{n,j+1}) - S_n(\check{\beta}_{n,j}) \right\}. \tag{3.82}$$

Therefore

$$\| S_n(\beta^*) - S_n(\beta_n) \| \le p \cdot \max_j \left\| S_n(\check{\beta}_{n,j+1}) - S_n(\check{\beta}_{n,j}) \right\| \tag{3.83}$$

In the right hand side of (3.83), the change only occurs in the $(j+1)$th coordinate, and the other coordinates are fixed. Consequently, the uniform convergence for $\beta \in \mathbb{R}^p$ will follow if we can show (3.81) for $\beta \in \mathbb{R}$.

For the rest of this subsection, without loss of generality, assume that $\beta \in \mathbb{R}$. Recall that $V_i = \min(T_i, C_i)$. Define the residuals r_i as

$$r_i \equiv r(V_i, \beta, \beta_0, Z_i) = \varphi^{-1}(V_i, \beta, Z_i) = J\left(\varepsilon_i, \beta_0, \beta, Z_i\right),$$

We can complete the proof of (3.81) by putting a probabilistic bound on the maximum change of $S_n(\beta^*)$ as β^* varies from β_n to $\beta_n + \delta n^{-1/2}$. Recall that

$$S_n(\beta) = \sum_{i=1}^{n} \Delta_i \left\{ Z_i - \frac{\sum_{j=1}^{n} Z_j I\left\{V_j \geq \varphi(\varepsilon_i, Z_j, \beta_i)\right\}}{\sum_{j=1}^{n} I\left\{V_j \geq \varphi(\varepsilon_i, Z_j, \beta_i)\right\}} \right\} \tag{3.84}$$

so $S_n(\beta^*)$ is a function of the ranks of residuals r_i, hence change in $S_n(\beta^*)$ occurs whenever the change of β^* from β_n to $\beta_n + \delta n^{-1/2}$ leads to a change of r_i, $i = 1, \ldots, n$. Therefore the maximum change of $S_n(\beta^*)$ can be calculated by computing $F_1 \times F_2$, where

$$\begin{aligned} L_1 &= \#[\text{pairs of interchanged ranks }], \\ L_2 &= \left[\begin{array}{c} \text{the maximum change of } S_n(\beta^*) \\ \text{for each such interchange} \end{array} \right] \end{aligned} \tag{3.85}$$

In Lemma 3.4.6, we investigate the two factors of (3.85). Then, using Lemma 3.4.6, we prove the uniform linearity as stated in Theorem 3.3.2.

Lemma 3.4.6 *Let T^* be a value such that $P\left\{V_i \geq T^* + \xi\right\} \geq \psi > 0$. Define*

$$\mathcal{B}_n \equiv \left\{ \boldsymbol{X}_\infty : \frac{1}{n} \sum_{i=1}^{n} I\left\{V_i \geq T^* + \xi\right\} \geq \frac{\psi}{2} \right\}, \tag{3.86}$$

where $\boldsymbol{X}_\infty = \{X_i\}_{i=1}^n$. *Under assumption (A.4) and (A.8), for any* $\varepsilon > 0$, *there exists* N_ε *such that for any* $n > N_\varepsilon$,

(i) $P\{\mathcal{B}_n\} > 1 - \varepsilon$;

(ii) *for* $n > N_\varepsilon$ *and* $\boldsymbol{X}_\infty \in \mathcal{B}_n$, $P\{nL_2 \leq 6c/\psi\} \geq 1 - \varepsilon$.

(iii) $L_1 = \sum_{i=1}^n \sum_{j \neq i}^n I(A_{ij})$, *where* A_{ij} *is the event* $|V_i - V_j| \leq c \cdot \delta n^{-1/2}$

Proof: The conclusion (i) in is guaranteed by the Law of Large Numbers. Next, let us consider (ii) which shows how to bound L_2 in (3.85). Note that whenever the change of β^* from β_n to $\beta_n + \delta n^{-1/2}$ causes an interchange in ranks of the residual, the interchange must happen between two adjacent order statistics of r_i, $i = 1, \ldots, n$. Let $\left\{r_{(i)}\right\}_{i=1}^n$ be the set of order statistics of $\{r_i\}_{i=1}^n$, and denote the corresponding covariate and failure indicator by $Z_{(i)}(\beta^*)$ and $\Delta_{(i)}(\beta^*)$, then $S_n(\beta^*)$ can be written as

$$\sum_{i=1}^n \Delta_i(\beta^*)\left\{Z_{(i)}(\beta^*) - \bar{Z}_{(i)}\right\}, \quad \text{where } \bar{Z}_{(i)}(\beta^*) = \sum_{k=i}^n \frac{Z_{(k)}(\beta^*)}{n-i+1}. \tag{3.87}$$

Now, assume that the change of β^* from β_n to $\beta_n + \delta n^{-1/2}$ causes an interchange in ranks between two adjacent order statistics $r_{(j)}$ and $r_{(j+1)}$, then the new $S_n(\beta^{*+})$ is

$$\begin{aligned}
&\sum_{i=1}^{j-1} \Delta_i(\beta^*)\left\{Z_{(i)}(\beta^*) - \bar{Z}_{(i)}(\beta^*)\right\} + \Delta_{(j+1)}(\beta^*)\left\{Z_{(j+1)}(\beta^*) - \bar{Z}_{(j)}(\beta^*)\right\} \\
&\quad + \Delta_{(j)}(\beta^*)\left\{Z_{(j)}(\beta^*) - \frac{\bar{Z}_{(j+2)}(\beta^*)(n-j-1) + Z_{(j)}(\beta^*)}{n-j}\right\} \\
&\quad + \sum_{i=j+2}^n \Delta_i(\beta^*)\left\{Z_{(i)}(\beta^*) - \bar{Z}_{(i)}(\beta^*)\right\}.
\end{aligned} \tag{3.88}$$

Hence the difference of $S_n(\beta^*)$ before and after the interchange in $r_{(j)}$ and $r_{(j+1)}$ is (3.87) minus (3.88), which equals to

$$\begin{aligned}\left\{\Delta_{(j+1)}(\beta^*) - \Delta_{(j)}(\beta^*)\right\}&\left\{\frac{\bar{Z}_{(j+2)}(\beta^*)(n-j-1)}{n-j} - \bar{Z}_{(j)}(\beta^*)\right\} \\ &+\frac{\Delta_{(j+1)}(\beta^*)Z_{(j+1)}(\beta^*)}{n-j} - \frac{\Delta_{(j)}(\beta^*)Z_{(j)}(\beta^*)}{n-j},\end{aligned} \tag{3.89}$$

where we use the fact that

$$\bar{Z}_{(j+1)}(\beta^*) = \frac{\sum_{k=j+1}^{n} Z_{(k)}(\beta^*)}{n-j} = \frac{\bar{Z}_{(j+2)}(\beta^*)(n-j-1) + Z_{j+1}(\beta^*)}{n-j}.$$

Then (3.89) is equal to

$$\begin{cases} \dfrac{Z_{(j+1)} - Z_{(j)}}{n-j} & \text{if } \Delta_{(j)} = \Delta_{(j+1)} = 1 \\ -\dfrac{Z_{(j)}}{(n-j)(n-j+1)} + \dfrac{Z_{(j+1)}}{n-j+1} + \dfrac{\bar{Z}_{(j+2)}(n-j-1)}{(n-j)(n-j+1)} & \text{if } \Delta_{(j)} = 1 \text{ and } \Delta_{(j+1)} = 0 \\ \dfrac{\bar{Z}_{(j+1)} - Z_{(j)}}{n-j+1} & \text{if } \Delta_{(j)} = 0 \text{ and } \Delta_{(j+1)} = 1 \end{cases} \tag{3.90}$$

For any of the three cases in (3.90), since $\|Z_i\|$'s are bounded by a constant c with probability 1 under assumption (A.4), the change in $S_n(\beta^*)$ is bounded by $c/(n-j)$, where $n-j$ is the number of r_i's at risk at the point where the interchange occurs.

Consider $\boldsymbol{X}_\infty \in \mathcal{B}_n$ on which

$$\frac{1}{n}\sum_{i=1}^{n} I\left\{V_i \geq T^* + \xi\right\} \geq \frac{\psi}{2}.$$

Since $S_n(\beta)$ is computed for r_i's that are less than $\varphi^{-1}(T^*, \beta_n, Z_i)$, for $\boldsymbol{X}_\infty \in \mathcal{B}_n$, the number of the r_i's at risk will exceed $n\psi/2$ if an interchange occurs, i.e., $n-j \geq n\psi/2$. Consequently, the change in $S_n(\beta)$ is bounded by $(6c/\psi)n^{-1}$.

Then we try to find L_1 in (3.85), i.e., the number of interchanges when β^* varies from β_n to $\beta_n + \delta n^{-1/2}$. An interchange between (i, j) will occur for β_{ij} if $\varepsilon_i = \varepsilon_j$. Therefore by assumption (A.4) and (A.8), an interchange occurring for values of β^* between β_n and $\beta_n + \delta n^{-1/2}$ implies

$$|V_i - V_j| = \left|\varphi\left(\varepsilon_i, \beta_n, Z_i\right) - \varphi\left(\varepsilon_i, \beta_n + \frac{\delta}{\sqrt{n}}, Z_j\right)\right| \leq C_2 \cdot \frac{\delta}{\sqrt{n}} \tag{3.91}$$

The total number of interchanges equates

$$L_1 = \sum_{i=1}^{n} \sum_{j \neq i}^{n} I(A_{ij}),$$

where A_{ij} denotes the event in (3.91). □

Proof of Theorem 3.3.2 In Lemma 3.4.6, we have shown that the maximum change of $S_n(\beta)$ after each interchange in ranks is bounded by $(4/\psi)n^{-1}$, and the the number of interchanges as β^* varies from β_n to $\beta_n + \delta n^{-1/2}$ is L_1. If we can show that

$$\lim_{n \to \infty} P\left\{n^{-3/2} L_1 \geq \varepsilon\right\} = 0 \tag{3.92}$$

for some $\delta > 0$ that is to be chosen properly, then the proof is complete, i.e.,

$$\lim_{n \to \infty} P\left\{\sup_{0 \leq \|\beta^* - \beta_n\| \leq \delta n^{-1/2}} n^{-1/2} \|S_n(\beta^*) - S_n(\beta_n)\| \geq \varepsilon\right\} = 0.$$

For $1 \leq i < j \leq n$, Let

$$W_{ij} = I(A_{ij}) + I(A_{ji}),$$

$$U_i = \sum_{j \neq i}^{n} \left\{E(W_{ij}|V_i) - E(W_{ij})\right\},$$

$$U_{ij} = W_{ij} - E(W_{ij}|V_i) - E(W_{ij}|V_j) + E(W_{ij}).$$

Then

$$L\_E\left\{L_1\right\} = \sum_{i=1}^{n} U_i + \sum_{i<j}^{n}\sum^{n} U_{ij}$$

is the sum of pairwise uncorrelated random variables. Let $f_i^*(u)$ be the density function of V_i, hence $f_i(u) = f(u)H(u) + h(u)S(u)$. Let

$$P\left\{A_{ij}|V_i\right\} = \int_0^{C_2\delta/\sqrt{n}} f_j^*(u)du \leq \frac{2(K_1+K_2)C_2\delta}{\sqrt{n}},$$

where the inequality can be attained by assumption A and B. Similarly, $P\left\{A_{ji}|V_i\right\} \leq 2(K_1+K_2)C_2\delta/\sqrt{n}$, so

$$\left|E\left\{W_{ij}|V_i\right\}\right| \leq 4(K_1+K_2)C_2\delta n^{-1/2},$$

therefore

$$E\left\{L_1\right\} \leq 2(K_1+K_2)C_2\delta n^{3/2}, \qquad \sigma^2(M) = \sum_{i=1}^{n}\sigma^2(U_i) + \sum_{i<j}^{n}\sum^{n}\sigma^2(V_{ij}) = O(n^2)$$

Since in (3.92), $P\left\{n^{-3/2}L_1 \geq \varepsilon\right\}$ is bounded by $P\left\{|L_1 - E(L_1)| \geq n^{3/2}\varepsilon - E(L_1)\right\}$, using the Chebyshev's inequality,

$$\begin{aligned} P\left\{|L_1 - E(L_1)| \geq n^{3/2}\varepsilon - E(L_1)\right\} &\leq \frac{\sigma^2(L_1)}{(n^{3/2}\varepsilon - E(L_1))^2} \\ &\leq \frac{Cn^2}{(n^{3/2}\varepsilon - 2(K_1+K_2)C_2\delta n^{3/2})^2}, \end{aligned} \tag{3.93}$$

where C is a constant. Let $\delta = \varepsilon/\left\{3C_2(K_1+K_2)\right\}$, then the probability in (3.93) is bounded by $3C/(n\varepsilon^2)$, hence we have shown (3.85). □

3.5 List of Notations

$T = \varphi(\varepsilon, \beta_0, Z)$ page 34

$J(u, \beta_1, \beta_2, z) \equiv \varphi^{-1}(\varphi(u, \beta_1, z), \beta_2, z)$ page 38

$\gamma(u, \beta_0, z) = \nabla_2 J(u, \beta, \beta_0, z)|_{\beta=\beta_0} = \dfrac{\nabla_2 \varphi(u, \beta_0, z)}{\nabla_1 \varphi(u, \beta_0, z)}.$ page 38

$\bar{Z}(u, \beta) = \dfrac{\sum_{j=1}^{n} Z_j Y_j \{\varphi(u, \beta, Z_j)\}}{\sum_{j=1}^{n} Y_j \{\varphi(u, \beta, Z_j)\}}$ page 39

$S_n(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{T^*} dN_i^T \{\varphi(u, \beta, Z_i)\} \left\{Z_i - \bar{Z}(u, \beta)\right\}$ page 41

$\rho(u, \beta, \beta_0, Z_i) = \nabla_1 J(u, \beta, \beta_0, Z_i) = \dfrac{\nabla_1 \varphi(u, \beta, Z_i)}{\nabla_1 \left(\varphi(J(u, \beta, \beta_0, Z_i), \beta_0, Z_i)\right)}$ page 41

$M_i \left(\varphi(u, \beta, Z_i)\right) = N_i \left(\varphi(u, \beta, Z_i)\right) - \int_{-\infty}^{J(u,\beta,\beta_0,Z_i)} \dfrac{\lambda(x) Y_i \left(\varphi(x, \beta_0, Z_i)\right)}{\rho(x, \beta_0, \beta, Z_i)} dx$ page 41

$\tilde{S}_n(\beta) = S_n(\beta_0) + g(\beta_0)(\beta - \beta_0)$ page 44

Chapter 4: Technical Results I: Owen's Lemmas & Empirical Process

Consider d-dimensional *independently identically distributed* (i.i.d.) random vectors from a common distribution family

$$X_1, \ldots, X_n \sim P_\theta, \quad \text{where } \theta = (\beta, \lambda). \tag{4.1}$$

The parameter θ is consisted of the structural part $\beta \in \mathbb{R}^p$, and the nuisance part $\lambda \in \mathcal{H}$ that is infinite dimensional. *Classic Estimating Equation* (CEE) assumes there exists function $m(x, \beta)$ satisfying Definition 1.1.1 in Section 1.2. We extended the CEE to the *Extended Estimating Equation* (EEE) in Definition 1.2.1, in which the estimating function is denoted by $m_n(x, \boldsymbol{x}, \beta)$. A direct way of constructing an estimator from the CEE and EEE is solving

$$\tilde{\beta}_n : \quad S_n(\beta) = 0, \tag{4.2}$$

where $S_n(\beta) = \sum_{i=1}^{n} m(X_i, \beta)$ or $S_n(\beta) = \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)$.

Under the CEE setting, another way to construct an estimator for β_0 is through the *Empirical Likelihood* (EL) method using $S_n(\beta) = 0$ as a constraint. The method

is first to define the probability vector $\{\hat{p}_i(\beta, X_i, \boldsymbol{X}, \beta_0)\}_{i=1}^n$ that solves

$$\begin{cases} \max_{\boldsymbol{p}} \prod p_i, \quad \text{where } \boldsymbol{p} = (p_1, \ldots, p_n); \\ \text{subject to } \sum_{i=1}^n p_i = 1, p_i \in (0,1), \sum_{i=1}^n p_i m(X_i, \beta) = 0. \end{cases} \tag{4.3}$$

Then $\hat{\beta}_n$ can be defined via maximizing the *Profile Empirical Likelihood* (pEL), or equivalently, minimizing the negative logarithm of the pEL, i.e.

$$\hat{\beta}_n = \arg\min_{\beta} l(\beta), \quad \text{where } l(\beta) = -\sum_{i=1}^n \ln\left(\hat{p}_i(\beta, X_i, \boldsymbol{X}, \beta_0)\right) \tag{4.4}$$

However, to prove that there exists a neighborhood U_{β_0} of β_0 such that for any $\beta \in U_{\beta_0}$, there exists a unique solution to (4.3), and to establish the asymptotic behavior of $\hat{\beta}_n$, we need to prove lemmas parallel to Lemma 11.2 and 11.4 in [32] by Owen, and Lemma 1 and Theorem 1 by Qin in [33]. Since the major difference between the CEE and the EEE is the appearance of the higher order summations in the latter, the simplest torms of the *Law of Large Numbers* (LLN) and *Central Limit Theorem* (CLT) cannot be applied directly in the EEE setting. To handle more general forms of summation, we introduce the concepts of Donsker and GlivenKo-Cantelli classes in the empirical process theory using the series of examples in Chapter 19 of [41].

4.1 Assumptions and Notations

Let X_i, for $i = 1, \ldots, n$ be the i.i.d. sample defined in (4.1) with common distribution function $F_X(x, \theta)$ supported on $\mathbb{R}^d$. Let U_{β_0} be a neighbourhood of β_0 that

will be defined in Lemma 5.2.2 and (5.48), and define

$$m_n(X_i, \boldsymbol{X}, \beta) = Q(X_i, \beta) \left\{ C(X_i) - \frac{\sum_{j=1}^n C(X_j) k(X_i, X_j, \beta)}{\sum_{j=1}^n k(X_i, X_j, \beta)} \right\}. \tag{4.5}$$

$$\bar{m}_n(\boldsymbol{X}, \beta) = \frac{1}{n} \sum_{i=1}^n m_n(X_i, \boldsymbol{X}, \beta). \tag{4.6}$$

$$Z_n^*(\boldsymbol{X}, \beta) = \max_{1 \le i \le n} \|m_n(X_i, \boldsymbol{X}, \beta)\|; \tag{4.7}$$

$$\overline{k_c}(X_i, \beta) = E\left\{C(X_j) k(X_i, X_j, \beta) | X_i\right\}, \quad \text{for } i \neq j; \tag{4.8}$$

$$\bar{k}(X_i, \beta) = E\left\{k(X_i, X_j, \beta) | X_i\right\}, \quad \text{for } i \neq j; \tag{4.9}$$

$$\eta(X_i, X_j, \beta) = C(X_j) k(X_i, X_j, \beta) - \overline{k_c}(X_i, \beta), \quad \text{for } i \neq j \tag{4.10}$$

$$E\left\{\cdot\right\} \equiv E_{\theta_0}\left\{\cdot\right\}; \quad P\left\{\cdot\right\} \equiv P_{\theta_0}\left\{\cdot\right\} \tag{4.11}$$

Note that $\eta(X_i, X_j, \beta)$ is defined such that $E_\theta\left\{\eta(X_i, X_j, \beta) | X_i\right\} \equiv 0$, for all $\theta \in \Theta$.

In this Chapter, we also use $O(\cdot)$, $o(\cdot)$ to represent the almost sure magnitude, and $O_P(\cdot)$ and $o_P(\cdot)$ the magnitude in probability. A more detailed definition of these four notations can be found in Chapter 1 of [36]. In general, for a vector $\upsilon \in \mathbb{R}^r$, $\|\upsilon\|$ denotes the Euclidean norm.

Assumptions

(A.1) $E_{\beta_0, \lambda}\left\{m_n(X, \boldsymbol{X}, \beta_0)\right\} = 0$, for all $\lambda \in \mathcal{H}$.

(A.2) For all x and $\beta \in U_{\beta_0}$, $\|C(x)\| < b < \infty$, $|Q(x, \beta)| < M < \infty$.

(A.3) $E_\theta\left\{\nabla_\beta m(X, \boldsymbol{X}, \beta)\right\}$ is nonsingular for $\theta \in U_{\beta_0} \times \mathcal{H}$.

(A.4) For any $(x, \theta) \in \mathbb{R}^d \times (U_{\beta_0} \times \mathcal{H})$, $k(\cdot, \cdot, \cdot) > 0$; for positive γ that is close to 1, $E\left\{\bar{k}(X_i, \beta_0)^{-\gamma}\right\} < \infty$

(A.5) For each x and $\boldsymbol{x}$, $\nabla_\beta m_n(x, \boldsymbol{x}, \beta)$ exists for β in U_{β_0}, and is continuous at β_0.

(A.6) $E\{\nabla_\beta m_n(x, \boldsymbol{x}, \beta_0)\}$ is of full rank p.

(A.7) $E[m_n(X, \boldsymbol{X}, \beta_0) m_n^{tr}(X, \boldsymbol{X}, \beta_0)]$ is positive definite.

(A.8) There exists a constant M such that the jth component η_j of $\eta = \eta(x, X, \beta_0)$ in (4.10) satisfies $|\eta_j| \leq M$, and

$$E\left\{e^{|\eta_j|/M} - 1 - \frac{\eta_j}{M}\right\} M^2 \leq \frac{1}{2} var(\eta_j).$$

(A.9) There exists M such that for any β, $\|k(x, y, \beta)\| < M$.

(A.10) Let $\alpha = (x, \beta) \in \mathbb{R}^d \times U_{\beta_0}$ with Euclidean norm, and $k(x, y, \beta) = k_\alpha(y)$. There exists a measurable function $b(y) : \mathbb{R}^d \mapsto \mathbb{R}$ such that for any $\alpha_1 \neq \alpha_2$,

$$|k_{\alpha_1}(y) - k_{\alpha_2}(y)| \leq b(y)\|\alpha_1 - \alpha_2\|,$$

and $E\{|b(X_1)|\} < \infty$.

(A.11) Let $\gamma > 0$ be close to 1, then $E\left\{\bar{k}(X_1, \beta_0)^{-\gamma}\right\} < \infty$.

4.2 Lemmas Parallel to Owen, and Qin and Lawless

In this section, we provide some lemmas that are parallel to Lemmas 11.2 through 11.4 in Chapter 11 of [32], which are used by Owen to establish the *Nonarametric Maximum Empirical Likelihood* (NPMELE) estimator and the Wilks type theorem when the dimension of the estimating function r equals the dimension of the structural parameter p. Qin and Lawless also applied these lemmas when generalizing

Owen's work to the case when $r > p$. We will show that under regularity conditions, even though estimating functions of *Extended Estimating Equations* (EEE) denoted by $m_n(X_i, \boldsymbol{X}, \beta)$ are no longer i.i.d., the parallel versions of these Lemmas continue to hold.

The way to overcome loss of independence in EEE is to refer to the tools in empirical process theory. Compare the following two expressions,

$$m_n(X_i, \boldsymbol{X}, \beta) = Q(X_i, \beta) \left\{ C(X_i) - \frac{\sum_{j=1}^n C(X_j) k(X_i, X_j, \beta)}{\sum_{j=1}^n k(X_i, X_j, \beta)} \right\} \tag{4.12}$$

versus

$$V(X_i, \beta) = Q(X_i, \beta) \left\{ C(X_i) - \frac{E\left\{C(Y) k(X_i, Y, \beta) | X_i\right\}}{E\left\{k(X_i, Y, \beta) | X_i\right\}} \right\}. \tag{4.13}$$

Applying the LLN to the numerator and denominator in equation (4.12), then applying Slutsky's lemma, we can see that (4.12) and (4.13) are close when n is large. In order to take advantage of the similarity of (4.12) and (4.13) uniformly over indices i, we need to show the higher order summations in the former converge to the corresponding terms in the latter in probability, and uniformly in X_i and $\beta \in U_{\beta_0}$. More strictly,

Lemma 4.2.1 *Suppose $k(x, y, \beta) : \mathbb{R}^d \times \mathbb{R}^d \times U_{\beta_0} \mapsto \mathbb{R}^+$ and $C(x) : \mathbb{R}^d \mapsto \mathbb{R}^p$ satisfies assumptions (A.1), (A.2), (A.9) and (A.10), where U_{β_0} is a open and bounded set in Θ that contains the true parameter value β_0. Then*

$$\sup_{(x,\beta) \in \mathcal{X} \times U_{\beta_0}} \left| \frac{1}{n} \sum_{j=1}^n k(x, X_j, \beta) - E\left\{k(x, X, \beta)\right\} \right| \xrightarrow{P} 0 \tag{4.14}$$

and

$$\sup_{(x,\beta)\in\mathcal{X}\times U_{\beta_0}} \left| \frac{1}{n}\sum_{j=1}^{n} C(X_j)k(x,X_j,\beta) - E\left\{C(X)k(x,X,\beta)\right\} \right| \xrightarrow{P} 0 \tag{4.15}$$

The proof of the lemma involves constructing Glivenko-Cantelli classes, and an application of parametric class discussed by Van der Vaart in Chapter 19 of [41]. The following Proposition is parallel to Lemma 11.2 in Chapter 11 of [32]

Proposition 4.2.1 *Let* $Z_n^*(\boldsymbol{X},\beta) = \max_{1\le i\le n} \|m_n(X_i,\boldsymbol{X},\beta)\|$. *Under assumptions (A.3), (A.4) and (A.6), for any fixed* $\beta\in\Theta$,

$$Z_n^*(\boldsymbol{X},\beta) = o(n^{1/2}). \tag{4.16}$$

To prove Proposition 4.2.1, we decompose $m_n(X_i,\boldsymbol{X},\beta)$ into an i.i.d part and a non-i.i.d. quotient part. Then the conclusion can be drawn by applying Lemma 11.2 in [32] to the two parts separately.

The following two lemmas play roles that are equivalent to the *Law of Iterated Logarithm* (LIL) in [32]. They will be applied to prove the EEE version of Lemma 11.4 in [32].

Lemma 4.2.2 *Let* $0<\delta<1/2$, *then under assumptions (A.2), (A.4), (A.5), (A.8), (A.9) and (A.10)*

$$n^{-1/2}\sum_{i=1}^{n} m_n(X_i,\boldsymbol{X},\beta_0) = O_P(n^{\delta}) \tag{4.17}$$

Lemma 4.2.3 *Let* $0<\delta<1/2$, *then under assumptions (A.2), (A.4), (A.5), (A.8), (A.9) and (A.10)*

$$\frac{1}{n}\sum_{i=1}^{n} \|m_n(X_i,\boldsymbol{X},\beta_0)\|^2 = O(n^{\delta}). \tag{4.18}$$

The proofs of Lemmas 4.2.2 and 4.2.3 involve two steps. First, $m_n(X_i, \boldsymbol{X}, \beta_0)$ is split into the summation of an i.i.d. part and a higher order summation quotient part. The magnitude of the i.i.d. term is given by the LIL, and the quotient bounded by Berstein's Inequality listed as a proposition in Section 4.3. We also prove both lemmas first for univariate β, then generalize the conclusion to the multivariate case. With the conclusion in Lemma 4.2.3, we present Proposition 4.2.2 which is parallel to Lemma 11.3 in Chapter 11 of [32].

Proposition 4.2.2 *Under assumptions of Lemma 4.2.3,*

$$n^{-1} \sum_{i=1}^{n} \|m_n(X_i, \boldsymbol{X}, \beta_0)\|^3 = o(n^{1/2}). \tag{4.19}$$

The proof of Proposition 4.2.2 is a direct application of Proposition 4.2.1 and Lemma 4.2.3.

4.3 Some Proofs

Proof of Lemma 4.2.1 The convergence of (4.15) and (4.14) for fixed (x, β) is guaranteed by the LLN. Therefore the main concern is to prove the uniformity in parameter (x, β). Let $\alpha = (x, \beta)$, $k_\alpha(y) = k(x, y, \beta)$ and $\mathcal{K} = \{k_\alpha(y), \alpha \in \mathcal{X} \times U_{\beta_0}\}$, where $\mathcal{X}$ and U_{β_0} are bounded subsets of $\mathbb{R}^d$ and $\mathbb{R}^p$ as defined in Definition 1.2.1. Under assumption (A.10), $\mathcal{K}$ forms a parametric class mentioned in Example 19.7 on page 271 of [41], which refers to a class of functions that are Lipschitz in a finite-dimensional parameter on a bounded region, but the function domain may be unbounded. Therefore $\mathcal{K}$ is a Donsker class, and as in Theorem A.0.1, is also a

Glivenko-Cantelli class. Hence the convergence in (4.15) is uniform in (x, β).

As for (4.14), note that $\mathcal{KC} = \{C(y)k_\alpha(y), \alpha \in \mathcal{X} \times U_{\beta_0}\}$, for any $\alpha_1, \alpha_2 \in \mathcal{X} \times U_{\beta_0}$,

$$\|C(y)k_{\alpha_1}(y) - C(y)k_{\alpha_2}(y)\| \leq \|C(y)\| \cdot \|k_{\alpha_1}(y) - k_{\alpha_2}(y)\| \tag{4.20}$$

Under assumption (A.2), the right hand side of (4.20) is bounded by $\|C(y)\| \cdot b(y)\|\alpha_1 - \alpha_2\|$. Under assumption (A.2) and (A.10),

$$E|\, b(Y) \cdot \|C(Y)\| \,| = E|b(Y)| \cdot E\|C(Y)\| < \infty \tag{4.21}$$

Therefore $\mathcal{KC}$ also forms a parametric class mentioned in Theorem A.0.2 of Appendix A. So with the same reasoning we made for (4.15), the convergence (4.14) is also uniform in $(x, \beta) \in \mathcal{X} \times U_{\beta_0}$. □

Proof of Proposition 4.2.1 Let $Z_n^*(\boldsymbol{X}, \beta) = \max_{1 \leq i \leq n} \|m_n(X_i, \boldsymbol{X}, \beta)\|$. Since

$$m_n(X, \boldsymbol{X}, \beta) = Q(X, \beta)\left(C(X) - \frac{\sum_{j=1}^n C(X_j)k(X, X_j, \beta)}{\sum_{j=1}^n k(X, X_j, \beta)}\right),$$

For a fixed β, $Z_n^*(\boldsymbol{X}, \beta)$ is bounded by the sum of A and B, where

$$A = \max_i \|Q(X_i, \beta)C(X_i)\|,$$

and

$$B = \max_i \left\| Q(X_i, \beta)\frac{\sum_{j=1}^n C(X_j)k(X_i, X_j, \beta)}{\sum_{j=1}^n k(X_i, X_j, \beta)} \right\|.$$

By Lemma 11.2 in [32], under assumption (A.6), $A = o(n^{1/2})$. As for B, it is bounded by

$$\max_i |Q(X_i, \beta)| \cdot \max_i \left\| \frac{\sum_{j=1}^n C(X_j)k(X_i, X_j, \beta)}{\sum_{j=1}^n k(X_i, X_j, \beta)} \right\|.$$

Since the $k(\cdot,\cdot,\beta)$'s are assumed to be nonnegative in assumption (A.4),

$$\max_i \left\| \frac{\sum_{j=1}^n C(X_j)k(X_i,X_j,\beta)}{\sum_{j=1}^n k(X_i,X_j,\beta)} \right\| \leq \max_i \left(\frac{\sum_{j=1}^n k(X_i,X_j,\beta)\cdot \max_j \|C(X_j)\|}{\sum_{j=1}^n k(X_i,X_j,\beta)} \right)$$

$$= \max_j(\|C(X_j)\|).$$

Then under the boundedness assumption in (A.2),

$$B \leq \max_i |Q(X_i,\beta)| \max_i \|C(X_i)\| = o(n^{1/2}).$$

Therefore for any fixed $\beta \in U_{\beta_0}$, $Z_n^*(\boldsymbol{X},\beta) = o(n^{1/2})$. □

Before giving the proof of Lemma 4.2.3, we state the Bernstein Inequality as follows. This is a well-known theorem and can be found in references like [42].

Proposition 4.3.1 (Bernstein's Inequality) *Let $X_1,\ldots,X_n$ be independent variables with zero mean such that $E|X_i|^m \leq m!M^{m-2}v_i/2$, for every $m \geq 2$ and all i and some constant M and v_i. Then*

$$P(|X_1+\cdots+X_n| > x) \leq 2\exp\left\{-\frac{1}{2}\cdot\frac{x^2}{v+Mx}\right\},$$

for $v \geq v_1+\cdots+v_n$.

Proof of Lemma 4.2.2 Let us start with the univariate case that $\beta_0 \in \Theta \subset \mathbb{R}$ and $C(x) : \mathbb{R}^d \mapsto \mathbb{R}$. We can split $\sqrt{n}\bar{m}_n(\boldsymbol{X},\beta_0)$ as the difference of the following two terms,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n Q(X_i,\beta_0)\left\{C(X_i) - \frac{\overline{k_c}(X_i,\beta_0)}{\bar{k}(X_i,\beta_0)}\right\} \tag{4.22}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n Q(X_i,\beta_0)\left\{\frac{\sum_{j=1}^n C(X_j)k(X_i,X_j,\beta_0)}{\sum_{j=1}^n k(X_i,X_j,\beta_0)} - \frac{\overline{k_c}(X_i,\beta_0)}{\bar{k}(X_i,\beta_0)}\right\}. \tag{4.23}$$

Note that (4.22) is a summation of i.i.d. terms, therefore under assumption (A.5), by the LIL, it is $O(\sqrt{\ln\ln n})$, and is asymptotically normal by CLT. Next, we evaluate the order of magnitude of (4.23). To begin with, let us rewrite (4.23) as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Q(X_i,\beta_0)}{\bar{k}(X_i,\beta_0)\sum_{j=1}^{n}k(X_i,X_j,\beta_0)} \times\left\{\bar{k}(X_i,\beta)\sum_{j=1}^{n}C(X_j)k(X_i,X_j,\beta_0)-\overline{k_c}(X_i,\beta_0)\sum_{j=1}^{n}k(X_i,X_j,\beta_0)\right\} \quad (4.24)$$

Then we can split (4.24) into the difference of the following two terms,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Q(X_i,\beta_0)}{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}k(X_i,X_j,\beta_0)}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(C(X_j)k(X_i,X_j,\beta_0)-\bar{k}_c(X_i,\beta_0)\right) \quad (4.24.a)$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Q(X_i,\beta_0)\overline{k_c}(X_i,\beta_0)}{\bar{k}(X_i,\beta_0)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}k(X_i,X_j,\beta_0)}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(k(X_i,X_j,\beta_0)-\bar{k}(X_i,\beta_0)\right) \quad (4.24.b)$$

Let us consider the following term in the numerator of (4.24.a),

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{C(X_j)k(X_i,X_j,\beta_0)-\overline{k_c}(X_i,\beta_0)\right\} \quad (4.24.a.1)$$

$$=\frac{1}{\sqrt{n}}\sum_{j:j\neq i}^{n}\left\{C(X_j)k(X_i,X_j,\beta_0)-\overline{k_c}(X_i,\beta_0)\right\}+\frac{1}{\sqrt{n}}\left\{C(X_i)k(X_i,X_i,\beta_0)-\overline{k_c}(X_i,\beta_0)\right\}$$

Note that by Proposition 4.2.1, for any $i=1,\ldots,n$, the norm of the second term in equation (4.24.a.1) is bounded by

$$\frac{1}{\sqrt{n}}\max_i\|C(X_i)k(X_i,X_i,\beta_0)-\overline{k_c}(X_i,\beta_0)\|=\frac{1}{\sqrt{n}}\cdot o(n^{1/2})=o(1).$$

Therefore for $i=1,\ldots,n$, we can rewrite (4.24.a.1) as

$$\frac{1}{\sqrt{n}}\sum_{j:j\neq i}\left\{C(X_j)k(X_i,X_j,\beta_0)-\overline{k_c}(X_i,\beta_0)\right\}+o(1). \quad (4.24.a.2)$$

Then by substituting (4.24.a.2) for (4.24.a.1), we know that (4.24.a) is equal to

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}&\frac{Q(X_i,\beta_0)}{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}k(X_i,X_j,\beta_0)}\\&\times\left\{\frac{1}{\sqrt{n}}\sum_{j:j\neq i}\Big(C(X_j)k(X_i,X_j,\beta_0)-\bar{k}_c(X_i,\beta_0)\Big)+o(1)\right\}\end{aligned}\tag{4.24.a.3}$$

By (4.8),

$$E\left\{\overline{k_c}(X_i,\beta_0)\right\}=E\left\{C(X_j)k(X_i,X_j,\beta_0)\right\}.$$

Therefore the terms inside the summation of (4.24.a.2) have expectation zero. Moreover, conditioned on X_i, (4.24.a.2) is the summation of i.i.d. terms with zero mean, under assumption (A.8), we can apply Bernstein's inequality in Proposition 4.3.1 with the choice of constant x equals to $k_n=cn^{(\delta+1)/2}$. Since δ is between 0 and 1/2,

$$\frac{k_n^2}{n\sigma+Mk_n}=\frac{n^{1+\delta}}{\sigma n+Mn^{(1+\delta)/2}}\sim\frac{n^{\delta}}{\sigma}\qquad\text{as}\qquad n\to\infty$$

where $\sigma=\sup_{x\in\mathcal{X}}var\left\{C(X_1)k(x,X_1,\beta_0)-\bar{k}_c(x,\beta_0)\right\}$, and $\sim$ means that the ratio of the two expressions converges to 1.Hence by Proposition 4.3.1, for $i=1,\ldots,n$,

$$P\left\{\left|\sum_{j:j\neq i}C(X_j)k(X_i,X_j,\beta_0)-\overline{k_c}(X_i,\beta_0)\right|\geq k_n\middle|X_i\right\}\leq 2e^{-cn^{\delta}}.$$

By putting together all such sets for different i, we get

$$P\left\{\max_i\left|\sum_{j:j\neq i}C(X_j)k(X_i,X_j,\beta_0)-\overline{k_c}(X_i,\beta_0)\right|\geq k_n\right\}\leq 2n\,e^{-cn^{\delta}}.$$

Therefore

$$\begin{aligned}\sum_{n=1}^{\infty}P\left\{\max_i\frac{1}{\sqrt{n}}\left|\sum_{j:j\neq i}C(X_j)k(X_i,X_j,\beta_0)-\overline{k_c}(X_i,\beta_0)\right|\geq\frac{k_n}{\sqrt{n}}\right\}&\\\leq 2\sum_{n=1}^{\infty}ne^{-cn^{\delta}}<\infty.&\end{aligned}\tag{4.25}$$

By Borel-Cantelli lemma, the inequality above implies that for all n sufficiently large,

$$\max_i \left| \frac{1}{\sqrt{n}} \sum_{j \neq i} C(X_j) k(X_i, X_j, \beta_0) - \overline{k_c}(X_i, \beta_0) \right| \leq \frac{k_n}{\sqrt{n}} = n^{\delta}, \quad \text{a.s.} \tag{4.26}$$

Now, let us discuss the denominator in (4.24.a), namely,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} k(X_i, X_j, \beta_0). \tag{4.27}$$

In the following proof, we would like to show that under the assumptions that we have made up to now and assumption (A.11), (4.27) is $O(n^{\delta})$. Split (4.27) in to

$$\frac{1}{\sqrt{n}} \sum_{j: i \in \mathcal{J}} k(X_i, X_j, \beta_0) + \frac{1}{\sqrt{n}} \sum_{j: i \notin \mathcal{J}} k(X_i, X_j, \beta_0), \tag{4.28}$$

where $\mathcal{J} \equiv \left\{ \bar{k}(X_i, \beta_0) < c' n^{(\delta-1)/2} \right\}$. With the same argument used to deduct (4.25), by the Berstein inequality, for $\delta^* = \delta/3$,

$$\max_i \left| \frac{1}{\sqrt{n}} \sum_{j: j \neq i} k(X_i, X_j, \beta) - \bar{k}(X_i, \beta_0) \right| \leq c n^{\delta^*/2}, , \quad \text{a.s.} \tag{4.29}$$

which implies that we can re-write (4.28) as

$$\frac{1}{\sqrt{n}} \sum_{\mathcal{J}} k(X_i, X_j, \beta_0) + I \left\{ \bar{k}(X_i, \beta_0) \geq c' n^{(\delta-1)/2} \right\} \cdot \left\{ \bar{k}(X_i, \beta_0) + O(n^{(\delta^*-1)/2}) \right\} \tag{4.30}$$

Next, we discuss the order of magnitude of the two terms in (4.30). Since $k(\cdot, \cdot, \cdot)$ is bounded assumed to be bounded by M in assumption (A.9), and the first term in (4.30) is symmetric with respect to i, we know that the first term in (4.30)

$$\begin{aligned} \frac{2M}{\sqrt{n}} E \{ \text{ number of } i \in \mathcal{J} \} &= 2M\sqrt{n} P \left\{ \bar{k}(X_i, \beta_0) < c' n^{(\delta-1)/2} \right\} \\ &= 2M\sqrt{n} P \left\{ \bar{k}(X_i, \beta_0)^{-\gamma} > c'^{-\gamma} n^{\gamma(1-\delta)/2} \right\} \\ &\leq 2M E \left\{ \bar{k}(X_i, \beta_0)^{-\gamma} \right\} n^{\gamma \cdot \frac{\delta-1}{2} \cdot \frac{1}{2}} \leq n^{\delta} \end{aligned} \tag{4.31}$$

where γ is the positive constant that is close to 1 defined in (A.11) and the inequality is attained by the Chebyshev's inequality. So in (4.30), the first term is $O_P(n^\delta)$. Combining (4.26) - (4.31) we know that (4.24.a.3) is bounded in probability by

$$\frac{2}{n}\sum_{i=1}^{n}\frac{Q(X_i,\beta_0)}{O_P(n^\delta)+I\left\{\bar{k}(X_i,\beta_0)\geq c'n^{(\delta-1)/2}\right\}\bar{k}(X_i,\beta_0)+O(n^{(\delta^*-1)/2})}n^\delta. \tag{4.32}$$

With the choice of $\delta^*=\delta/3$, by the LIL for i.i.d. summands, the order of (4.32) is given by $O(\sqrt{\ln\ln n/n})$.

By far we have shown that (4.24.a.3) is $O(n^\delta)$. Since (4.24.a.3) equals to (4.24.a), the latter is also $O(n^\delta)$. Using the same strategy, we can also prove that (4.24.b) is $O(n^\delta)$. Therefore $\sqrt{n}\bar{m}_n(\boldsymbol{X},\beta_0)=O(n^\delta)$.

Next, we generalize our conclusion to vector valued β_0 and $C(x)$. It suffices to show that (4.26) holds for $\beta_0\in\mathbb{R}^p$ and $C(x):\mathbb{R}^d\mapsto\mathbb{R}^p$. Define a p-dimensional vector $\eta_j=(\eta_{j1},\ldots,\eta_{jp})^{tr}$, where

$$\eta_j=C(X_j)k(X_i,X_j,\beta_0)-\overline{k_c}(X_i,\beta_0),\quad j=1,\ldots,n \text{ and } i\neq j.$$

Then by (4.26), for each component η_{jk} of η_j,

$$P\left\{\limsup_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{j:j\neq i}\eta_{jk}\geq\frac{k_n}{\sqrt{n}}\right\}=0,\quad k=1,\ldots,p. \tag{4.33}$$

Let $t=(t_1,\ldots,t_p)^{tr}$ be a unit vector in $\mathbb{R}^p$,

$$\left|\sum_{j:j\neq i}^{n}t^{tr}\eta_i\right|=\left|\sum_{j:j\neq i}^{n}\sum_{k=1}^{p}t_k\eta_{jk}\right|=\left|\sum_{k=1}^{p}t_k\left(\sum_{j:j\neq i}^{n}\eta_{jk}\right)\right|\leq\sum_{k=1}^{p}|t_k|\left|\sum_{i=1}^{n}\eta_{jk}\right| \tag{4.34}$$

where we attain the inequality using triangle inequality. Since $\|t\|=1$, $|t_k|\leq 1$ for $k=1,\ldots p$, then together with (4.34),

$$\left|\sum_{j=1}^{n}t^{tr}\eta_i\right|\leq\sum_{k=1}^{p}\left|\sum_{j:j\neq i}^{n}\eta_{jk}\right| \tag{4.35}$$

Note that (4.35) holds for any unit vector in $\mathbb{R}^p$, which indicates that

$$P\left\{\sup_{t\in\mathbb{R}^p,\|t\|=1}\left|\sum_{j:j\neq i}^{n} t^{tr}\eta_j\right|\geq k_n\right\}\leq P\left\{\sum_{k=1}^{p}\left|\sum_{j:j\neq i}^{n}\eta_{jk}\right|\geq k_n\right\}$$
$$\leq p\max_k P\left\{\left|\sum_{j:j\neq i}^{n}\eta_{jk}\right|\geq \frac{k_n}{p}\right\} \tag{4.36}$$

Then combine (4.33) and (4.36), we know that

$$P\left\{\limsup_{n\to\infty}\sup_{t\in\mathbb{R}^p,\|t\|=1}\left|\frac{1}{\sqrt{n}}\sum_{j:j\neq i}^{n} t^{tr}\eta_j\right|\geq \frac{k_n}{\sqrt{n}}\right\}=0 \tag{4.37}$$

□

Proof of Lemma 4.2.3 Note that we can split the left hand side of (4.18) as the summation of the following three expressions,

$$\frac{1}{n}\sum_{i=1}^{n} Q(X_i,\beta_0)^2 C(X_i)^{tr}C(X_i)=O\left(\sqrt{\frac{\ln\ln n}{n}}\right), \tag{4.18.a}$$

$$-\frac{2}{n}\sum_{i=1}^{n} Q(X_i,\beta_0)^2 C(X_i)^{tr}\frac{\sum_{j=1}^{n}C(X_j)k(X_i,X_j,\beta_0)}{\sum_{j=1}^{n}k(X_i,X_j,\beta_0)}=O(n^{\delta}), \tag{4.18.b}$$

$$\frac{1}{n}\sum_{i=1}^{n} Q(X_i,\beta_0)^2\frac{\left[\sum_{j=1}^{n}C(X_j)k(X_i,X_j,\beta_0)/n\right]^{tr}\left[\sum_{j=1}^{n}C(X_j)k(X_i,X_j,\beta_0)/n\right]}{\left[\sum_{j=1}^{n}k(X_i,X_j,\beta_0)/n\right]^2}, \tag{4.18.c}$$

where we get the order of (4.18.a) and (4.18.b) applying the LIL and the same strategy used in the proof of (4.26), respectively. Next, we investigate the order of (4.18.c). By the proof of Lemma 4.2.2, an almost sure upper bound of (4.18.c),

$$\frac{4}{n}\sum_{i=1}^{n}\frac{Q(X_i,\beta_0)^2}{\bar{k}(X_i,\beta_0)^2}\left[\frac{1}{n}\sum_{j=1}^{n}C(X_j)k(X_i,X_j,\beta_0)\right]^{tr}\left[\frac{1}{n}\sum_{j=1}^{n}C(X_j)k(X_i,X_j,\beta_0)\right] \tag{4.38}$$

Similar to the proof of Lemma 4.2.2, let us assume that $C(X)$ and β_0 are scalar valued then generalize the conclusion into vector valued case. Applying the same

method in the proof of Lemma 4.2.2 where equation (4.24.a.2) was attained, we can show that for $i = 1, \ldots, n$,

$$\frac{1}{n}\sum_{j=1}^{n} C(X_j)k(X_i, X_j, \beta_0) = \frac{1}{n}\sum_{j:j\neq i}^{n} C(X_j)k(X_i, X_j, \beta_0) + o(1).$$

Next, we show that for $k_n = O(n^{\frac{1}{2}\delta+\frac{5}{4}})$.

$$\frac{1}{n}\sum_{j:j\neq i}^{n} C(X_j)k(X_i, X_j, \beta_0) \leq \frac{k_n}{n}, \quad \text{a.s., for } i = 1, \ldots, n.$$

Case 1 If $C(X)k(x, X, \beta_0)$ is centered at 0 under β_0, Let $k_n = O(n^{\frac{1}{2}\delta+\frac{5}{4}})$ be the constant mentioned in Bernstein's Inequality. Since δ is between 0 and 1/2,

$$O\left(\frac{x^2}{n\sigma + Mx}\right) = O\left(\frac{n^{\delta+\frac{5}{2}}}{n\sigma + Mn^{\frac{1}{2}\delta+\frac{5}{4}}}\right) = O\left(n^{\frac{1}{2}\delta+\frac{5}{4}}\right).$$

Hence by Proposition 4.3.1, for $i = 1, \ldots, n$

$$P\left\{\sum_{j:j\neq i}^{n} C(X_j)k(X_i, X_j, \beta_0) \geq k_n \;\middle|\; X_i\right\} \leq e^{-k_n^2/(n\sigma+Mk_n)} = e^{-cn^{\frac{1}{2}\delta+\frac{5}{4}}}.$$

By putting together all such sets for $i = 1, \ldots, n$, we get

$$P\left\{\sum_{j:j\neq i}^{n} C(X_j)k(X_i, X_j, \beta_0) \geq k_n\right\} \leq e^{-k_n^2/(n\sigma+Mk_n)} = e^{-cn^{\frac{1}{2}\delta+\frac{5}{4}}}.$$

Therefore

$$\sum_{i=1}^{n} P\left\{\frac{1}{n}\sum_{j:j\neq i}^{n} C(X_j)k(X_i, X_j, \beta_0) \geq \frac{k_n}{n}\right\} \leq \sum_{i=1}^{n} e^{-n^{\frac{1}{2}\delta+\frac{5}{4}}} < \infty$$

By Borel-Cantelli lemma, the inequality above indicates that

$$P\left\{\limsup_{n\to\infty} \frac{1}{n}\sum_{j:j\neq i}^{n} C(X_j)k(X_i, X_j, \beta_0) \geq \frac{k_n}{n}\right\} = 0, \tag{4.39}$$

which shows that

$$\frac{1}{n}\sum_{j:j\neq i}^{n} C(X_j)k(X_i, X_j, \beta_0) = O\left(\frac{k_n}{n}\right) = O\left(n^{\frac{1}{2}\delta+\frac{1}{4}}\right),$$

and that

$$\frac{1}{n}\sum_{j:j\neq i}^{n} C(X_j)k(X_i, X_j, \beta_0) \leq \frac{k_n}{n}, \quad \text{a.s.} \tag{4.40}$$

Case 2 $C(X)k(x, X, \beta_0)$ is centered at $E\{C(X)k(x, X, \beta_0)\} = \mu(x, \beta_0) \neq 0$, then by (4.39)

$$\frac{1}{n}\sum_{j:j\neq i}^{n} \{C(X_j)k(X_i, X_j, \beta_0) - \mu(X_i, \beta_0)\} = O\left(\frac{k_n}{n}\right) = O\left(n^{\frac{1}{2}\delta+\frac{1}{4}}\right), \tag{4.41}$$

and

$$\frac{1}{n}\sum_{j:j\neq i}^{n} \{C(X_j)k(X_i, X_j, \beta_0) - \mu(X_i, \beta_0)\} \leq \frac{k_n}{n}, \quad \text{a.s.} \tag{4.42}$$

From (4.41) and (4.42), we know that

$$\frac{1}{n}\sum_{j:j\neq i}^{n} C(X_j)k(X_i, X_j, \beta_0) = \mu(X_1, \beta_0) + O(n^{\frac{1}{2}\delta+\frac{1}{4}}) = O(n^{\frac{1}{2}\delta+\frac{1}{4}}) \tag{4.43}$$

and

$$\frac{1}{n}\sum_{j\neq i}^{n} C(X_j)k(X_i, X_j, \beta_0) \leq \frac{k_n + n\mu(X_1, \beta)}{n} = \frac{k_n}{n}, \quad \text{a.s.} \tag{4.44}$$

Plug (4.44) into (4.38), yielding an almost sure upper bound for (4.18.c)

$$\frac{4}{n}\sum_{i=1}^{n} \frac{Q(X_i, \beta_0)^2}{\bar{k}(X_i, \beta_0)^2} \cdot \frac{k_n}{n} \cdot \frac{k_n}{n} = O\left(\sqrt{\frac{\ln\ln n}{n}}\right) \cdot O\left(\frac{n^{\frac{1}{2}\delta+\frac{5}{4}}}{n}\right) \cdot O\left(\frac{n^{\frac{1}{2}\delta+\frac{5}{4}}}{n}\right) = O(n^{\delta}) \tag{4.45}$$

Therefore $\sum_{i=1}^{n} \|m_n(X_i, \boldsymbol{X}, \beta)\|^2/n$ is also $O(n^{\delta})$. This conclusion can be extended for $\beta_0 \in \mathbb{R}^p$ using the same strategy as the proof of Lemma 4.2.2. □

Proof of Proposition 4.2.2 Since $Z_n^*(\boldsymbol{X}, \beta_0) = \max_{1\leq i\leq n} \|m_n(X_i, \boldsymbol{X}, \beta_0)\|$, write

$$\frac{1}{n}\sum_{i=1}^{n} \|m_n(X_i, \boldsymbol{X}, \beta)\|^3 \leq Z_n^*(\boldsymbol{X}, \beta_0) \cdot \frac{1}{n}\sum_{i=1}^{n} \|m_n(X_i, \boldsymbol{X}, \beta_0)\|^2, \tag{4.46}$$

By Proposition 4.2.1, $Z_n^*(\boldsymbol{X}, \beta_0) = o(n^{1/2})$. By Lemma 4.2.3, the second factor on the right hand side of (4.46) is $O(n^\delta)$ with δ between 0 and 1/2. Therefore

$$\frac{1}{n}\sum_{i=1}^{n} \|m_n(X_i, \boldsymbol{X}, \beta_0)\|^3 = o(n^{1/2}).$$

$\square$

Chapter 5: Technical Results II: Zero in the Convex Hull Theorems

In this chapter, we continue to discuss the technical results for *Classic Estimating Equations* (CEE) and *Extended Estimating Equations* (EEE), with the same notations and assumptions in Chapter 4, from pages 65 - 68. We define a neighbourhood of β_0 such that for all β in this neighbourhood, the *Empirical Likelihood* (EL) method has a unique maximizer with probability approaching 1.

Under the CEE setting, consider random samples $X_1, \ldots, X_n$ from distribution family P_θ, where $\theta = (\beta, \lambda) \in \Theta \times \mathcal{H}$, $\Theta \subset \mathbb{R}^p$ and $\mathcal{H}$ infinite dimensional. Using the Lagrange multiplier method, it can be shown that for fixed β, if there exists a unique solution to

$$\begin{cases} \max_{\boldsymbol{p}} \prod_{i=1}^{n} p_i \quad \text{where } \boldsymbol{p} = (p_1, \ldots, p_n); \\ \text{subject to } \sum_{i=1}^{n} p_i = 1, p_i \in (0,1), \sum_{i=1}^{n} p_i m(X_i, \beta) = 0, \end{cases} \tag{5.1}$$

then the solution is given by

$$\begin{aligned} &\hat{p}_i(\beta, X_i, \boldsymbol{X}) = \frac{1}{n} \cdot \frac{1}{1 + t^{tr} m(X_i, \beta)}, \\ &\text{where } t = t(\beta, \boldsymbol{X}) \text{ solves } \sum_{i=1}^{n} \frac{m(X_i, \beta)}{1 + t^{tr} m(X_i, \beta)} = 0. \end{aligned} \tag{5.2}$$

A typical interpretation of $\hat{p}_i(\beta, X_i, \boldsymbol{X})$ is that the distribution of X_1 is approximated by the modified empirical measure $\sum_{i=1}^{n} p_i \delta_{X_i}$. This is why the condition

$$E_{\theta_0}\{m(X_1, \beta_0)\} = 0$$

is rendered through the approximating distribution of X_1 as $\sum_{i=1}^{n} p_i m(X_i, \beta_0) = 0$ connecting the p_i's and β.

Owen in [32], and Qin and Lawless in [33], claimed that for a fixed β, a sufficient condition for problem (5.2) to have a local unique solution is "zero in the convex hull", i.e., for fixed $\beta \in U_{\beta_0}$, $\mathbf{0} \in \mathrm{Conv}(\beta) \subset \mathbb{R}^r$, where

$$\mathrm{Conv} \equiv \left\{ \sum_{i=1}^{n} p_i m(X_i, \beta) : \sum_{i=1}^{n} p_i = 1, p_i \in (0,1), \beta \in U_{\beta_0} \right\}. \tag{5.3}$$

However, they did not state explicitly how the set U_{β_0} in (5.3) is constructed, and whether (5.3) is a deterministic fact, or an asymptotic result.

In the following section, we answer the two questions in the previous paragraph. Then we generalize these conclusions to discontinuous estimating function $m(x, \cdot)$. In Section 5.2, we prove the existence and uniqueness of solutions to (5.1) under the EEE setting with estimating function $m_n(x, \boldsymbol{x}, \beta)$ that is continuous with respect to β. For the discontinuous case, we prove the conclusion for the φ-transformation model, which to our knowledge, is the broadest class of semiparametric models satisfying the EEE definitions.

5.1 Classic Estimating Equation

In this section, under the CEE setting, we establish the uniqueness of solutions to maximization problem (5.1) initially for continuous estimating functions, then for discontinuous ones.

5.1.1 Continuous Criterion Function

We prove the existence and uniqueness of solutions to (5.1) in the following steps. First, we demonstrate that (5.3) is true when $\beta = \beta_0$ and that $\hat{p}_i(\beta_0, X_i, \boldsymbol{X})$ is the calculus maximizer of (5.1). Then, using the continuity of $m(x, \cdot)$, we apply Rolle's theorem on the gradient of the Lagrangian of the negative logarithm of the *Profile Empirical Likelihood* (pEL) function to prove the uniqueness of solution to the second equation in (5.2). Finally, combining the results in the previous two steps with a continuation method, we prove that there exists a neighbourhood of β_0 on which (5.1) has a calculus maximum with probability approaching 1.

Let us start with the first step described in the previous paragraph, which is showing $0 \in \mathrm{Conv}(\beta_0)$ with probability approaching 1.

Lemma 5.1.1 *For i.i.d. random variables $X_1, \ldots, X_n$, assume that $m(x, \cdot)$ is con-*

tinuous, and

$$E\left\{m(X_1,\beta_0)\right\}=0; \tag{5.4}$$

$$E\left\{m(X_1,\beta_0)^{\otimes 2}\right\}=\Sigma \text{ is positive definite.} \tag{5.5}$$

$$E\|m(X_1,\beta_0)\|^3<\infty. \tag{5.6}$$

For any constant $K>0$, let $\boldsymbol{X}=\{X_i\}_{i=1}^{\infty}$, t_0 be a unit vector in $\mathbb{R}^p$.

(a) *Let λ_0 be the smallest eigenvalue of Σ, and $I\in\mathbb{R}^{p\times p}$ be the identity matrix, define*

$$\mathcal{A}_{K,n}^{(1)}=\left\{\boldsymbol{X}:\inf_{t_0}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}t_0^{tr}m(X_i,\beta_0)\geq -K\right\}, \tag{5.7}$$

$$\mathcal{A}_{\varepsilon,n}^{(2)}=\left\{\boldsymbol{X}:\left\|\frac{1}{n}\sum_{i=1}^{n}m(X_i,\beta_0)^{\otimes 2}-\Sigma\right\|_2\leq\varepsilon\right\}, \tag{5.8}$$

$$\mathcal{A}_{n}^{(3)}=\left\{\boldsymbol{X}:\frac{1}{n}\sum_{i=1}^{n}\|m(X_i,\beta_0)\|^3\leq 2\cdot E\|g(X_1,\beta_0)\|^3\right\}, \tag{5.9}$$

$$\mathcal{A}_{n}^{(4)}=\left\{\boldsymbol{X}:\frac{1}{n}\sum_{i=1}^{n}m(X_i,\beta_0)^{\otimes 2}\geq\frac{1}{2}\lambda_0 I\right\}. \tag{5.10}$$

Then for any $\varepsilon\in(0,\lambda_0)$, there exist K_ε and N_ε, such that

(i) for any $n\geq N_\varepsilon$, $P\left\{\mathcal{A}_{K_\varepsilon,n}^{(1)}\right\}\geq 1-\varepsilon$;

(ii) the following limits are all identical to 1

$$\lim_{n_1\to\infty}P\left\{\bigcap_{n\geq n_1}\mathcal{A}_{\varepsilon,n}^{(2)}\right\},\quad \lim_{n_1\to\infty}P\left\{\bigcap_{n\geq n_1}\mathcal{A}_{n}^{(3)}\right\},\quad \lim_{n_1\to\infty}P\left\{\bigcap_{n\geq n_1}\mathcal{A}_{n}^{(4)}\right\}.$$

(b) *Let $Conv(\beta_0)$ be the convex hull of $m(X_i,\beta_0)$. For any $\varepsilon>0$, there exist K_ε and N_ε such that $P\left\{\mathcal{A}_\varepsilon^*\right\}\geq 1-\varepsilon$, where*

$$\mathcal{A}_\varepsilon^*=\bigcap_{n\geq N_\varepsilon}\left\{\mathcal{A}_{K_\varepsilon,n}^{(1)}\cap\mathcal{A}_{\varepsilon,n}^{(2)}\cap\mathcal{A}_{n}^{(3)}\cap\mathcal{A}_{n}^{(4)}\right\}, \tag{5.11}$$

and for $n > N_\varepsilon$ and $\boldsymbol{X} \in \mathcal{A}^*_\varepsilon$, $P\{0 \in int\,(Conv_n(\beta_0)) \mid \boldsymbol{X} \in \mathcal{A}^*_\epsilon\} \geq 1-\varepsilon$, where in general, $int(\cdot)$ denotes the interior of a set.

Now, let us go back to the EL problem, which is essentially solving

$$\begin{cases} \max_{\boldsymbol{p}} \prod_{i=1}^n p_i, \qquad \text{where } \boldsymbol{p} = (p_1, \ldots, p_n), \\ \text{subject to} \quad p_i > 0, \ \sum_{i=1}^n p_i = 1, \ \sum_{i=1}^n p_i m(X_i, \beta) = 0, \end{cases} \tag{5.12}$$

for any fixed β. Given β, by the concavity of $\sum_{i=1}^n \ln(p_i)$ in $\boldsymbol{p}$, a unique maximum exists provided that 0 is in the interior of $\mathrm{Conv}_n(\beta)$, the convex hull of $m(X_1, \beta), \ldots, m(X_n, \beta)$. By Lemma 5.1.1, for any ε, there exists N_ε such that for any $n > N_\varepsilon$ and $\boldsymbol{X}_\infty \in \mathcal{A}_\varepsilon$,

$$P\{0 \in int\,(\mathrm{Conv}(\beta_0)) \mid \boldsymbol{X} \in \mathcal{A}^*_\epsilon\} > 1 - \varepsilon.$$

Thus for $n > N_\varepsilon$, with $\beta = \beta_0$, there exists a unique solution $\hat{\boldsymbol{p}}_0$ to the maximization problem in (5.12) with probability greater than $1 - \varepsilon$.

To conclude that the unique maximizer over $\boldsymbol{p}$'s for fixed $\beta = \beta_0$ is a calculus maximizer, we need to show that there is a ball of dimension $n-1-p$ for $\boldsymbol{p}$'s within which to take the derivative, where p is the dimension of $m(X_1, \beta_0)$. Let $\mathbf{1}$ be an $1 \times n$ vector with all entries equal to 1, $\mathbf{0}$ be a $p \times 1$ vector with all entries equal to 0, $M_0 = (m(X_1, \beta_0), \ldots, m(X_n, \beta_0))$ be a $p \times n$ matrix, and $M = (\mathbf{1}^{tr}, M_0^{tr})^{tr}$ be a $(p+1) \times n$ matrix. Lemma 5.1.1 guarantees the existence and uniqueness of solutions to maximization problem (5.12). Use $\hat{\boldsymbol{p}}_0$ to denote that solution, then $\hat{\boldsymbol{p}}_0$

belongs to

$$\mathcal{P} \equiv \left\{ \boldsymbol{p} : \ M\boldsymbol{p} = (1, \mathbf{0}^{tr})^{tr}, \quad \text{all} \ \ p_i > 0, \sum_{i=1}^{n} p_i = 1 \right\} \tag{5.13}$$

For any $\boldsymbol{v}$ in the null space of M and $\alpha \in \mathbb{R}$, define

$$\boldsymbol{p} = \frac{1}{1 + \alpha \boldsymbol{v} \cdot \mathbf{1}} \cdot (\hat{\boldsymbol{p}}_0 + \alpha \boldsymbol{v}),$$

we know that $M\boldsymbol{p} = (1, \mathbf{0}^{tr})^{tr}$, hence for small α such that the entries for $\boldsymbol{p}$ are all positive, we know that $\boldsymbol{p}$ belongs to $\mathcal{P}$. Furthermore, $\boldsymbol{v}$ is in the null space of M, and $\text{rank}(M) = p + 1$, so $\dim(\mathcal{P}) = n - p - 1$. Observing that $\boldsymbol{X}_\infty \in \mathcal{A}^{(2)}_{\varepsilon,n}$ for $\varepsilon < \lambda_0$, there is a relative open set $\mathcal{P}$ of dimension $n - p - 1$ within which we can take derivative.

Next, we give the form of the unique solution to (5.12) when $\beta = \beta_0$ and $n > N_\varepsilon$, using Lagrange multipliers $\lambda \in \mathbb{R}$ and $t \in \mathbb{R}^p$. Define $G(\boldsymbol{p}, \lambda, t)$ as

$$G = \sum_{i=1}^{n} \ln p_i - \lambda \left\{ \sum_{i=1}^{n} p_i - 1 \right\} - nt^{tr} \sum_{i=1}^{n} p_i m(X_i, \beta_0)$$

To maximize the concave function G, differentiate G with respect to p_i, $i = 1, \ldots, n$, then set the derivatives to be zero,

$$\frac{\partial G}{\partial p_i} = \frac{1}{p_i} - \lambda - nt^{tr} m(X_i, \beta_0) = 0, \qquad i = 1, \ldots, n. \tag{5.14}$$

Multiply the equations above by p_i, then add them together,

$$n - \lambda \sum_{i=1}^{n} p_i - nt^{tr} \sum_{i=1}^{n} p_i m(X_i, \beta_0) = 0 \tag{5.15}$$

By the constraints in (5.12), (5.15) implies that $n - \lambda = 0$, so $n = \lambda$. Therefore from equation (5.14),

$$\hat{p}_i = \frac{1}{n \left\{ 1 + t_0^{tr} m(X_i, \beta_0) \right\}} \tag{5.16}$$

Since $\sum_{i=1}^{n} \hat{p}_i m(X_i, \beta_0) = \mathbf{0}$, the vector t must solve the following equation,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{m(X_i, \beta_0)}{1 + t^{tr} m(X_i, \beta_0)} = \mathbf{0}. \tag{5.17}$$

Important Event In Lemma 5.1.1, we proved that for any ε, there exists N_ε such that $P\left\{\mathcal{A}_\varepsilon^*\right\} \geq 1 - \varepsilon$, where

$$\mathcal{A}_\varepsilon^* = \bigcap_{n \geq N_\varepsilon} \left\{ \mathcal{A}_{K_{\varepsilon,n}}^{(1)} \cap \mathcal{A}_{\varepsilon,n}^{(2)} \cap \mathcal{A}_n^{(3)} \cap \mathcal{A}_n^{(4)} \right\}.$$

Furthermore, let $\Sigma(\beta, \beta_0) = E_{\beta_0}\left\{m(X_1, \beta)^{\otimes 2}\right\}$. Since $\Sigma(\beta, \beta_0)$ is continuous, and

$$\Sigma = E_{\beta_0}\left\{m(X_1, \beta_0)^{\otimes 2}\right\} = \Sigma(\beta_0, \beta_0)$$

is positive definite, there exists U_{β_0}, a neighbourhood of β_0 such that for all β in U_{β_0}, matrix $\Sigma(\beta, \beta_0)$ is positive definite, i.e.,

$$U_{\beta_0} \equiv \left\{\beta : \Sigma(\beta, \beta_0) \text{ is positive definite}\right\}. \tag{5.18}$$

Then, consider the following class of functions,

$$\mathcal{G} \equiv \left\{m(x, \beta) : \beta \in U_{\beta_0}\right\},$$

In Chapter 4, we assumed

(A.11) Let $\alpha = (x, \beta) \in \mathbb{R}^d \times \Theta$ with Euclidean norm, and $k(x, y, \beta) = k_\alpha(y)$. There exists a measurable function $b(y) : \mathbb{R}^d \mapsto \mathbb{R}$ such that for any $\alpha_1 \neq \alpha_2$,

$$|k_{\alpha_1}(y) - k_{\alpha_2}(y)| \leq b(y)\|\alpha_1 - \alpha_2\|,$$

Since $m(x,\beta)$ is Lipschitz continuous with respect to β, by Example 19.7 on page 271 of [41], we know that $\mathcal{G}$ is a Donsker class, which implies that

$$\sup_{\beta\in U_{\beta_0}}\left\|\frac{1}{n}\sum_{i=1}^{n}m(X_i,\beta)^{\otimes 2}-\Sigma(\beta,\beta_0)\right\|_2\xrightarrow{a.s.}0,$$

therefore $P\left\{\mathcal{A}_{\varepsilon,n}^{(5)}\right\}\geq 1-\varepsilon$, where

$$\mathcal{A}_{\varepsilon,n}^{(5)}=\left\{\boldsymbol{X}:\sup_{\beta\in U_{\beta_0}}\left\|\frac{1}{n}\sum_{i=1}^{n}m(X_i,\beta)^{\otimes 2}-\Sigma\right\|_2\leq\frac{\varepsilon}{2}\right\},$$

Let λ_0 be the smallest eigenvalue of Σ, and $0<\varepsilon<2\lambda_0/3$. For Lemma 5.1.2 through Theorem 5.1.1, we restrict our discussion to $\boldsymbol{X}\in\mathcal{A}_\varepsilon^+$, where the set $\mathcal{A}_\varepsilon^+=\mathcal{A}_\varepsilon^*\bigcap\mathcal{A}_{\varepsilon,n}^{(5)}$ and $P\left\{\mathcal{A}_\varepsilon^+\right\}>1-2\varepsilon$.

The following lemma states the uniquness of $t=t(\beta)$ for $\beta\in U_{\beta_0}$ and $\boldsymbol{X}\in A_\varepsilon^+$. The proof is given by contradiction using Rolle's theorem.

Lemma 5.1.2 *Under the assumptions of Lemma 5.1.1, and (A.11), for $\boldsymbol{X}\in\mathcal{A}_\varepsilon^+$ and any fixed β in set U_{β_0} defined in (5.18), if there exists a solution t to equation $g(t,\boldsymbol{X},\beta)=0$, then the solution must be unique a.s., where*

$$g(t,\boldsymbol{X},\beta)=\frac{1}{n}\sum_{i=1}^{n}\frac{m(X_i,\beta)}{1+t^{tr}m(X_i,\beta)},$$

Note that the unique solution of (5.12), denoted by $\hat{p}_i$ in (5.16), must lie in $(0,1)$, therefore $1+t^{tr}m(X_i,\beta)>1/n$, for all i. Define

$$D_\beta(\boldsymbol{X})=\left\{t:\ \text{for all } i,\ 1+t^{tr}m(X_i,\beta)>\frac{1}{n}\right\}. \tag{5.19}$$

Our goal is to show that for any fixed $\beta\in U_{\beta_0}$, there is a unique solution $t(\beta)$ to the maximization problem (5.12).

Theorem 5.1.1 *Under the assumptions of Lemma 5.1.1, for any $\beta \in U_{\beta_0}$ and for any $\boldsymbol{X} \in \mathcal{A}_\varepsilon^+$, there exists a unique solution $t(\beta) \in D_\beta(\boldsymbol{X})$ to the maximization problem (5.12).*

5.1.2 Discontinuous Criterion Function

When the criterion function $m(x, \beta)$ is no longer continuous with respect to β, many methods we used in the the continuous case no longer applies, for example, Rolle's theorem in Lemma 5.1.2 and the continuation method in Theorem 5.1.1. Therefore we seek different ways to attain the existence and uniqueness of solutions to maximization problem (5.1) for β in some neighbourhood of β_0. We prove a more general version of the "zero in the convex hull" theorem in the previous section, namely, we demonstrate that the statement is true not only for convex hull $\mathrm{Conv}(\beta_0)$, but also true for $\mathrm{Conv}(\beta)$ with probability approaching 1, when β belongs to some neighbourhood of β_0 that is to be specified, i.e.,

$$\mathbf{0} \in \mathrm{Conv}(\beta), \quad \text{for } \beta \in U^*_{\beta_0} \tag{5.20}$$

If we can prove (5.20), then for any $\beta \in U^*_{\beta_0}$, the maximization problem (5.1) is guaranteed to have a unique solution for any $\beta \in U^*_{\beta_0}$.

To reach this goal, instead of assigning $m(X_i, \beta)$ with the constant probability mass $1/n$ in Lemma 5.1.1, we construct a random probability vector as follows. Let W_i be i.i.d. random variables, $i = 1, \ldots, n$ that follow an exponential distribution

with $E(W_1) = 1$, and also assume that $\{W_i\}_{i=1}^n$ is independent of $\{X_i\}_{i=1}^n$. Let

$$q(x, \beta) = \frac{dP_\beta(x)}{dP_{\beta_0}(x)} \tag{5.21}$$

be the Radon-Nikodym derivative of P_β with respect to P_{β_0}, and define

$$W_i^* = W_i q(X_i, \beta), \quad i = 1, \ldots, n \tag{5.22}$$

Then

$$\begin{aligned} E\left\{W_1^*\right\} &= \iint w \frac{dP_\beta(x)}{dP_{\beta_0}(x)} dP_{\beta_0}(x) dP_W(w) = \int w \left\{ \int dP_\beta(x) \right\} dP_W(w) \\ &= \int w dP_W(w) = 1 \end{aligned}$$

and

$$E\left\{W_1^* | X_1\right\} = E\left\{W_1 q(X_1, \beta) | X_1\right\} = q(X_1, \beta) \tag{5.23}$$

Define $V = (V_1, \ldots, V_n)$ be a vector in the simplex Δ_n defined in (5.72), where

$$V_i = \frac{W_i^*}{\sum\limits_{i=1}^n W_i^*}. \tag{5.24}$$

We would like to show that $\mathbf{0} \in in\left(\mathrm{Conv}_n(\beta)\right)$ with weights in (5.23). Assume that

(A.12) Assume that $\Sigma_2(\beta, \beta_0)$ and $\Sigma_3(\beta, \beta_0)$ are two continuous functions given by

$$\Sigma_2(\beta, \beta_0) = E_{\beta_0}\left\{q^2(X_1, \beta) m(X_1, \beta)^{\otimes 2}\right\} \tag{5.25}$$

$$\Sigma_3(\beta, \beta_0) = E_{\beta_0}\left\{q^3(X_1, \beta) \|m(X_1, \beta)\|^3\right\} \tag{5.26}$$

with $\Sigma_2(\beta_0, \beta_0) = E_{\beta_0}\left\{m(X_1, \beta_0)^{\otimes 2}\right\}$ and $\Sigma_3(\beta, \beta_0) = E_{\beta_0}\left\{\|m(X_1, \beta_0)\|^3\right\}$

Since $\Sigma_2(\beta, \beta_0)$ and $\Sigma_3(\beta, \beta_0)$ are both continuous with respect to β, we define

$$U_{\beta_0}^* \equiv \left\{\beta : \Sigma_2(\beta, \beta_0) \text{ is positive definite, and } \Sigma_3(\beta, \beta_0) \text{ is bounded}\right\}. \tag{5.27}$$

Now we are ready to establish the "zero in the convex hull" theorem for discontinuous estimating function. The following lemma is a generalization of Lemma 5.1.1, which essentially states $0 \in \mathrm{Conv}(\beta)$ for all $\beta \in U^*_{\beta_0}$.

Lemma 5.1.3 *Assume that (5.4)-(5.6) in Lemma 5.1.1, and (A.12) are satisfied. For $\beta \in U^*_{\beta_0}$ defined in (5.27), for any constant $K > 0$, let $\boldsymbol{X} = \{X_i\}_{i=1}^{\infty}$, t_0 be a unit vector in $\mathbb{R}^p$.*

(a) Let $\lambda_2(\beta, \beta_0)$ be the smallest eigenvalue of $\Sigma_2(\beta, \beta_0)$. Define

$$\mathcal{B}^{(1)}_{K,n} = \left\{ \boldsymbol{X} : \inf_{t_0} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} q(X_i, \beta) m(X_i, \beta) \geq -K \right\}, \tag{5.28}$$

$$\mathcal{B}^{(2)}_{\varepsilon,n} = \left\{ \boldsymbol{X} : \left\| \frac{1}{n} \sum_{i=1}^{n} q(X_i, \beta) m(X_i, \beta)^{\otimes 2} - \Sigma^*(\beta, \beta_0) \right\|_2 \leq \varepsilon \right\}, \tag{5.29}$$

$$\mathcal{B}^{(3)}_{n} = \left\{ \boldsymbol{X} : \frac{1}{n} \sum_{i=1}^{n} \| q(X_i, \beta) m(X_i, \beta) \|^3 \leq 2 \cdot \Sigma_3(\beta, \beta_0) \right\}, \tag{5.30}$$

$$\mathcal{B}^{(4)}_{n} = \left\{ \boldsymbol{X} : \frac{1}{n} \sum_{i=1}^{n} q(X_i, \beta) m(X_i, \beta)^{\otimes 2} \geq \frac{1}{2} \lambda_0^*(\beta, \beta_0) I \right\}. \tag{5.31}$$

Then for any ε, there exists K_ε and N_ε, such that

(i) for any $n \geq N_\varepsilon$, $P\left\{ \mathcal{B}^{(1)}_{K_\varepsilon,n} \right\} \geq 1 - \varepsilon$;

(ii) the following limits are all identical to 1,

$$\lim_{n_1 \to \infty} P\left\{ \bigcap_{n \geq n_1} \mathcal{B}^{(2)}_{\varepsilon,n} \right\} = \lim_{n_1 \to \infty} P\left\{ \bigcap_{n \geq n_1} \mathcal{B}^{(3)}_{n} \right\} = \lim_{n_1 \to \infty} P\left\{ \bigcap_{n \geq n_1} \mathcal{B}^{(4)}_{n} \right\}.$$

(b) *Let $Conv_n(\beta)$ be the convex hull of $m(X_i, \beta)$. For any $\varepsilon > 0$, there exist K_ε and N_ε such that $P\{\mathcal{B}^*_\varepsilon\} \geq 1 - \varepsilon$, where*

$$\mathcal{B}^*_\varepsilon = \bigcap_{n \geq N_\varepsilon} \left\{ \mathcal{B}^{(1)}_{K_\varepsilon,n} \cap \mathcal{B}^{(2)}_{\varepsilon,n} \cap \mathcal{B}^{(3)}_{n} \cap \mathcal{B}^{(4)}_{n} \right\}, \tag{5.32}$$

and for $n > N_\varepsilon$ and $\boldsymbol{X} \in \mathcal{B}^*_\varepsilon$, $P\{0 \in Conv^{\circ}_n(\beta) \,|\, \boldsymbol{X}_\infty \in \mathcal{B}^*_\varepsilon\} \geq 1 - \varepsilon$.

5.2 Extended Estimating Equation

In this section, we extend the conclusions in the previous section when the constraints of the EL maximization is an EEE. In Section 5.2.1, we prove the uniqueness of solutions to the EL maximization in (5.1) for $m_n(x, \boldsymbol{x}, \beta)$ that is continuous with respect to β. In Section 5.2.2, we discuss the case when $m_n(x, \boldsymbol{x}, \beta)$ is no longer continuous with respect to β.

5.2.1 Continuous Criterion Function

We can make the same conclusion for $m_n(X_i, \boldsymbol{X}, \beta_0)$ if we can show that (a) in the Lemma 5.1.1 is true. Note that the conclusions of (5.7) and (5.9) are guaranteed by the asymptotic normality of $n^{-1/2} \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0)$ and the order of $\sum_{i=1}^{n} \|m_n(X_i, \boldsymbol{X}, \beta_0)\|^3$, respectively, which has been proved in Lemma 4.2.2 and Proposition 4.2.2. As for the conclusion regarding (5.8) and (5.10), it suffices to show

$$\frac{1}{n} \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0)^{\otimes 2} \xrightarrow{P} \Sigma \tag{5.33}$$

We will discuss the conditions for (5.33) in the following proposition, then state and prove a result for continuous $m_n(x, \boldsymbol{x}, \beta)$ that is parallel to Lemma 5.1.1.

Proposition 5.2.1 *Let Y_i, $i = 1, 2$ be random variables that are i.i.d. as X_1, and are independent of $\boldsymbol{X} = (X_1, \ldots, X_n)$. Under assumptions (A.2), (A.4), (A.9), and (A.10),*

$$\frac{1}{n} \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0)^{\otimes 2} \xrightarrow{P} \Sigma, \tag{5.34}$$

where

$$\Sigma = E\left\{Q(Y_1,\beta_0)^2\left[C(Y_1) - \frac{E\left\{C(Y_2)k(Y_1,Y_2,\beta_0)|Y_1\right\}}{E\left\{k(Y_1,Y_2,\beta_0)|Y_1\right\}}\right]^{\otimes 2}\right\}. \tag{5.35}$$

Next, we present a result that is parallel to Lemma 5.1.1 for extended estimating equations.

Lemma 5.2.1 *For i.i.d. random variables $\boldsymbol{X}_n = (X_1,\ldots,X_n)$, assume that the estimating function $m_n(X_i,\boldsymbol{X},\beta_0)$ satisfies (A.2), (A.9) and (A.10). For any constant $K>0$, let $\boldsymbol{X}_\infty = \{X_i\}_{i=1}^\infty$, and t_0 be a unit vector in $\mathbb{R}^p$.*

(a) *Let λ_0 be the smallest eigenvalue of Σ, and $I \in \mathbb{R}^{p\times p}$ be the identity matrix, define*

$$\mathcal{C}_{K,n}^{(1)} = \left\{\boldsymbol{X}_\infty : \inf_{t_0} \frac{1}{\sqrt{n}}\sum_{i=1}^n t_0^{tr} m_n(X_i,\boldsymbol{X},\beta_0) \geq -K\right\}, \tag{5.36}$$

$$\mathcal{C}_{\varepsilon,n}^{(2)} = \left\{\boldsymbol{X}_\infty : \left\|\frac{1}{n}\sum_{i=1}^n m_n(X_i,\boldsymbol{X},\beta_0)^{\otimes 2} - \Sigma\right\|_2 \leq \varepsilon\right\}, \tag{5.37}$$

$$\mathcal{C}_{n}^{(3)} = \left\{\boldsymbol{X}_\infty : \frac{1}{n}\sum_{i=1}^n \|m_n(X_i,\boldsymbol{X},\beta_0)\|^3 \leq C\right\}, \tag{5.38}$$

$$\mathcal{C}_{n}^{(4)} = \left\{\boldsymbol{X}_\infty : \frac{1}{n}\sum_{i=1}^n m_n g(X_i,\boldsymbol{X},\beta_0)^{\otimes 2} \geq \frac{1}{2}\lambda_0 I\right\}. \tag{5.39}$$

where C in (5.38) is a constant in $\mathbb{R}$ that is greater than $E\|m_n(X_1,\boldsymbol{X},\beta_0)\|^3$. Then for any ε, there exists K_ε and N_ε, such that

(i) for any $n \geq N_\varepsilon$, $P\left\{\mathcal{C}_{K_\varepsilon,n}^{(1)}\right\} \geq 1-\varepsilon$;

(ii) the following limits are all identical to 1,

$$\lim_{n_1\to\infty} P\left\{\bigcap_{n\geq n_1}\mathcal{C}_{\varepsilon,n}^{(2)}\right\} = \lim_{n_1\to\infty} P\left\{\bigcap_{n\geq n_1}\mathcal{C}_{n}^{(3)}\right\} = \lim_{n_1\to\infty} P\left\{\bigcap_{n\geq n_1}C_{n}^{(4)}\right\}.$$

(b) *Let* $\mathrm{Conv}(\beta_0)$ *be the convex hull of* $m_n(X_i, \boldsymbol{X}, \beta_0)$*. For any* $\varepsilon > 0$*, there exist* K_ε *and* N_ε *such that* $P\{\mathcal{C}_\varepsilon^*\} \geq 1 - \varepsilon$*, where*

$$\mathcal{C}_\varepsilon^* = \bigcap_{n \geq N_\varepsilon} \left\{ \mathcal{C}_{K_\varepsilon,n}^{(1)} \cap \mathcal{C}_{\varepsilon,n}^{(2)} \cap \mathcal{C}_n^{(3)} \cap \mathcal{C}_n^{(4)} \right\}, \tag{5.40}$$

and for $n > N_\varepsilon$ *and* $\boldsymbol{X}_\infty \in C_\varepsilon^*$*,* $P\{0 \in Conv_n(\beta_0)^{\mathrm{o}}\} \geq 1 - \varepsilon$*.*

Now, let us go back to the EL problem, which for any fixed β, is essentially the following maximization problem

$$\begin{cases} \max_{\boldsymbol{p}} \prod_{i=1}^n p_i, \qquad \text{where } \boldsymbol{p} = (p_1, \ldots, p_n) \\ \text{subject to} \quad p_i \geq 0, \ \sum_{i=1}^n p_i = 1, \ \sum_{i=1}^n p_i m_n(X_i, \boldsymbol{X}, \beta) = 0 \end{cases} \tag{5.41}$$

Note that for a given β, a unique maximum exists provided that 0 is in the interior of set $\mathrm{Conv}(\beta)$, the convex hull of $m_n(X_1, \boldsymbol{X}, \beta), \ldots, m_n(X_n, \boldsymbol{X}, \beta)$. Since by Lemma 5.2.1, for any ε, there exists N_ε such that for any $n > N_\varepsilon$ and $\boldsymbol{X}_\infty \in \mathcal{C}_\varepsilon$, $P\{0 \in \mathrm{Conv}^{\mathrm{o}}(\beta_0)\} > 1 - \varepsilon$, we know that for $n > N_\varepsilon$, there exist a unique solution denoted by $\hat{\boldsymbol{p}}_0$ to the maximization problem in (5.41) with $\beta = \beta_0$, with probability greater than $1 - \varepsilon$.

To know that the unique maximizer is a calculus maximizer, we need to know that there is a ball of dimension $n - 1 - p$ for $\boldsymbol{p}$'s within which to take the derivative, where p is the dimension of $m_n(X_1, \boldsymbol{X}, \beta)$. Let $\mathbf{1}$ be an $1 \times n$ vector with all entries equal to 1, and $\mathbf{0}$ be a $p \times 1$ vector with all entries equal to 0,

$$M_0 = (m_n(X_1, \boldsymbol{X}, \beta), \ldots, m_n(X_n, \boldsymbol{X}, \beta))$$

be a $p \times n$ matrix, and $M = (\mathbf{1}^{tr}, M_0^{tr})^{tr}$ be a $(p+1) \times n$ matrix. Suppose there exists one solution to the maximization problem (5.41) denoted by $\hat{\boldsymbol{p}}_0$. Then $\hat{\boldsymbol{p}}_0$ belongs to

$$\mathcal{P} \equiv \left\{ \boldsymbol{p} : \ M\boldsymbol{p} = (1, \mathbf{0}^{tr})^{tr}, \quad \text{all} \ \ p_i > 0, \sum_{i=1}^{n} p_i = 1 \right\} \tag{5.42}$$

For any $\boldsymbol{v}$ in the null space of M and $\alpha \in \mathbb{R}$, define

$$\boldsymbol{p} = \frac{1}{1 + \alpha \boldsymbol{v} \cdot \mathbf{1}} \cdot (\hat{\boldsymbol{p}}_0 + \alpha \boldsymbol{v}),$$

we know that $M\boldsymbol{p} = (1, \mathbf{0}^{tr})^{tr}$, hence for small α such that the entries for $\boldsymbol{p}$ are all positive, we know that $\boldsymbol{p}$ belongs to $\mathcal{P}$. Furthermore, $\boldsymbol{v}$ is in the null space of M, and $\dim(M) = p + 1$, so $\dim(\mathcal{P}) = n - p - 1$. Therefore we know that there is a relative open set $\mathcal{P}$ of dimension $n - p - 1$ within which we can take derivative.

Next, we give the form of the unique solution to (5.41) when $\beta = \beta_0$ and $n > N_\varepsilon$, using Lagrange multipliers $\lambda \in \mathbb{R}$ and $t \in \mathbb{R}^p$. Define $G(\boldsymbol{p}, \lambda, t)$ as

$$G = \sum_{i=1}^{n} \ln p_i - \lambda \left\{ \sum_{i=1}^{n} p_i - 1 \right\} - nt^{tr} \sum_{i=1}^{n} p_i m_n(X_i, \boldsymbol{X}, \beta_0)$$

To maximize the concave function G, differentiate G with respect to p_i, $i = 1, \ldots, n$, then set the derivatives to be zero,

$$\frac{\partial G}{\partial p_i} = \frac{1}{p_i} - \lambda - nt^{tr} m_n(X_i, \boldsymbol{X}, \beta_0) = 0, \qquad i = 1, \ldots, n. \tag{5.43}$$

Multiply the equations above by p_i, then add them together,

$$n - \lambda \sum_{i=1}^{n} p_i - nt^{tr} \sum_{i=1}^{n} p_i m_n(X_i, \boldsymbol{X}, \beta_0) = 0 \tag{5.44}$$

By the constraintts in (5.41), (5.44) indicates that $n - \lambda = 0$, so $n = \lambda$. Therefore from equation (5.43),

$$\hat{p}_i = \frac{1}{n \left\{ 1 + t_0^{tr} m_n(X_i, \boldsymbol{X}, \beta_0) \right\}} \tag{5.45}$$

Since $\sum_{i=1}^{n} \hat{p}_i m_n(X_i, \boldsymbol{X}, \beta_0) = \mathbf{0}$, the vector t must solve the following equation,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{m_n(X_i, \boldsymbol{X}, \beta_0)}{1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta_0)} = \mathbf{0}. \tag{5.46}$$

Important Event In Lemma 5.2.1, we proved that for any ε, there exists N_ε such that $P\{\mathcal{C}_\varepsilon^*\} \geq 1 - \varepsilon$, where

$$\mathcal{C}_\varepsilon^* = \bigcap_{n \geq N_\varepsilon} \left\{ \mathcal{C}_{K_{\varepsilon,n}}^{(1)} \cap \mathcal{C}_{\varepsilon,n}^{(2)} \cap \mathcal{C}_n^{(3)} \cap \mathcal{C}_n^{(4)} \right\}.$$

Furthermore, let

$$\Sigma(\beta, \beta_0) = E\left\{ Q(Y_1, \beta)^2 \left[C(Y_1) - \frac{E\{C(Y_2)k(Y_1, Y_2, \beta)|Y_1\}}{E\{k(Y_1, Y_2, \beta)|Y_1\}} \right]^{\otimes 2} \right\}. \tag{5.47}$$

Since $\Sigma(\beta, \beta_0)$ is continuous, and by (5.35),

$$\Sigma = E\left\{ Q(Y_1, \beta_0)^2 \left[C(Y_1) - \frac{E\{C(Y_2)k(Y_1, Y_2, \beta_0)|Y_1\}}{E\{k(Y_1, Y_2, \beta_0)|Y_1\}} \right]^{\otimes 2} \right\} = \Sigma(\beta_0, \beta_0)$$

is positive definite, there exists U_{β_0}, a neighbourhood of β_0 such that for all β in U_{β_0}, matrix $\Sigma(\beta, \beta_0)$ is positive definite, i.e.,

$$U_{\beta_0} \equiv \{\beta : \Sigma(\beta, \beta_0) \text{ is positive definite}\}. \tag{5.48}$$

Then, consider the following class of functions,

$$\mathcal{G} \equiv \{m_n(X_i, \boldsymbol{X}, \beta) : \beta \in U_{\beta_0}\},$$

By Example 19.7 on Page 271 of [41] and Lemma 4.2.1, we know that family $\mathcal{G}$ is a Donsker class, which implies that

$$\sup_{\beta \in U_{\beta_0}} \left\| \frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)^{\otimes 2} - \Sigma(\beta, \beta_0) \right\|_2 \xrightarrow{a.s.} 0,$$

therefore $P\left\{\mathcal{C}_{\varepsilon,n}^{(5)}\right\} \geq 1-\varepsilon$, where

$$\mathcal{C}_{\varepsilon,n}^{(5)} = \left\{\boldsymbol{X} : \sup_{\beta \in U_{\beta_0}} \left\| \frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)^{\otimes 2} - \Sigma \right\|_2 \leq \frac{\varepsilon}{2} \right\},$$

Let λ_0 be the smallest eigenvalue of Σ, and $0 < \varepsilon < \lambda_0$. For Lemma 5.2.2 through Theorem 5.2.1, we restrict our discussion to $\boldsymbol{X}_\infty \in \mathcal{C}_\varepsilon^+$, where the set

$$\mathcal{C}_\varepsilon^+ = \mathcal{C}_\varepsilon^* \bigcap \mathcal{C}_{\varepsilon,n}^{(5)} \tag{5.49}$$

and $P\left\{\mathcal{C}_\varepsilon^+\right\} > 1-2\varepsilon$.

Lemma 5.2.2 *Under the assumptions of Lemma 4.2.1 and Lemma 5.1.1, for $\boldsymbol{X}_\infty \in \mathcal{C}_\varepsilon^+$ and any β in set U_{β_0} defined in (5.48), if there exists a solution t to equation $g(t, \boldsymbol{X}, \beta) = 0$, then the solution must be unique, where*

$$g(t, \boldsymbol{X}, \beta) = \frac{1}{n}\sum_{i=1}^{n} \frac{m_n(X_i, \boldsymbol{X}, \beta)}{1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta)}, \tag{5.50}$$

Note that the unique solution of (5.41), denoted by $\hat{p}_i$ in (5.45), must lie in $(0,1)$, therefore $1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta_0) > 1/n$, for all i. Define

$$D_\beta(\boldsymbol{X}) = \left\{ t : \text{ for all } i,\ 1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta) > \frac{1}{n} \right\}. \tag{5.51}$$

Our goal is to show that for any fixed $\beta \in U_{\beta_0}$, there is a unique solution $t(\beta)$ to the maximization problem (5.41).

Theorem 5.2.1 *Under assumptions of Lemma 5.2.1, for any $\beta \in U_{\beta_0}$ and $\boldsymbol{X} \in \mathcal{C}_\varepsilon^+$, there exists a unique solution $t(\beta)$ to the maximization problem (5.41).*

5.2.2 Discontinuous Criterion Function

Similar to the strategy with used in the proof under CEE settings, when the criterion function $m_n(X_i, \boldsymbol{X}, \beta)$ is no longer continuous with respect to β, we use random variables instead of the fixed constant $1/n$ as the probability mass assigned to each $m_n(X_i, \boldsymbol{X}, \beta)$, $i = 1, \ldots, n$. More specifically, let $P_{\beta_0}(x)$ be the cumulative density function of X_1, and $q(x, \beta)$ be the Radon-Nicodym derivative of $P_\beta(x)$ with respect to $P_{\beta_0}(x)$, i.e.,

$$q(x, \beta) = \frac{dP_\beta(x)}{dP_{\beta_0}(x)}, \tag{5.52}$$

Let $Y_1 \sim P_{\beta_0}$ be independent of $\boldsymbol{X}_\infty = \{X_i\}_{i=1}^\infty$, and

$$V(X_i, \beta) = Q(X_i, \beta) \left\{ C(X_i) - \frac{E\left\{C(Y_1)k(X_i, Y_1, \beta)|X_i\right\}}{E\left\{k(X_i, Y_1, \beta)|X_i\right\}} \right\} \tag{5.53}$$

We claim that $E\left\{V(X_i, \beta_0)\right\} = 0$ because,

$$\begin{aligned} 0 &= E\left\{m_n(X_i, \boldsymbol{X}, \beta_0)\right\} = \lim_{n\to\infty} E\left\{m_n(X_i, \boldsymbol{X}, \beta_0)\right\} \\ &= E\left\{\lim_{n\to\infty} m_n(X_i, \boldsymbol{X}, \beta_0)\right\} = E\left\{E\left\{\lim_{n\to\infty} m_n(X_i, \boldsymbol{X}, \beta_0)|X_1\right\}\right\} \\ &= E\left\{V(X_1, \beta_0)\right\}, \end{aligned} \tag{5.54}$$

therefore

$$E_{\beta_0}\left\{q(X_1, \beta)V(X_1, \beta)\right\} = \int V(x, \beta) \cdot \frac{dP_\beta(x)}{dP_{\beta_0}(x)} dP_{\beta_0}(x) = 0.$$

Assume that

$$E_{\beta_0}\left\{q^2(X_1, \beta)\right\} < \infty, \qquad \text{for all } \beta \in U_{\beta_0}, \tag{5.55}$$

then by the CLT, for any $t_0 \in \mathbb{R}^p$ with $\|t_0\| = 1$ and $\Sigma^*(\beta, , \beta_0) = E_{\beta_0}\left\{q^2(X_1)V(X_1, \beta)^{\otimes 2}\right\}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n q(X_i) t_0^{tr} V(X_i, \beta) \xrightarrow{D} N(0, t_0^{tr} \Sigma^*(\beta, , \beta_0) t_0) \tag{5.56}$$

At this point, we would like to point two facts. First, by (5.54), $V(X_i, \beta)$ is eligible as a criterion function of CEE, and based on the discussion in the previous sections, there exists U_{β_0} such that $0 \in \text{Conv}(\beta)$, for all $\beta \in U_{\beta_0}$. Another heuristic observation is that $m_n(X_i, \boldsymbol{X}, \beta)$ and $V(X_i, \beta)$ are "close" in some sense because they only distinct by the quotient term, and by LLN and Slutsky's lemma, the quotient term in $m_n(X_i, \boldsymbol{X}, \beta)$ converges to that in $V(X_i, \beta)$ in probability.

However, in order to pass the desirable feature of $V(X_i, \beta)$ to $m_n(X_i, \boldsymbol{X}, \beta)$, for any fixed $\beta \in U_{\beta_0}$, we need $n^{-1/2}\|m_n(x, \boldsymbol{X}) - V(x, \beta)\|$ to be bounded uniform in x, i.e.,

$$\sup_{x \in \mathcal{X}} \|m_n(x, \boldsymbol{X}, \beta) - V(x, \beta)\| = O_P(n^{-1/2}). \tag{5.57}$$

By Example 19.11 in [41] by Van der Vaart, if we know that

(i) *function $k(x, \cdot, \cdot)$ has bounded variation*

(ii) *for any $\varepsilon > 0$ and $\delta_\varepsilon > 0$, let $\mathcal{X}_\varepsilon \equiv \{x : E\{k(x, X, \beta)\} > \delta_\varepsilon\}$, then for any constant $c_\varepsilon > 0$, there exists N_ε such that for any $n > N_\varepsilon$,*

$$P\left\{\left\|\frac{1}{\sqrt{n}} \sum_{i: X_i \in \mathcal{X}_\varepsilon^c} m_n(X_i, \boldsymbol{X}, \beta)\right\| \geq c_\varepsilon\right\} < \varepsilon, \tag{5.58}$$

then (5.57) is guaranteed.

Since up to now, the broadest class of semiparametric models for which the structural parameter can be defined via EEE is the φ-transformation model discussed in Chapter 3, for the rest of this chapter, we restrict our attention to the $m_n(X_i, \boldsymbol{X}, \beta)$ under the φ-transformation model assumptions.

Recall that under the φ-transformation model assumption, $k(x,\cdot,\cdot)$ is an indicator function, therefore (ii) on page 100 is satisfied. Let us consider a martingale assumption that guarantees (ii) in the previous lemma. Assume that $X_i = (T_i, C_i, Z_i)$, where the lifetime T_i and the right censoring variable C_i are independent conditional on covariate Z_i. Recall that under the φ-transformation model assumption,

$$T_i = \varphi(\varepsilon_i, \beta_0^{tr} Z_i), \tag{5.59}$$

where $\varphi(x,\cdot)$ is strictly increasing, ε_i is the residual with distribution function $F(x)$ and hazard rate $\lambda(x)$. Let ζ_i be the residual when lifetime is censored, i.e., $C_i = \varphi(\zeta_i, \beta_0^{tr} Z_i)$. Use the classic method in [21] and construct the compensated counting process martingales

$$M_i\left\{\varphi(u, \beta_0^{tr} Z_i)\right\} = N_i\left\{\varphi(u, \beta_0^{tr} Z_i)\right\} - \int_{-\infty}^{u} \lambda(x) Y_i\left\{\varphi(x, \beta_0^{tr} Z_i)\right\} dx \tag{5.60}$$

is a martingale with respect to measure $\mathcal{P}_{\beta_0}$ and filtration

$$\mathcal{F}_n(u) = \sigma(X_i, N_i(\varphi(s, \beta_0^{tr} Z_i)), Y_i(\varphi(s, \beta_0^{tr} Z_i)) : \ s \le u, \ i = 1, \ldots, n)$$

where $V_i = \min(T_i, C_i)$ is the observed time, $\Delta_i = I\{T_i \le C_i\}$ is the non-censored indicator, $N_i(x) = I\{T_i \le x, \Delta_i = 1\}$, $Y_i(x) = I\{V_i \ge x\}$ is the at-risk indicator,

$$Q(X_i, \beta) = \Delta_i, \qquad C(X_i) = C(Z_i), \qquad k(X_i, X_j, \beta) = Y_j(\varphi(\varepsilon_i, \beta_0^{tr} Z_j)), \tag{5.61}$$

Then we can write

$$\begin{aligned} S_n(\beta_0) &= \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0) \\ &= \sum_{i=1}^{n} Q(X_i, \beta) \left\{ C(X_i) - \frac{\sum_{j=1}^{n} C(X_j) k(X_i, X_j, \beta)}{\sum_{j=1}^{n} k(X_i, X_j, \beta)} \right\} \\ &= \sum_{i=1}^{n} \Delta_i \left\{ C(Z_i) - \frac{\sum_{j=1}^{n} C(Z_j) Y_j\left\{\varphi(\varepsilon_i, \beta_0^{tr} Z_j)\right\}}{\sum_{j=1}^{n} Y_j\left\{\varphi(\varepsilon_i, \beta_0^{tr} Z_j)\right\}} \right\} \end{aligned} \tag{5.62}$$

By (5.60), together with the model assumption (5.59), on the residual scale,

$$\begin{aligned} S_n(\beta_0) &= \sum_{i=1}^{n} \int dN_i \left\{\varphi(u, \beta_0^{tr} Z_i)\right\} \left\{C(Z_i) - \bar{C}(u, \beta_0)\right\} \\ &= \sum_{i=1}^{n} \int dM_i \left\{\varphi(u, \beta_0^{tr} Z_i)\right\} \left\{C(Z_i) - \bar{C}(u, \beta_0)\right\} \end{aligned} \tag{5.63}$$

where

$$\bar{C}(u, \beta_0) = \frac{\sum_{j=1}^{n} C(Z_j) Y_j \left\{\varphi(u, \beta_0^{tr} Z_j)\right\}}{\sum_{j=1}^{n} Y_j \left\{\varphi(u, \beta_0^{tr} Z_j)\right\}}$$

Note that $E_{\beta_0}\left\{S_n(\beta_0)\right\} = 0$ due to the martingale property. Recall that the predictable variation process $\langle dM_i, dM_i \rangle = \lambda(u) Y_i \left\{\varphi(u, \beta_0^{tr} Z_i)\right\} du$ as mentioned in [1], hence

$$Var\left\{S_n(\beta_0)\right\} = E\left\{ \int \sum_{i=1}^{n} \left\{C(Z_i) - \bar{C}(u, \beta_0)\right\}^2 \lambda(u) P\left\{V_i \geq \varphi(u, \beta_0^{tr} Z_i)\right\} du \;\middle|\; \boldsymbol{Z} \right\}$$

Under assumption (A.2), $\|C(x)\| < b$, for all $x \in \mathcal{X}$, therefore

$$\left\| Var\left\{n^{-1/2} S_n(\beta_0)\right\} \right\| \leq \frac{4b^2}{n} \sum_{i=1}^{n} \int \lambda(u) S(u) du \leq 4b^2.$$

Now that we have established (5.57), for any $\varepsilon > 0$, there exists C_ε and N_ε such that for any $n > N_\varepsilon$,

$$\sup_x \|m_n(x, \boldsymbol{X}, \beta) - V(X_i, \beta)\| \leq \frac{C_\varepsilon}{\sqrt{n}}, \text{ w.p. no less than } 1 - \varepsilon$$

therefore $P\left\{\mathcal{C}_{\varepsilon,n}^{(6)}\right\} > 1 - \varepsilon$, where

$$\mathcal{C}_{\varepsilon,n}^{(6)} \equiv \left\{ \boldsymbol{X} : \sup_x \|m_n(x, \boldsymbol{X}, \beta) - V(X_i, \beta)\| \leq \frac{K_\varepsilon}{\sqrt{n}} \right\}, \tag{5.64}$$

Let $q(x, \beta)$ be he Radon-Nikodym derivative defined in (5.52), and as we have assumed in (5.55), $E\left\{q(X_1, \beta)^2\right\} < \infty$, for any $\beta \in U_{\beta_0}$. Then by the law of large

numbers, any fixed $\beta \in U_{\beta_0}$, we know that $P\left\{\mathcal{C}_{\varepsilon,n}^{(7)}\right\} > 1-\varepsilon$, where

$$\mathcal{C}_{\varepsilon,n}^{(7)} \equiv \left\{ \boldsymbol{X} : \frac{1}{n}\sum_{i=1}^{n} q(X_i,\beta)^2 < E_{\beta_0}\left\{q(X_1,\beta)^2\right\} \right\}, \tag{5.65}$$

Now, we can update the definition of C_ε^+ in (5.49) by including $C_{\varepsilon,n}^{(6)}$ and $C_{\varepsilon,n}^{(7)}$, namely,

$$\mathcal{C}_\varepsilon^+ \equiv \mathcal{C}_\varepsilon^* \bigcap \left\{ \bigcap_{i=5}^{7} \mathcal{C}_{\varepsilon,n}^{(i)} \right\}, \tag{5.66}$$

and $P\left\{\mathcal{C}_\varepsilon^+\right\} > 1-4\varepsilon$ for $n > N_\varepsilon$.

Theorem 5.2.2 *Under assumptions of (5.57), for any $\beta \in U_{\beta_0}$ and $\boldsymbol{X}_\infty \in \mathcal{C}_\varepsilon^+$, 0 belongs to int $(Conv(\beta))$.*

5.3 Some Proofs

5.3.1 Proofs Under the CEE Setting

Let us state two widely used lemmas without proof. The Lyapunov Central Limit Theorem was found in Chapter 27 of [6], and the Hyperplane Separation Theorem was found in Chapter 2 of [7].

Lemma 5.3.1 (Lyapunov Central Limit Theorem) *Suppose $Z_1, \ldots, Z_n$ are independent random variables, each with finite expected value μ_i and variance σ_i^2. Define*

$$s_n^2 = \sum_{i=1}^{n} \sigma_i^2$$

If for some $\delta > 0$, the Lyapunovs condition

$$\frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} E\big[\, |Z_i - \mu_i|^{2+\delta} \,\big] \to 0 \tag{5.67}$$

as $n \to \infty$ is satisfied, then a sum of $(Z_i - \mu)/s_n$ converges in distribution to a standard normal random variable, as n goes to infinity:

$$\frac{1}{s_n} \sum_{i=1}^{n} (Z_i - \mu_i) \xrightarrow{D} N(0,\ 1). \tag{5.68}$$

Lemma 5.3.2 (Hyperplane Separation Theorem) *Let A and B be two disjoint nonempty convex sets. If A is open, then there exist a nonzero vector v and real number c such that*

$$\langle x, v \rangle > c \ \text{ and } \langle y, v \rangle \leq c$$

for all x in A and y in B.

In general, for two matrices A and B in $\mathbb{R}^{p\times p}$

$$A > B \;\Leftrightarrow A - B \text{ is positive definite.}$$

Proof of Lemma 5.1.1 First, let us show that (a) is true. Under assumptions (5.4)-(5.6), by law of large numbers, we know that (ii) is true. Note that by the multivariate central limit theorem,

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(X_i, \beta_0) \xrightarrow{D} S_\infty \sim N(0, \Sigma). \tag{5.69}$$

Since S_∞ follows a multivariate normal distribution, for any $\varepsilon > 0$ and any unit vector $t_0 \in \mathbb{R}^p$, there exists K_ε such that

$$P\left\{\inf_{t_0} t_0^{tr} S_\infty < -K_\varepsilon\right\} < \frac{\varepsilon}{2}. \tag{5.70}$$

By (5.69), there exists N_ε such that for any $n > N_\varepsilon$,

$$\left| P\left\{\inf_{t_0} t_0^{tr} S_n < -K_\varepsilon\right\} - P\left\{\inf_{t_0} t_0^{tr} S_\infty < -K_\varepsilon\right\} \right| < \frac{\varepsilon}{2} \tag{5.71}$$

Combining (5.70) and (5.71), we conclude that $P\left\{\inf_{t_0} t_0^{tr} S_n < -K_\varepsilon\right\} < \varepsilon$, hence the conclusion for (i) is also true.

Next, we prove (b) using Lemma 5.3.2. Let $\mathrm{Conv}_n(\beta_0)$ be the convex hull of $m(X_i, \beta_0)$, then we want to show that the two sets

$$A = \{0\}, \quad B = int\left(\mathrm{Conv}_n(\beta_0)\right) = \left\{\sum_{i=1}^{n} v_i m(X_i, \beta_0); v \in \Delta_n\right\}$$

are not separated, where

$$\Delta_n = \left\{v = (v_1, \ldots, v_n) : \; \sum_{i=1}^{n} v_i = 1, v_i \in (0, 1)\right\}. \tag{5.72}$$

By Lemma 5.3.2, the statement $0 \in \mathrm{Conv}_n^{\mathrm{o}}(\beta_0)$ is equivalent to

$$\begin{aligned} &\forall t_0 \in \mathbb{R}^p \text{ with } \|t_0\| = 1 \text{ and } \forall a \leq 0, \\ &\qquad\qquad \exists z \in int\,(\mathrm{Conv}_n(\beta_0)) \text{ such that } t_0^{tr} z > a. \end{aligned} \tag{5.73}$$

To prove (5.73), we will show in the following paragraphs that for any $\varepsilon > 0$, nonpositive constant $a = -c/\sqrt{n}$ and vector $t_0 \in \mathbb{R}^p$ with $\|t_0\| = 1$, there exists N_ε such that for $n > N_\varepsilon$, given $\boldsymbol{X} \in \mathcal{A}_\varepsilon^*$

$$\inf_{t_0} P\left\{ t_0^{tr} \sum_{i=1}^n V_i m(X_i, \beta_0) \geq -\frac{c}{\sqrt{n}} \;\middle|\; \boldsymbol{X} \right\} > 0, \quad \text{for } \boldsymbol{X} \in \mathcal{A}_\varepsilon^*, \tag{5.74}$$

where we consider $V = v$ as a continuously distributed random vector in symplex denoted by Δ_n, and z in (5.73) for each t_0 is a value $\sum_{i=1}^n V_i(\omega) m(X_i, \beta_0)$ for ω in the event where the probability is positive in (5.74). Note that (5.74) is true if we can prove the case when $a = 0$, namely, for $n > N_\varepsilon$,

$$\inf_{t_0} P\left\{ t_0^{tr} \sum_{i=1}^n V_i m(X_i, \beta_0) \geq 0 \;\middle|\; \boldsymbol{X} \right\} > 0, \quad \text{for } \boldsymbol{X} \in \mathcal{A}_\varepsilon^* \tag{5.75}$$

We prove (5.75) by constructing V in the following way. Let W_i be i.i.d. random variables that follow an exponential distribution with mean that equals to 1, and let $V = (V_1, \ldots, V_n)$ with

$$V_i = \frac{W_i}{\sum_{i=1}^n W_i}. \tag{5.76}$$

Let $\overline{W} = \sum_{i=1}^{n} W_i/n$ and t_0 be a unit vector in $\mathbb{R}^p$, then the probability in (5.75) can be written as

$$\begin{aligned}
&P\left\{ t_0^{tr} \sum_{i=1}^{n} V_i m(X_i, \beta_0) \geq 0 \;\middle|\; \boldsymbol{X} \right\} \\
=&P\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} W_i m(X_i, \beta_0) \geq 0 \;\middle|\; \boldsymbol{X} \right\} \\
=&P\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} m(X_i, \beta_0) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} (W_i - 1) m(X_i, \beta_0) \geq 0 \;\middle|\; \boldsymbol{X} \right\} \\
=&P\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} (W_i - 1) m(X_i, \beta_0) \geq -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} m(X_i, \beta_0) \;\middle|\; \boldsymbol{X} \right\}
\end{aligned} \tag{5.77}$$

Combining (5.77) and the conclusion (i) in (a), we obtain for $n > N_\varepsilon$, any nonnegative constant K_ε, and $\boldsymbol{X} \in \mathcal{A}_\varepsilon^*$

$$\begin{aligned}
&P\left\{ t_0^{tr} \sum_{i=1}^{n} V_i m(X_i, \beta_0) \geq 0 \;\middle|\; \boldsymbol{X} \right\} \\
\geq\; &P\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} (W_i - 1) m(X_i, \beta_0) \geq K_\varepsilon \;\middle|\; \boldsymbol{X} \right\}
\end{aligned} \tag{5.78}$$

Next, we show that the term $n^{-1/2} \sum_{i=1}^{n} t_0^{tr} (W_i - 1) m(X_i, \beta_0)$ in (5.78) satisfies (5.67) in Lemma 5.3.1, with the choice of $\delta = 1$, $Z_i = t_0^{tr}(W_i - 1) m(X_i, \beta_0)$, and

$$\mu_i = E\left\{ (W_i - 1) \cdot t_0^{tr} m(X_i, \beta_0) | X_i \right\} = 0;$$

$$\sigma_i^2 = var\left\{ (W_i - 1) \cdot t_0^{tr} m(X_i, \beta_0) | X_i \right\} = t_0^{tr} m(X_i, \beta_0)^{\otimes 2} t_0.$$

Let $s_n^2 = t_0^{tr} \left\{ \sum_{i=1}^{n} m(X_i, \beta_0)^{\otimes 2} \right\} t_0$, then the left hand side of (5.67) becomes

$$\begin{aligned}
&\frac{1}{s_n^3} \sum_{i=1}^{n} E\left\{ \|(W_i - 1) \cdot t_0^{tr} m(X_i, \beta)\|^3 | X_i \right\} \\
\leq &\left(\frac{2}{n^{-3/2} s_n^3} \right) \cdot \left(\frac{1}{n} \sum_{i=1}^{n} \|m(X_i, \beta_0)\|^3 \right) \cdot \left(\frac{1}{\sqrt{n}} \right)
\end{aligned} \tag{5.79}$$

By Lemma 5.3.1, given $\boldsymbol{X} \in \mathcal{A}_\varepsilon^*$, it follows that $n^{-1/2}\sum_{i=1}^{n} t_0^{tr}(W_i - 1)m(X_i, \beta_0)$ is asymptotically normal, i.e., for any unit vector $s \in \mathbb{R}^p$,

$$n^{-1/2}\sum_{i=1}^{n}(W_i - 1)\cdot s^{tr}m(X_i, \beta_0) \xrightarrow{D} N\left(0, s^{tr}\Sigma s\right), \quad \text{with probability 1.} \tag{5.80}$$

Note that by the strong law of large numbers, $\overline{W}$ converges to $EW_1 = 1$ almost surely, therefore for $n > N_\varepsilon$, constant $K_\varepsilon > 0$, and $\boldsymbol{X} \in \mathcal{A}_\varepsilon^*$

$$P\left\{\inf_s \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(W_i - 1)s^{tr}m(X_i, \beta_0)\right] \geq K_\varepsilon \;\middle|\; \boldsymbol{X}\right\} > 0. \tag{5.81}$$

Combining (5.77), (5.78) and (5.81), we get that for any $n > N_\varepsilon$, constant $a = -c/\sqrt{n}$, unit vector $t_0 \in \mathbb{R}^p$, and $\boldsymbol{X} \in B_\varepsilon^*$,

$$\begin{aligned} & P\left\{t_0^{tr}\sum_{i=1}^{n} V_i m(X_i, \beta_0) \geq 0 \;\middle|\; \boldsymbol{X}\right\} \\ \geq\; & P\left\{\inf_{s\in\mathbb{R}^p:\|s\|=1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(W_i - 1)s^{tr}m(X_i, \beta_0) \geq K_\varepsilon \;\middle|\; \boldsymbol{X}\right\} > 0, \end{aligned} \tag{5.82}$$

and hence we complete the proof. □

Lemma 5.3.3 (Implicit Function Theorem) *Let $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$ be a continuously differentiable function, and let $\mathbb{R}^{n+m}$ have coordinates (x, y). Fix a point $(a, b) = (a_1, \ldots, a_n, b_1, \ldots, b_m)$ with $f(a, b) = c$, where $c \in \mathbb{R}^m$. If the Jacobian matrix $J_{f,y}(a, b) = [(\partial f_i/\partial y_j)(a, b)]$is invertible, then there exists an open set U containing a, an open set V containing b, and a unique continuously differentiable function $g : U \to V$ such that*

$$\{(x, g(x))|x \in U\} = \{(x, y) \in U \times V|f(x, y) = c\}.$$

Proof of Lemma 5.1.2 For ε smaller than λ_0, the smallest eigenvalue of Σ, on the event $\mathcal{A}_\varepsilon^+$, for any fixed $\beta \in U_{\beta_0}$,

$$\left\| \frac{1}{n} \sum_{i=1}^{n} m(X_i, \beta)^{\otimes 2} - \Sigma(\beta, \beta_0) \right\|_2 < \frac{\varepsilon}{2}$$

Since $\beta \in U_{\beta_0}$, we know that $\Sigma(\beta, \beta_0)$ is positive definite, hence

$$\frac{1}{n} \sum_{i=1}^{n} m(X_i, \beta)^{\otimes 2} \text{ is positive definite for } \beta \in U_{\beta_0}. \tag{5.83}$$

By contradiction, we can show that the conclusion is true for any fixed $\beta \in U_{\beta_0}$. Suppose that for a fixed $\beta \in U_{\beta_0}$, there exist distinct t_1 and t_2 such that

$$g(t_1, \boldsymbol{X}, \beta) = g(t_2, \boldsymbol{X}, \beta) = 0.$$

Therefore by Rolle's theorem from [15], there exists $s \in (0, 1)$ such that for $t_3 = st_1 + (1-s)t_2$, $\partial g(t_3, \boldsymbol{X}, \beta)/\partial s = 0$, i.e.,

$$(t_1 - t_2)^{tr} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{m(X_i, \beta)^{\otimes 2}}{(1 + t_3^{tr} m(X_i, \beta))^2} \right\} (t_1 - t_2) = 0, \tag{5.84}$$

which implies

$$\frac{1}{n} \sum_{i=1}^{n} m(X_i, \beta)^{\otimes 2} \quad \text{is singular,}$$

contradicting (5.83). □

Proof of Theorem 5.1.1 Define $\mathcal{C}$ in the following way:

$$\begin{aligned} \mathcal{C} \equiv \{\beta^* \in U_{\beta_0} : \exists\, r_{\beta^*} > 0 \text{ such that for any } \beta \in B_{r_{\beta^*}}(\beta^*), \\ \text{there exists } t = t(\beta) \text{ such that } m(t, \boldsymbol{X}, \beta) = 0\}, \end{aligned} \tag{5.85}$$

where

$$g(t, \boldsymbol{X}, \beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{m(X_i, \beta)}{1 + t^{tr} m(X_i, \beta)}. \tag{5.86}$$

First, we show that for any fixed β^* such that the equation $g(t, \boldsymbol{X}, \beta^*) = 0$ has a solution $t = t^*$, then $\beta^* \in \mathcal{C}$. By (5.86), $g(t^*, \boldsymbol{X}, \beta^*) = 0$ implies $\mathbf{0}$ belongs to $int\,(\mathrm{Conv}_n(\beta^*))$, consequently, there exists a unique solution to the maximizing problem (5.12) with the choice of $\beta = \beta^*$. Since this unique solution is bounded by 0 and 1, i.e.,

$$\frac{1}{n} \cdot \frac{1}{1 + t^{*tr} m(X_i, \beta^*)} \in (0, 1) \qquad \text{for } i = 1, \ldots, n,$$

we know that $t^* \in D_{\beta^*}(\boldsymbol{X})$. Since $g(t^*, \boldsymbol{X}, \beta^*) = 0$, $\partial g(t, \boldsymbol{X}, \beta)/\partial t|_{(t^*, \beta^*)}$ is negative definite. By the implicit function theorem, there exists r_{β^*} such that for $\beta \in B_{r_{\beta^*}}(\beta^*)$, equation $g(t, \boldsymbol{X}, \beta) = 0$ has a solution $t = t(\beta)$. Therefore $t^* \in \mathcal{C}$.

In particular, since we have shown in (5.16) that when $\beta = \beta_0$, there exists t_0 such that $g(t_0, \boldsymbol{X}, \beta_0) = 0$, we know that the conclusion in the previous paragraph is true for there is $\beta = \beta_0$, i.e., $\beta_0 \in \mathcal{C}$.

Next, we show that $\mathcal{C} = U_{\beta_0}$, i.e., for any $\beta^* \in U_{\beta_0}$, β^* also belongs to $\mathcal{C}$. Let

$$\rho_{\max} = \sup \left\{ s : \beta_0 + r(\beta^* - \beta_0) \in \mathcal{C}, \text{ for all } 0 < r < s \right\}, \tag{5.87}$$

By contradiction, we can show that $\rho_{\max} \geq 1$. Otherwise, suppose $\rho_{\max} < 1$. By equation (5.87), we know that

$$\beta = \beta_0 + r(\beta^* - \beta_0) \in \mathcal{C}, \qquad \text{for all } r < \rho_{\max} \tag{5.88}$$

Let $\{r_k\}_{k=1}^{\infty}$ be a series of increasing positive numbers that are bounded by $\rho_{\max}$, and

$$\beta_k = \beta_0 + r_k(\beta^* - \beta_0), \qquad \text{for } 0 < r_k < \rho_{\max}. \tag{5.89}$$

Therefore by (5.88), $\beta_k \in \mathcal{C}$, which together with (5.85) - (5.87) implies for each β_k,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{m(X_i,\beta_k)}{1+t(\beta_k)^{tr}m(X_i,\beta_k)}=0 \qquad \text{and} \qquad 1+t^{tr}(\beta_k)m(X_i,\beta_k)\geq\frac{1}{n} \tag{5.90}$$

Let

$$\beta_{\max}=\lim_{k\to\infty}\beta_k=\beta_0+\rho_{\max}(\beta^*-\beta_0) \tag{5.91}$$

then by taking the limit of (5.90) as k goes to ∞,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{m(X_i,\beta_{\max})}{1+t(\beta_{\max})^{tr}m(X_i,\beta_{\max})}=0. \tag{5.92}$$

Therefore $\mathbf{0}$ belongs to $int\,(\mathrm{Conv}_n(\beta_{\max}))$. Consequently, there exists a unique solution to the maximizing problem (5.12) with $\beta=\beta_{\max}$. Since the unique solution is in $(0,1)$, we know that $t(\beta_{\max})\in D_{\beta_{\max}}(\boldsymbol{X})$. Similar to the previous proof, $\beta_{\max}\in\mathcal{C}$.

By the definition of $\mathcal{C}$, there exists $r_{\max}$ such that for all $\beta\in B_{r_{\max}}(\beta_{\max})$, we know that $t=t(\beta)$ solves $g(t,\boldsymbol{X},\beta)=0$. Therefore $\beta=\beta_0+r(\beta^*-\beta_0)\in\mathcal{C}$, where

$$r=\rho_{\max}+r_{\max}/4>\rho_{\max}$$

contradicts the definition of $\rho_{\max}$, hence $\rho_{\max}\geq 1$ and we conclude that $\beta^*\in\mathcal{C}$. $\square$

Proof of Lemma 5.1.3 First, let us show that (a) is true. By the LLN, we know that statement (ii) is true. Note that

$$E\left\{q(X_1,\beta)m(X_1,\beta)\right\}=\int m(x,\beta)\frac{dP_\beta(x)}{dP_{\beta_0}(x)}dP_{\beta_0}(x)=0,$$

by the multivariate central limit theorem,

$$S_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}q(X_i,\beta)m(X_i,\beta)\xrightarrow{D_{\beta_0}}S_\infty\sim N(0,\Sigma^*(\beta,\beta_0)). \tag{5.93}$$

Since S_∞ follows a multivariate normal distribution, for any $\varepsilon > 0$ and any unit vector $t_0 \in \mathbb{R}^p$, there exists K_ε such that

$$P\left\{\inf_{t_0} t_0^{tr} S_\infty < -K_\varepsilon\right\} < \frac{\varepsilon}{2}. \tag{5.94}$$

By (5.93), there exists N_ε such that for any $n > N_\varepsilon$,

$$\left| P\left\{\inf_{t_0} t_0^{tr} S_n < -K_\varepsilon\right\} - P\left\{\inf_{t_0} t_0^{tr} S_\infty < -K_\varepsilon\right\} \right| < \frac{\varepsilon}{2} \tag{5.95}$$

Combining (5.94) and (5.95), we conclude that $P\left\{\inf_{t_0} t_0^{tr} S_n < -K_\varepsilon\right\} < \varepsilon$, hence the conclusion for (i) is also true.

Next, we prove (b) using Lemma 5.3.2. Let $\mathrm{Conv}_n(\beta)$ be the convex hull of $m(X_i, \beta)$, then we want to show that the two sets

$$A = \{0\}, \quad B = \mathrm{Conv}_n(\beta) = \left\{\sum_{i=1}^n v_i m(X_i, \beta); v \in \Delta_n\right\}$$

are not separated, where

$$\Delta_n = \left\{v = (v_1, \ldots, v_n) : \ \sum_{i=1}^n v_i = 1, v_i \in (0, 1)\right\}.$$

By Lemma 5.3.2, the statement $0 \in int\left(\mathrm{Conv}_n(\beta)\right)$ is equivalent to

$$\begin{aligned} &\forall t_0 \in \mathbb{R}^p \text{ with } \|t_0\| = 1 \text{ and } \forall a \le 0, \\ &\qquad\qquad \exists z \in int\left(\mathrm{Conv}_n(\beta)\right) \text{ such that } t_0^{tr} z > a. \end{aligned} \tag{5.96}$$

To prove (5.96), we will show in the following paragraphs that for any $\varepsilon > 0$, nonpositive constant $a = -c/\sqrt{n}$ and vector $t_0 \in \mathbb{R}^p$ with $\|t_0\| = 1$, there exists N_ε such that for $n > N_\varepsilon$, given $\boldsymbol{X} \in \mathcal{B}_\varepsilon^*$

$$\inf_{t_0} P\left\{t_0^{tr} \sum_{i=1}^n V_i m(X_i, \beta) \ge -\frac{c}{\sqrt{n}} \,\middle|\, \boldsymbol{X}\right\} > 0, \quad \text{for } \boldsymbol{X} \in \mathcal{B}_\varepsilon^*, \tag{5.97}$$

where we consider $V = v$ as a continuously distributed random vector in Δ_n, and z in equation (5.96) for each t_0 is a value $\sum_{i=1}^{n} V_i(\omega)m(X_i, \beta)$ for ω in the event where the probability is positive in (5.97). Note that (5.97) is true if we can prove the case when choosing $a = 0$, namely, for $n > N_\varepsilon$,

$$\inf_{t_0} P\left\{ t_0^{tr} \sum_{i=1}^{n} V_i m(X_i, \beta) > 0 \;\middle|\; \boldsymbol{X} \right\} > 1 - \varepsilon, \qquad \text{for } \boldsymbol{X} \in \mathcal{B}_\varepsilon^* \tag{5.98}$$

We prove (5.98) using V defined in (5.24). Let $\overline{W^*} = \sum_{i=1}^{n} W_i^*/n$ and t_0 be a unit vector in $\mathbb{R}^p$, then the probability in (5.98) can be written as

$$\begin{aligned}
&P\left\{ t_0^{tr} \sum_{i=1}^{n} V_i m(X_i, \beta) \geq 0 \;\middle|\; \boldsymbol{X} \right\} \\
=&P\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} W_i^* m(X_i, \beta) \geq 0 \;\middle|\; \boldsymbol{X} \right\} \\
=&P\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} q(X_i, \beta) m(X_i, \beta) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} \left\{ W_i^* - q(X_i, \beta) \right\} m(X_i, \beta) \geq 0 \;\middle|\; \boldsymbol{X} \right\} \\
=&P\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} \left\{ W_i^* - q(X_i, \beta) \right\} m(X_i, \beta) \geq -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} q(X_i, \beta) m(X_i, \beta) \;\middle|\; \boldsymbol{X} \right\},
\end{aligned} \tag{5.99}$$

Combining (5.99) and conclusion (i) in (a), we obtain for $n > N_\varepsilon$, any nonnegative constant c and K_ε, and $\boldsymbol{X} \in \mathcal{B}_\varepsilon^*$

$$\begin{aligned}
&P\left\{ t_0^{tr} \sum_{i=1}^{n} V_i m(X_i, \beta) \geq 0 \;\middle|\; \boldsymbol{X} \right\} \\
\geq\; &P\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} t_0^{tr} \left\{ W_i^* - q(X_i, \beta) \right\} m(X_i, \beta) \geq K_\varepsilon \;\middle|\; \boldsymbol{X} \right\}
\end{aligned} \tag{5.100}$$

Next, we show that the term $n^{-1/2} \sum_{i=1}^{n} t_0^{tr} \left\{ W_i^* - q(X_i, \beta) \right\} m(X_i, \beta)$ in (5.100) satisfies equation (5.67) in Lemma 5.3.1, with the choice of $\delta = 1$,

$$Z_i = t_0^{tr} \left\{ W_i^* - q(X_i, \beta) \right\} m(X_i, \beta),$$

by (5.23)

$$\mu_i = E\left\{\{W_i^* - q(X_i,\beta)\}\cdot t_0^{tr} m(X_i,\beta)|X_i\right\} = 0;$$

$$\sigma_i^2 = var\left\{\{W_i^* - q(X_i,\beta)\}\cdot t_0^{tr} m(X_i,\beta)|X_i\right\} = t_0^{tr} q(X_i,\beta)^2 m(X_i,\beta)^{\otimes 2} t_0.$$

Let $s_n^2 = t_0^{tr}\left\{\sum_{i=1}^n q(X_i,\beta)^2 m(X_i,\beta)^{\otimes 2}\right\} t_0$, then the left hand side of (5.67) becomes

$$\begin{aligned}&\frac{1}{s_n^3}\sum_{i=1}^n E\left\{\|\{W_i^* - q(X_i,\beta)\}\cdot t_0^{tr} m(X_i,\beta)\|^3|X_i\right\}\\ \le &\left(\frac{2}{n^{-3/2}s_n^3}\right)\cdot\left(\frac{1}{n}\sum_{i=1}^n \|q(X_i,\beta)m(X_i,\beta)\|^3\right)\cdot\left(\frac{1}{\sqrt{n}}\right)\end{aligned} \tag{5.101}$$

By Lemma 5.3.1, given $\boldsymbol{X}\in\mathcal{B}_\varepsilon^*$, it follows that $n^{-1/2}\sum_{i=1}^n t_0^{tr}\{W_i^* - q(X_i,\beta)\}m(X_i,\beta)$ is asymptotically normal, i.e., for any unit vector $s\in\mathbb{R}^p$,

$$n^{-1/2}\sum_{i=1}^n \{W_i^* - q(X_i,\beta)\}\cdot s^{tr} m(X_i,\beta) \xrightarrow{D_{\beta_0}} N\left(0, s^{tr}\Sigma^*(\beta,\beta_0)s\right), \tag{5.102}$$

with probability 1. Note that by the strong law of large numbers, as n goes to infinity, $\overline{W^*}$ converges to $EW^*{}_1 = 1$ almost surely, therefore for any $n > N_\varepsilon$, positive constant K_ε, and $\boldsymbol{X}\in\mathcal{B}_\varepsilon^*$

$$P\left\{\inf_s\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \{W_i^* - q(X_i,\beta)\}s^{tr}m(X_i,\beta)\right]\ge K_\varepsilon \,\middle|\, \boldsymbol{X}\right\} > 0. \tag{5.103}$$

Combining (5.99), (5.100) and (5.103), we get that for any $n > N_\varepsilon$, constant $a = -c/\sqrt{n}$, unit vector $t_0\in\mathbb{R}^p$, and $\boldsymbol{X}\in\mathcal{B}_\varepsilon^*$,

$$\begin{aligned}&P\left\{t_0^{tr}\sum_{i=1}^n V_i m(X_i,\beta)\ge 0 \,\middle|\, \boldsymbol{X}\right\}\\ \ge\ &P\left\{\inf_{s\in\mathbb{R}^p:\|s\|=1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \{W_i^* - q(X_i,\beta)\}s^{tr}m(X_i,\beta)\ge K_\varepsilon \,\middle|\, \boldsymbol{X}\right\} > 0,\end{aligned} \tag{5.104}$$

and hence we complete the proof. □

5.3.2 Proofs Under the EEE Setting

In order to prove Proposition 5.2.1, we first prove the following lemma.

Lemma 5.3.4 *Let $\{U_{i,j}\}_{i,j=1}^n$ and $\{V_i\}_{i=1}^n$ be p-dimensional random vectors, and $\{Z_{i,j}\}_{i,j=1}^n$ and $\{W_i\}_{i=1}^n$ be random variables with finite means and variances, respectively. Suppose that $W_i > 0$ a.s., and that as $n \to \infty$,*

$$\sup_i \frac{1}{n}\sum_{j=1}^n (U_{i,j} - V_i) \xrightarrow{P} 0 \quad \textit{and} \quad \sup_i \frac{1}{n}\sum_{j=1}^n (Z_{i,j} - W_i) \xrightarrow{P} 0, \tag{5.105}$$

If for any i, there exists a constant $b \in \mathbb{R}$ such that

$$\frac{\left\|\sum_{j=1}^n U_{i,j}\right\|}{\left|\sum_{j=1}^n Z_{i,j}\right|} \le b \quad \textit{and} \quad \frac{\|V_i\|}{|W_i|} \le b \tag{5.106}$$

then for any i, as $n \to \infty$,

$$\left\|\frac{\sum_{j=1}^n U_{i,j}}{\sum_{j=1}^n Z_{i,j}} - \frac{V_i}{W_i}\right\| \xrightarrow{P} 0. \tag{5.107}$$

Proof: First, by the triangle inequality and (5.106), we know that for any i, the left hand side of (5.107) must be bounded; also, note that it equates to

$$\left\|\frac{W_i \sum_{j=1}^n (U_{i,j} - V_i)/n - V_i \sum_{j=1}^n (Z_{i,j} - W_i)/n}{W_i \sum_{j=1}^n Z_{i,j}/n}\right\|,$$

which is bounded by

$$\left\|\frac{\sum_{j=1}^n (U_{i,j} - V_i)/n}{\sum_{j=1}^n Z_{i,j}/n}\right\| + \left\|\frac{V_i}{W_i}\right\| \cdot \left\|\frac{\sum_{j=1}^n (Z_{i,j} - W_i)/n}{\sum_{j=1}^n Z_{i,j}/n}\right\| \tag{5.108}$$

Moreover, notice that by the law of large numbers, for any $i = 1, 2, \ldots,$

$$\frac{1}{n}\sum_{j=1}^n Z_{i,j} \xrightarrow{P} W_i, \quad \text{a.s.}, \tag{5.109}$$

so the conclusion follows by (5.105), (5.106) and (5.109). □

Proof of Proposition 5.2.1 Note that we can write the left hand side of (5.34) as

$$\frac{1}{n}\sum_{i=1}^{n} Q(X_i,\beta_0)^2 \left\{ C(X_i) - \frac{\sum_{j=1}^{n} C(X_j)k(X_i,X_j,\beta_0)}{\sum_{j=1}^{n} k(X_i,X_j,\beta_0)} \right\}^{\otimes 2} \tag{5.110}$$

$$= \frac{1}{n}\sum_{i=1}^{n} Q(X_i,\beta_0)^2 C(X_i)^{\otimes 2} \tag{5.111}$$

$$- \frac{1}{n}\sum_{i=1}^{n} Q(X_i,\beta_0)^2 \frac{\sum_{j=1}^{n} C(X_i)C(X_j)^{tr}k(X_i,X_j,\beta_0)}{\sum_{j=1}^{n} k(X_i,X_j,\beta_0)} \tag{5.112}$$

$$- \frac{1}{n}\sum_{i=1}^{n} Q(X_i,\beta_0)^2 \frac{\sum_{j=1}^{n} C(X_j)C(X_i)^{tr}k(X_i,X_j,\beta_0)}{\sum_{j=1}^{n} k(X_i,X_j,\beta_0)} \tag{5.113}$$

$$+ \frac{1}{n}\sum_{i=1}^{n} Q(X_i,\beta_0)^2 \frac{\left\{\sum_{j=1}^{n} C(X_i)k(X_i,X_j,\beta_0)\right\}^{\otimes 2}}{\left\{\sum_{j=1}^{n} k(X_i,X_j,\beta_0)\right\}^{2}}, \tag{5.114}$$

and the right hand side of (5.34) as

$$\Sigma = E\left\{ Q(Y_1,\beta_0)^2 \left[C(Y_1) - \frac{E\left\{C(Y_2)k(Y_1,Y_2,\beta_0)|Y_1\right\}}{E\left\{k(Y_1,Y_2,\beta_0)|Y_1\right\}} \right]^{\otimes 2} \right\}$$

$$= E\left\{Q(Y_1,\beta_0)^2 C(Y_1)^{\otimes 2}\right\} \tag{5.111'}$$

$$- E\left\{ Q(Y_1,\beta_0)^2 \frac{E\left\{C(Y_1)C(Y_2)^{tr}k(Y_1,Y_2,\beta_0)|Y_1\right\}}{E\left\{k(Y_1,Y_2,\beta_0)|Y_1\right\}} \right\} \tag{5.112'}$$

$$- E\left\{ Q(Y_1,\beta_0)^2 \frac{E\left\{C(Y_2)C(Y_1)^{tr}k(Y_1,Y_2,\beta_0)|Y_1\right\}}{E\left\{k(Y_1,Y_2,\beta_0)|Y_1\right\}} \right\} \tag{5.113'}$$

$$- E\left\{ Q(Y_1,\beta_0)^2 \frac{\left[E\left\{C(Y_2)^{tr}k(Y_1,Y_2,\beta_0)|Y_1\right\}\right]^{\otimes 2}}{\left[E\left\{k(Y_1,Y_2,\beta_0)|Y_1\right\}\right]^{2}} \right\} \tag{5.114'}$$

By law of large numbers, (5.111) converges to (5.111') in probability. Now, let us show that (5.112) converges to (5.112') in probability. Note that by adding and subtracting the term

$$Q(X_i,\beta_0)^2 \frac{E\left\{C(X_i)C(Y_2)^{tr}k(X_i,Y_2,\beta_0)|X_i\right\}}{E\left\{k(X_i,Y_2,\beta_0)|X_i\right\}}, \tag{5.115}$$

we can write (5.112) as the summation of

$$A = \frac{1}{n}\sum_{i=1}^{n} Q(X_i,\beta_0)^2 \Bigg\{ \frac{\sum_{j=1}^{n} C(X_i)C(X_j)^{tr}k(X_i,X_j,\beta_0)}{\sum_{j=1}^{n} k(X_i,X_j,\beta_0)} - \frac{E\left\{C(X_i)C(Y_2)^{tr}k(X_i,Y_2,\beta_0)|X_i\right\}}{E\left\{k(X_i,Y_2,\beta_0)|X_i\right\}} \Bigg\} \tag{5.112.a}$$

and

$$B = \frac{1}{n}\sum_{i=1}^{n} Q(X_i,\beta_0)^2 \left\{ \frac{E\left\{C(X_i)C(Y_2)^{tr}k(X_i,Y_2,\beta_0)|X_i\right\}}{E\left\{k(X_i,Y_2,\beta_0)|X_i\right\}} \right\} \tag{5.112.b}$$

By LLN, (5.112.b) converges to (5.112') in probability. Next, we show that (5.112.a) converges to zero in probability. Let

$$d(X_i,\boldsymbol{X},\beta_0) = \frac{\sum_{j=1}^{n} C(X_i)C(X_j)^{tr}k(X_i,X_j,\beta_0)}{\sum_{j=1}^{n} k(X_i,X_j,\beta_0)} - \frac{E\left\{C(X_i)C(Y_2)^{tr}k(X_i,Y_2,\beta_0)|X_i\right\}}{E\left\{k(X_i,Y_2,\beta_0)|X_i\right\}}. \tag{5.116}$$

First, we apply Lemma 5.3.4 to (5.116) by setting

$$\begin{aligned} U_{i,j} &= C(X_i)C(X_j)^{tr}k(X_i,X_j,\beta_0), \\ V_i &= E\left\{C(X_i)C(Y_2)^{tr}k(X_i,Y_2,\beta_0)|X_i\right\} \\ Z_{i,j} &= k(X_i,X_j,\beta_0), \\ W_i &= E\left\{k(X_i,Y_2,\beta_0)|X_i\right\}. \end{aligned} \tag{5.117}$$

The assumption (5.106) in Lemma 5.3.4 is guaranteed by Lemma 4.2.1 under assumptions (A.2), (A.9) and (A.10). By (A.2), there exists a constant b such that $\|C(x)\| \le b$ for all x, hence

$$\frac{\|\sum_{j=1}^{n} U_{i,j}\|}{\|\sum_{j=1}^{n} Z_{i,j}\|} = \frac{\|\sum_{j=1}^{n} C(X_i)C(X_j)^{tr}k(X_i,X_j,\beta_0)\|}{\|\sum_{j=1}^{n} k(X_i,X_j,\beta_0)\|} \le b^2$$

$$\frac{\|V_i\|}{|W_i|} = \frac{\|E\left\{C(X_i)C(Y_2)^{tr}k(X_i,Y_2,\beta_0)|X_i\right\}\|}{E\left\{k(X_i,Y_2,\beta_0|X_i\right\}} \le b^2 \frac{|E\left\{k(X_i,Y_2,\beta_0)|X_i\right\}|}{E\left\{k(X_i,Y_2,\beta_0)|X_i\right\}} = b^2$$

Therefore the conclusion (5.107) of Lemma 5.3.4 holds. Moreover, notice that

(i) $C(x)$ and $k(x, y, \beta)$ are both bounded by a fixed constant under assumptions (A.2) and (A.9);

(ii) $\mathcal{X} \equiv$support(X_1) and U_{β_0} are both bounded in $\mathbb{R}^p$;

The convergences in (4.15) and (4.14) actually hold in L^p norm, where $p = 1, 2, \ldots,$ therefore

$$E\left\{\|d(X_i, \boldsymbol{X}, \beta_0)\| \mid X_i\right\} \xrightarrow{P} 0, \quad \text{for any } i. \tag{5.118}$$

Note that as a random variable, the left hand side of (5.118) also satisfied (i) and (ii), hence it also converges in L^r norm, and

$$\begin{aligned} &E\left\{\|d(X_i, \boldsymbol{X}, \beta_0)\|\right\} \to 0, \quad \text{for any } i. \\ &E\left\{\|d(X_i, \boldsymbol{X}, \beta_0)\|^2\right\} \to d < \infty, \quad \text{for any } i. \end{aligned} \tag{5.119}$$

Recall that under assumption (A.2), $|Q(x, \beta_0)| < M < \infty$, for any x, hence

$$\|A\| = \frac{1}{n}\sum_{i=1}^{n} Q(X_i, \beta_0)^2 \|d(X_i, \boldsymbol{X}, \beta_0)\| \leq M^2 \cdot \frac{1}{n}\sum_{i=1}^{n} \|d(X_i, \boldsymbol{X}, \beta_0)\| \tag{5.120}$$

Therefore for any $\varepsilon > 0$,

$$\begin{aligned} P\left\{\|A\| > \varepsilon\right\} &\leq P\left\{M^2 \cdot \frac{1}{n}\sum_{i=1}^{n} \|d(X_i, \boldsymbol{X}, \beta_0)\| > \varepsilon\right\} \\ &\leq \frac{\varepsilon}{M^2}\left(E\left\{\|d(X_i, \boldsymbol{X}, \beta_0)\|^2\right\}\right)^{1/2} \to 0, \end{aligned} \tag{5.121}$$

where we got the last inequality by Chebyshev's Inequality, and the convergence to zero by (5.119). Hence the conclusions follows by the definition of convergence in probability. □

Proof of Lemma 5.2.1 First, let us show that (a) is true. Under assumption (A.2) and (A.9), apply Proposition 5.2.1, we know that $n^{-1}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0)^{\otimes 2}$ converges to Σ in probability. Therefore the conclusions for (5.37) and (5.39) are true. By Proposition 4.2.2, the conclusion for (5.38) is true. As for the conclusion in (5.36), it is guaranteed by the asymptotic normality of $n^{-1/2}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0)$ shown in Lemma 4.2.2.

To prove (b), we follow the steps in the proof of part (b) in Lemma 5.1.1. Recall that by Lemma 5.3.2, the conclusion that 0 belongs to the $int\,(\mathrm{Conv}_n(\beta_0))$, which is the convex hull of $m_n(X_i, \boldsymbol{X}, \beta_0)$, can be drawn by showing for any $\varepsilon > 0$ and unit vector $t_0 \in \mathbb{R}^p$, there exists N_ε such that for $n > N_\varepsilon$, given $\boldsymbol{X}_\infty \in \mathcal{C}_\varepsilon^*$,

$$\inf_{t_0} P\left\{ t_0^{tr} \sum_{i=1}^{n} V_i m_n(X_i, \boldsymbol{X}, \beta_0) \geq 0 \middle| \boldsymbol{X}_\infty \right\} > 1 - \varepsilon, \qquad \text{for } \boldsymbol{X}_\infty \in \mathcal{C}_\varepsilon^*,$$

where $V = v$ is as constructed in (5.76). Note that with $g(X_i, \beta_0)$ replaced by the term $m_n(X_i, \boldsymbol{X}, \beta_0)$, (5.77) and (5.78) are both true , and (5.79) is true by Proposition 5.2.1 and Lemma 4.2.2. Therefore the conditions for the Lyapunov central limit theorem mentioned in Lemma 5.3.1 are verified with respect to $\{W_i\}$ variables under the extended estimating equation setting, and given $\boldsymbol{X}_\infty \in \mathcal{C}_\varepsilon^*$, it follows that

$$n^{-1/2} \sum_{i=1}^{n} t_0^{tr}(W_i - 1) m_n(X_i, \boldsymbol{X}, \beta_0)$$

is asymptotically normal, i.e., for any unit vector $s \in \mathbb{R}^p$,

$$n^{-1/2} \sum_{i=1}^{n} t_0^{tr}(W_i - 1) s^{tr} m_n(X_i, \boldsymbol{X}, \beta_0) \xrightarrow{D} N(0, s^{tr}\Sigma s), \qquad \text{with probability 1.}$$

Therefore (5.81) and (5.82) are both true with $g(X_i, \beta_0)$ replaced by $m_n(X_i, \boldsymbol{X}, \beta_0)$, and we finish the proof. $\square$

Proof of Lemma 5.2.2 For ε smaller than λ_0, the smallest eigenvalue of Σ, on the event $\mathcal{C}_\varepsilon^+$, for any fixed $\beta \in U_{\beta_0}$,

$$\left\| \frac{1}{n} \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)^{\otimes 2} - \Sigma(\beta, \beta_0) \right\|_2 < \frac{\varepsilon}{2}$$

Since $\beta \in U_{\beta_0}$, we know that $\Sigma(\beta, \beta_0)$ is positive definite, hence

$$\frac{1}{n} \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)^{\otimes 2} \text{ is positive definite for } \beta \in U_{\beta_0}. \tag{5.122}$$

By contradiction, we can show that the conclusion is true for any fixed $\beta \in U_{\beta_0}$. Suppose that for a fixed $\beta \in U_{\beta_0}$, there exist distinct t_1 and t_2 such that $g(t_1, \boldsymbol{X}, \beta) = g(t_2, \boldsymbol{X}, \beta) = 0$. Therefore by Rolle's theorem from [15], there exists $s \in (0, 1)$ such that for $t_3 = st_1 + (1-s)t_2$, $\partial g(t_3, \boldsymbol{X}, \beta)/\partial s = 0$, i.e.,

$$(t_1 - t_2)^{tr} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{m_n(X_i, \boldsymbol{X}, \beta)^{\otimes 2}}{(1 + t_3^{tr} m_n(X_i, \boldsymbol{X}, \beta))^2} \right\} (t_1 - t_2) = 0, \tag{5.123}$$

which implies

$$\frac{1}{n} \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)^{\otimes 2} \quad \text{is singular,}$$

contradicting (5.122). $\square$

Proof of Theorem 5.2.1 Define $\mathcal{C}$ in the following way:

$$\begin{aligned} \mathcal{C} \equiv \{\beta^* \in U_{\beta_0} : \exists\, r_{\beta^*} > 0 \text{ such that for any } \beta \in B_{r_{\beta^*}}(\beta^*), \\ \text{there exists } t = t(\beta) \text{ such that } g(t, \boldsymbol{X}, \beta) = 0\}, \end{aligned} \tag{5.124}$$

where

$$g(t, \boldsymbol{X}, \beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{m_n(X_i, \boldsymbol{X}, \beta)}{1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta)}. \tag{5.125}$$

First, we show that for any fixed β^* such that the equation $g(t, \boldsymbol{X}, \beta^*) = 0$ has a solution $t = t^*$, then $\beta^* \in \mathcal{C}$. By (5.125), $g(t^*, \boldsymbol{X}, \beta^*) = 0$ implies $\mathbf{0}$ belongs to $int\,(\mathrm{Conv}_n(\beta^*))$, consequently, there exists a unique solution to the maximizing problem (5.41) with the choice of $\beta = \beta^*$. Since this unique solution is bounded by 0 and 1, i.e.,

$$\frac{1}{n} \cdot \frac{1}{1 + t^{*tr} m_n(X_i, \boldsymbol{X}, \beta^*)} \in (0, 1) \qquad \text{for } i = 1, \ldots, n,$$

we know that $t^* \in D_{\beta^*}(\boldsymbol{X})$. Since $g(t^*, \boldsymbol{X}, \beta^*) = 0$, $\partial g(t, \boldsymbol{X}, \beta)/\partial t|_{(t^*, \beta^*)}$ is negative definite. By the implicit function theorem, there exists r_{β^*} such that for $\beta \in B_{r_{\beta^*}}(\beta^*)$, equation $g(t, \boldsymbol{X}, \beta) = 0$ has a solution $t = t(\beta)$. Therefore $t^* \in \mathcal{C}$.

In particular, since we have shown in (5.45) that when $\beta = \beta_0$, there exists t_0 such that $g(t_0, \boldsymbol{X}, \beta_0) = 0$, we know that the conclusion in the previous paragraph is true for there is $\beta = \beta_0$, i.e., $\beta_0 \in \mathcal{C}$.

Next, we show that $\mathcal{C} = U_{\beta_0}$, i.e., for any $\beta^* \in U_{\beta_0}$, β^* also belongs to $\mathcal{C}$. Let

$$\rho_{\max} = \sup \left\{ s : \beta_0 + r(\beta^* - \beta_0) \in \mathcal{C}, \text{ for all } 0 < r < s \right\}, \tag{5.126}$$

By contradiction, we can show that $\rho_{\max} \geq 1$. Otherwise, suppose $\rho_{\max} < 1$. By equation (5.126), we know that

$$\beta = \beta_0 + r(\beta^* - \beta_0) \in \mathcal{C}, \qquad \text{for all } r < \rho_{\max} \tag{5.127}$$

Let $\{r_k\}_{k=1}^{\infty}$ be a series of increasing positive numbers that are bounded by $\rho_{\max}$, and

$$\beta_k = \beta_0 + r_k(\beta^* - \beta_0), \qquad \text{for } 0 < r_k < \rho_{\max}. \tag{5.128}$$

Therefore by (5.127), $\beta_k \in \mathcal{C}$, which together with (5.85) - (5.87) implies that for each β_k,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{m_n(X_i,\boldsymbol{X},\beta_k)}{1+t(\beta_k)^{tr}m_n(X_i,\boldsymbol{X},\beta_k)}=0 \quad \text{and} \quad 1+t^{tr}(\beta_k)m_n(X_i,\boldsymbol{X},\beta_k)\geq\frac{1}{n} \tag{5.129}$$

Let

$$\beta_{\max}=\lim_{k\to\infty}\beta_k=\beta_0+\rho_{\max}(\beta^*-\beta_0) \tag{5.130}$$

then by taking the limit of (5.129) as k goes to ∞,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{m_n(X_i,\boldsymbol{X},\beta_{\max})}{1+t(\beta_{\max})^{tr}m_n(X_i,\boldsymbol{X},\beta_{\max})}=0. \tag{5.131}$$

Therefore $\mathbf{0}$ belongs to $int\,(\mathrm{Conv}_n(\beta_{\max}))$. Consequently, there exists a unique solution to the maximizing problem (5.41) with $\beta=\beta_{\max}$. Since the unique solution is in $(0,1)$, we know that $t(\beta_{\max})\in D_{\beta_{\max}}(\boldsymbol{X})$. Similar to the previous proof, $\beta_{\max}\in\mathcal{C}$.

By the definition of $\mathcal{C}$, there exists $r_{\max}$ such that for all $\beta\in B_{r_{\max}}(\beta_{\max})$, we know that $t=t(\beta)$ solves $g(t,\boldsymbol{X},\beta)=0$. Therefore $\beta=\beta_0+r(\beta^*-\beta_0)\in\mathcal{C}$, where

$$r=\rho_{\max}+r_{\max}/4>\rho_{\max}$$

contradicts the definition of $\rho_{\max}$, hence $\rho_{\max}\geq 1$ and we conclude that $\beta^*\in\mathcal{C}$. $\square$

Proof of Theorem 5.2.2 Let U_i be i.i.d. random variables that has a uniform distribution over the span of $t_0^{tr}\mathcal{X}$, and let U_i be independent of $\boldsymbol{X}_\infty$, hence by definition, there exists a constant M such that $|U_i|\leq M$. Define

$$V_i^*=\frac{U_i^*}{\sum_{i=1}^{n}U_i^*} \qquad \text{where } U_i^*=U_iq(X_i,\beta),\ i=1,\ldots,n \tag{5.132}$$

Let

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i^* t_0^{tr} m_n(X_i,\boldsymbol{X},\beta)\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i^* t_0^{tr} V(X_i,\beta)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i^* t_0^{tr}\left((m_n(X_i,\boldsymbol{X},\beta)-V(X_i,\beta)\right)\end{aligned} \tag{5.133}$$

Note that on the event $\mathcal{C}_\varepsilon^+$, by the discussion between page 100 and 102, and the boundedness of U_i^*, the second term in (5.133) is $O_P(1)$, namely,

$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i^* t_0^{tr}\left((m_n(X_i,\boldsymbol{X},\beta)-V(X_i,\beta)\right)\right|\leq\frac{1}{\sqrt{n}}\sum_{i=1}^{n}M\cdot\frac{K_\varepsilon}{\sqrt{n}}=MK_\varepsilon. \tag{5.134}$$

Therefore on event $\mathcal{C}_\varepsilon^+$, for a fixed $\beta\in U_{\beta_0}$ and any unit vector $t_0\in\mathbb{R}^p$

$$\begin{aligned}P\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i^* t_0^{tr} m_n(X_i,\boldsymbol{X},\beta)>0\right\}&\geq P\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i^* t_0^{tr} V(X_i,\beta)>MK_\varepsilon\right\}\\ &=P\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i q(X_i,\beta) t_0^{tr} V(X_i,\beta)>MK_\varepsilon\right\}\end{aligned} \tag{5.135}$$

Using the same procedure as in the proof of Lemma 5.1.3, for $n>N_\varepsilon$ and $\boldsymbol{X}\in\mathcal{C}_\varepsilon^+$, by the Lyapunov central limit theorem mentioned in Lemma 5.3.1

$$P\left\{\inf_{s\in\mathbb{R}^p:\|s\|=1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i q(X_i,\beta) t_0^{tr} V(X_i,\beta)>MK_\varepsilon\right\}>0 \tag{5.136}$$

Combine (5.135) and (5.136), yielding,

$$P\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i^* t_0^{tr} m_n(X_i,\boldsymbol{X},\beta)>0\right\}>0. \tag{5.137}$$

Note that equation (5.137) implies that under the condition of this lemma, $\mathbf{0}\in$ *int* $(\mathrm{Conv}_n(\beta))$ with weight V_i^* defined in (5.132) with a positive probability. Hence we complete the proof. □

Chapter 6: Empirical Likelihood Applied to Extended Estimating Equations

Consider d-dimensional *independently identically distributed* (i.i.d.) observations

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} P_\theta, \quad i = 1, \ldots, n \tag{6.1}$$

with support $\mathcal{X}$, and parameters

$$\theta = (\beta, \nu) \in \mathbb{R}^p \times \mathcal{H}, \tag{6.2}$$

and $\mathcal{H}$ is a infinite dimensional space such as function space. The *Empirical Likelihood* (EL) method is an estimation method that maximizes the empirical distribution subject to constraints. For example, in a classic setting, if there exists a estimating function

$$m(x, \beta) : \mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}^r \tag{6.3}$$

such that $E_{\beta_0}\{m(X_1, \beta_0)\} = 0$, then the EL method seeks solution to the maximization problem for a fix β in a neighbourhood of β_0

$$\begin{cases} \arg\max_{\boldsymbol{p}} \prod_{i=1}^{n} p_i, \quad \text{where } \boldsymbol{p} = (p_1, \ldots, p_n), \\ \text{subject to } \sum_{i=1}^{n} p_i = 1, p_i \in (0,1), \sum_{i=1}^{n} p_i m(X_i, \beta) = 0. \end{cases} \tag{6.4}$$

The solution of (6.4), given by $\hat{p}_i(\beta, X_i)$ can be used to construct the *Profile Empirical Likelihood*. A point estimator for β_0 can be constructed via maximizing the pEL, or equivalently, its negative logarithm value, i.e.,

$$\hat{\beta}_n = \arg\min_{\beta} l(\beta), \qquad \text{where } l(\beta) = \sum_{i=1}^{n} \ln \hat{p}_i(\beta, X_i). \tag{6.5}$$

EL method has been extensively researched in statistical literature. Owen in [32] established the Wilks type confidence region when the dimension of β_0 equals the dimension of $m(x, \beta)$, i.e., $r = p$. The Wilks type *Confidence Region* (CR) does not require calculating the variance covariance matrix of $m(X_1, \beta_0)$, and the $E\left\{\nabla_\beta m(X_1, \beta_0)\right\}$, and can usually provide a narrower CR then the Wald type statistics.

Another convenient feature of EL method is that it can by pass some regularity conditions that is essential to estimating equations. For example, the embedded constraints $m(x, \beta)$ can have a higher dimension than parameter, i.e., $r > p$. This result can be found in Qin and Lawless's work in [33].

Continuity is a necessary condition for the existence of solution to estimating equation $S_n(\beta) = 0$, where

$$S_n(\beta) = \sum_{i=1}^{n} m(X_i, \beta)$$

and thus cannot be ignored. Under the CEE setting, Owen, and Qin and Lawless both assumed that the criterion function $m(x, \cdot)$ is continuous in [32] and [33]. However, we found in some prominent right censored semiparametric model like

φ-transformation model and Cox model, $m_n(x, \boldsymbol{x}, \cdot)$ is usually discontinuous due to the appearance of indicator functions. Therefore in Chapter 5, we show that under *Classic Estimating Equation* (CEE) setting with continuous or discontinuous $m(x, \cdot)$, and *Extended Estimating Equation* (EEE) setting with continuous $m_n(x, \boldsymbol{x}, \cdot)$, there exists U_{β_0}, a non shrinking neighbourhood of β_0 such that for any $\beta \in U_{\beta_0}$, (6.4) has a unique solution with probability approaching 1. The same result can be shown for EEE when $m_n(X_i, \boldsymbol{X}, \cdot)$ loses continuity if we restrict to the φ-transformation model discussed in Chapter 3, which is the broadest type of semiparametric model of which the structural parameter can be constructed via EEE to our knowledge.

6.1 Empirical likelihood of Extended Estimating Equation

In this section, we consider i.i.d. observations

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} P_\theta, \qquad \theta = (\beta, \nu) \in \mathbb{R}^p \times \mathcal{H}, \tag{6.6}$$

described in (6.1) - (6.2). In Definition 1.2.1 of page 6, we defined EEE as

$$S_n(\beta) = \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta) = 0, \qquad \beta \in U_{\beta_0}, \tag{6.7}$$

where $Q(x, \beta) : \mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}$, $k(x, y, \beta) : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}^+$, $C(x) : \mathbb{R}^d \mapsto \mathbb{R}^p$,

$$m_n(X_i, \boldsymbol{X}, \beta) = Q(X_i, \boldsymbol{X}, \beta) \left\{ C(X_i) - \frac{\sum_{j=1}^n C(X_j) k(X_i, X_j, \beta)}{\sum_{j=1}^n k(X_i, X_j, \beta)} \right\}, \tag{6.8}$$

with $E_{\beta_0, \nu} \{m_n(X_1, \boldsymbol{X}, \beta_0)\} = 0$ for all $\nu \in \mathcal{H}$, and $\tilde{\beta}_n$ is the unique solution to (6.7). In the following sections, we investigate the asymptotic normality of $\hat{\beta}_n$, the pEL estimator, and compare its asymptotic variance matrix of $\tilde{\beta}_n$.

6.2 Asymptotic Normality Associated with the EEE

In this section, we discuss the asymptotic normality of $\tilde{\beta}_n$ via solving EEE, and $\hat{\beta}_n$ via minimizing the negative logarithm of pEL, i.e.,

$$\begin{aligned} \tilde{\beta}_n: &\quad \text{solution to } S_n(\hat{\beta}_n) = 0, \text{ where } S_n(\beta) = \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta); \\ \hat{\beta}_n: &\quad \arg\min_{\beta} \sum_{i=1}^{n} \ln\left(1 + t^{tr}(\beta) m_n(X_i, \boldsymbol{X}, \beta)\right) \end{aligned} \tag{6.9}$$

First compute the asymptotic variance matrix of $\sqrt{n}S_n(\beta_0)$, which lead to the sandwich form variance of $\tilde{\beta}_0$. Then we compute the variance for $\hat{\beta}_n$ following the idea of Qin and Lawless's Lemma 1 and Theorem 1 in [33]. At the end of this section, we show that when the dimension of the EEE r equals to the dimension of β_0, then the two variances from $\tilde{\beta}_n$ and $\hat{\beta}_n$ are identical.

Theorem 6.2.1 *Under assumptions of (A.1) - (A.10),* $n^{-1/2}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0) \to N(0, \Sigma)$*, where* $\Sigma = \Sigma_1 + \Sigma_2$*, and for*

$$\begin{aligned} \Sigma_1 &= Var\left\{Q(X_1, \beta_0)\left\{C(X_1) - \frac{E\left\{C(X_2)k(X_1, X_2, \beta_0)|X_1\right\}}{E\left\{k(X_1, X_2, \beta_0)|X_1\right\}}\right\}\right\} \\ \Sigma_2 &= Var\left\{q_1(X_1, \beta_0)C(X_1)\right\} + Var\left\{q_2(X_1, \beta_0)\right\} - 2\Sigma_{AB}, \end{aligned} \tag{6.10}$$

and

$$\begin{aligned} q_1(X_1, \beta_0) &= E\left\{\frac{Q(X_2, \beta_0)k(X_2, X_1, \beta_0)}{E\left\{k(X_2, X_3, \beta_0)|X_2\right\}} \;\middle|\; X_1\right\}, \\ q_2(X_1, \beta_0) &= E\left\{\frac{Q(X_2, \beta_0)k(X_2, X_1, \beta_0)E\left\{C(X_3)k(X_2, X_3, \beta_0)|X_2\right\}}{\left\{E\left\{k(X_2, X_3, \beta_0)|X_2\right\}\right\}^2} \;\middle|\; X_1\right\} \\ \Sigma_{AB} &= E\Bigg\{\frac{Q(X_1, \beta_0)}{\bar{k}(X_1, \beta_0)}\left\{C(X_2)k(X_1, X_2, \beta_0) - E\left\{C(X_4)k(X_1, X_4, \beta_0)|X_1\right\}\right\} \\ &\qquad \times \frac{Q(X_3, \beta_0)\overline{k_c}(X_3, \beta_0)}{\bar{k}(X_3, \beta_0)^2}\left\{k(X_3, X_2, \beta_0) - E\left\{k(X_3, X_4, \beta_0)|X_3\right\}\right\}\Bigg\}. \end{aligned}$$

Asymptotic Varaince Covariance of $\tilde{\beta}_n$ Using the conclusion in Theorem 6.2.1, we can calculate the asymptotic variance matrix of $\tilde{\beta}_n$. Let $\tilde{\beta}_n$ be the solution to $S_n(\beta) = 0$. We can derive the sandwich formed asymptotic variance covariance matrix for $\sqrt{n}(\tilde{\beta}_n - \beta_0)$ using the same Taylor expansion method in Section 5.3 of [41] by Van der Vaart to conclude

$$\sqrt{n}(\tilde{\beta}_n - \beta_0) \xrightarrow{D} N(0, \tilde{\Sigma}), \tag{6.11}$$

where for Σ defined in Theorem 6.2.1,

$$\tilde{\Sigma} = E\left\{\nabla_\beta m_n(X_1, \boldsymbol{X}, \beta_0)\right\}^{-1} \Sigma E\left\{\nabla_\beta m_n(X_1, \boldsymbol{X}, \beta_0)\right\}^{-1} \tag{6.12}$$

Next, we discuss the asymptotic normality of $\hat{\beta}_n$. The following lemma is parallel Lemma 1 by Qin and Lawless in [33].

Lemma 6.2.1 *Under assumptions (A.1)-(A.10), as $n \to \infty$, with probability, $l(\beta)$ is minimized at $\hat{\beta}_n$ in the interior of $\left\{\beta : \ \|\beta - \beta_0\| \leq n^{-1/3}\right\}$, with $\hat{\beta}_n$ and $\hat{t}_n = t(\hat{\beta}_n)$ given by*

$$Q_{1n}(\hat{\beta}_n, \hat{t}_n) = 0, \qquad Q_{2n}(\hat{\beta}_n, \hat{t}_n) = 0, \tag{6.13}$$

where

$$Q_{1n}(\beta, t) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta)} m_n(X_i, \boldsymbol{X}, \beta),$$
$$Q_{2n}(\beta, t) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta)} \left(\frac{\partial m_n(X_i, \boldsymbol{X}, \beta)}{\partial \beta}\right)^{tr} t.$$

The proof of this Lemma is almost identical to Lemma 1 by Qin and Lawless in [33] after we developed the proceding parallel lemmas in Chapter 4. With Lemma 6.2.1, we can show the asymptotic normality of $\sqrt{n}(\hat{\beta}_n - \beta_0)$.

Theorem 6.2.2 *Under assumptions (A.1) - (A.10),* $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, V)$, *where* V *is defined in (6.12), where*

$$V = \left(\hat{S}_{22.1}^{-1}\hat{S}_{21}\hat{\Sigma}_1^{-1}\right)\Sigma\left(\hat{S}_{22.1}^{-1}\hat{S}_{21}\hat{\Sigma}_1^{-1}\right)^{tr} \tag{6.14}$$

with Σ *and* Σ_1 *defined in Theorem 6.2.1,* $\hat{S}_{21} = E\left\{\nabla_\beta m_n(X_i, \boldsymbol{X}, \beta_0)\right\}$, *and*

$$\hat{S}_{22.1} = \hat{S}_{21}^{tr}\Sigma_1^{-1}\hat{S}_{21}. \tag{6.15}$$

This is an extension of Theorem 1 by Qin and Lawless in [33] under the CEE setting. We followed their idea of proof after establishing preceding parallel lemmas in Chapter 4.

Now, assume $r = p$ as in all the other chapters of this thesis, where r is the dimension of $m_n(X_i, \boldsymbol{X}, \beta)$ and p is the dimension of β_0. Also assume $\hat{S}_{21} = E\left\{\nabla_\beta m_n(X_i, \boldsymbol{X}, \beta_0)\right\}$ is non singular for $\beta \in U_{\beta_0}$, where U_{β_0} is defined in Theorem 5.2.1. Then (6.14) becomes

$$V = \hat{S}_{21}^{-1}\Sigma_1\hat{S}_{21}^{-1}\hat{S}_{21}\Sigma_1^{-1}\Sigma\Sigma_1^{-1}\hat{S}_{21}\hat{S}_{21}^{-1}\Sigma_1\hat{S}_{21}^{-1} = \hat{S}_{21}^{-1}\Sigma S_{21}^{-1}. \tag{6.16}$$

Comparing (6.12) and (6.16), we can see that $V = \tilde{\Sigma}$, i.e., the asymptotic variance covariance matrix for $\hat{\beta}_n$ is the same as that for $\tilde{\beta}_n$.

6.3 Some Proofs

Proof of Lemma 6.2.1 Note that $\sqrt{n}\bar{m}_n(\boldsymbol{X}, \beta_0)$ can be split into the difference of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q(X_i, \beta_0)\left\{C(X_i) - \frac{\overline{k_c}(X_i, \beta_0)}{\bar{k}(X_i, \beta_0)}\right\}. \tag{6.17}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Q(X_i,\beta_0)\left\{\frac{n^{-1/2}\sum_{j=1}^{n} C(X_j)k(X_i,X_j,\beta_0)}{n^{-1/2}\sum_{j=1}^{n} k(X_i,X_j,\beta_0)} - \frac{\overline{k_c}(X_i,\beta_0)}{\bar{k}(X_i,\beta_0)}\right\} \tag{6.18}$$

Since (6.17) is a summation of i.i.d. terms, under assumption (A.1)-(A.3) and (A.5)-(A.6), by the CLT, it converges to $N(0,\Sigma_1)$, where

$$\Sigma_1 = Var\left\{Q(X,\beta_0)\left\{C(X) - \frac{\overline{k_c}(X,\beta_0)}{\bar{k}(X,\beta_0)}\right\}\right\}. \tag{6.19}$$

Next, we show that (6.18) is also asymptotically normal. Using the same strategy in the proof of Lemma 4.2.2, first split (6.18) into the difference of the following two terms,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Q(X_i,\beta_0)}{n^{-1/2}\sum_{j=1}^{n} k(X_i,X_j,\beta_0)}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{C(X_j)k(X_i,X_j,\beta_0) - \overline{k_c}(X_i,\beta_0)\right\} \tag{6.18.a}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Q(X_i,\beta_0)\overline{k_c}(X_i,\beta_0)}{\bar{k}(X_i,\beta_0)n^{-1/2}\sum_{j=1}^{n} k(X_i,X_j,\beta_0)}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{k(X_i,X_j,\beta_0) - \bar{k}(X_i,\beta_0)\right\}. \tag{6.18.b}$$

By Lemma 4.2.1, we know that

$$\sup_{x\in\mathcal{X}}\left\|\frac{1}{n}\sum_{j=1}^{n} k(x,X_j,\beta_0) - \bar{k}(x,\beta_0)\right\| \xrightarrow{a.s.} 0$$

Therefore for large n, (6.18.a) and (6.18.b) equals to

$$A = \frac{1}{n}\sum_{i=1}^{n}\frac{Q(X_i,\beta_0)}{\bar{k}(X_i,\beta_0)}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{C(X_j)k(X_i,X_j,\beta_0) - \overline{k_c}(X_i,\beta_0)\right\} \tag{6.18.a.1}$$

and

$$B = \frac{1}{n}\sum_{i=1}^{n}\frac{Q(X_i,\beta_0)\overline{k_c}(X_i,\beta_0)}{\bar{k}(X_i,\beta_0)^2}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{k(X_i,X_j,\beta_0) - \bar{k}(X_i,\beta_0)\right\}, \tag{6.18.b.1}$$

respectively. Let

$$d\mu_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i},$$

and $d\mu = dF_X$, then (6.18.a.1) can be re-written as an integral with respect to the difference between μ_n and μ, i.e.

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Q(X_i,\beta_0)}{\bar{k}(X_i,\beta_0)}\int C(y)k(X_i,y,\beta_0)\sqrt{n}\left\{d\mu_n(y)-d\mu(y)\right\}. \tag{6.20}$$

Let

$$d\gamma_n(y) = \sqrt{n}\left\{d\mu_n(y)-d\mu(y)\right\}.$$

Recall that for $i = 1,\ldots,n$,

$$\bar{k}(X_i,\beta_0) = E\left\{k(X_i,X,\beta_0)|X_i\right\} = \int k(X_i,x,\beta_0)d\mu(x).$$

Continuing to use the integrated empirical process notation to replace the sum, we know that (6.20) is equal to

$$\int \frac{Q(z,\beta_0)}{\int k(z,x,\beta_0)d\mu(x)}d\mu_n(z)\int C(y)k(z,y,\beta_0)d\gamma_n(y) \tag{6.21}$$

Note that, as $n\to\infty$, $\mu_n\to\mu$, which indicates that the integral in (6.21) approaches

$$\iint \frac{Q(z,\beta_0)C(y)k(z,y,\beta_0)}{\int k(z,x,\beta_0)d\mu(x)}d\mu(z)d\gamma_n(y) \tag{6.22}$$

Let

$$q_1(y,\beta_0) = \int \frac{Q(z,\beta_0)k(z,y,\beta_0)}{\int k(z,x,\beta_0)d\mu(x)}d\mu(z), \tag{6.23}$$

then (6.22) equals to

$$\int q_1(y,\beta_0)C(y)d\gamma_n(y), \tag{6.24}$$

which converges to $N(0, Var\left\{q_1(Y,\beta_0)C(Y)\right\})$. Note that for large n, (6.18.a) is equal to (6.18.a.1) almost surely. Therefore combining (6.20)-(6.24), we conclude

that (6.18.a) converges to $N(0, Var\{q_1(Y,\beta_0)C(Y)\})$. Similarly, (6.18.b) can be re-written as

$$\iint Q(z,\beta_0)k(z,y,\beta_0)\frac{\int k(z,x,\beta_0)C(x)d\mu(x)}{\left\{\int k(z,x,\beta_0)d\mu(x)\right\}^2}d\mu(z)d\gamma_n(y), \tag{6.25}$$

which converges to $N(0, Var\{q_2(Y,\beta_0)\})$, where

$$q_2(y,\beta_0) = \int Q(z,\beta_0)k(z,y,\beta_0)\frac{\int k(z,x,\beta_0)C(x)d\mu(x)}{\left\{\int k(z,x,\beta_0)d\mu(x)\right\}^2}d\mu(z). \tag{6.26}$$

It remains to show the joint normality and the asymptotic covariance of (6.18.a) and (6.18.b), or equivalently, the covariance of A and B defined in (6.18.a.1) and (6.18.b.1). Note that

$$\begin{aligned} E(A\cdot B) = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j=1}^{n}\sum_{j'=1}^{n} E\Bigg\{ & \frac{Q(X_i,\beta_0)}{\bar{k}(X_i,\beta_0)}\left\{C(X_j)k(X_i,X_j,\beta_0) - \overline{k_c}(X_i,\beta_0)\right\} \\ & \times \frac{Q(X_{i'},\beta_0)\overline{k_c}(X_{i'},\beta_0)}{\bar{k}(X_{i'},\beta_0)^2}\left\{k(X_{i'},X_{j'},\beta_0) - \bar{k}(X_{i'},\beta_0)\right\}\Bigg\} \end{aligned} \tag{6.27}$$

Let us consider the following cases which are broken down according to the number of distinct number of elements in $\mathcal{I} = \{i, j, i', j'\}$,

Case 1 If the number of distinct elements in $\mathcal{I}$ is 1, then there are n identical terms to be added

Case 2 If the number of distinct elements in $\mathcal{I}$ is 2, then there are $\{C_4^1 + C_4^2\}\, n(n-1)$ identical terms to be added.

Case 3 If the number of distinct elements in $\mathcal{I}$ is 3, then there are $C_4^1 \cdot n(n-1)(n-2)$ identical terms to be added.

Case 4 If the number of distinct elements in $\mathcal{I}$ is 4, then due to the independence of X_i's, the expected value in the summand of (6.27) is 0.

Since (6.27) is the summation of groups of identical terms then divided by n^3, so we only need to consider Case 3. Assume that $i \neq j$ and $i' \neq j'$ as we did before, and consider the following subcases for Case 3,

Case 3a $i = i'$, then the expected values of (6.27) are all equal to

$$E\Bigg\{\frac{Q(X_1,\beta_0)}{\bar{k}(X_1,\beta_0)}\left\{C(X_2)k(X_1,X_2,\beta_0)-\overline{k_c}(X_1,\beta_0)\right\} \times \frac{Q(X_1,\beta_0)\overline{k_c}(X_1,\beta_0)}{\bar{k}(X_1,\beta_0)^2}\left\{k(X_1,X_3,\beta_0)-\bar{k}(X_1,\beta_0)\right\}\Bigg\}. \tag{6.28}$$

We can show that (6.28) is identical to 0 by conditioning on X_1.

Case 3b $i = j'$, then the expected values of (6.27) are all equal to

$$E\Bigg\{\frac{Q(X_1,\beta_0)}{\bar{k}(X_1,\beta_0)}\left\{C(X_2)k(X_1,X_2,\beta_0)-\overline{k_c}(X_1,\beta_0)\right\} \times \frac{Q(X_3,\beta_0)\overline{k_c}(X_3,\beta_0)}{\bar{k}(X_3,\beta_0)^2}\left\{k(X_3,X_1,\beta_0)-\bar{k}(X_3,\beta_0)\right\}\Bigg\} \tag{6.29}$$

We can show that (6.29) is identical to 0 by conditioning on X_1 and X_3.

Case 3c $j = i'$, then the expected values of (6.27) are all equal to

$$E\Bigg\{\frac{Q(X_1,\beta_0)}{\bar{k}(X_1,\beta_0)}\left\{C(X_2)k(X_1,X_2,\beta_0)-\overline{k_c}(X_1,\beta_0)\right\} \times \frac{Q(X_2,\beta_0)\overline{k_c}(X_2,\beta_0)}{\bar{k}(X_2,\beta_0)^2}\left\{k(X_2,X_3,\beta_0)-\bar{k}(X_2,\beta_0)\right\}\Bigg\} \tag{6.30}$$

We can show that (6.30) is identical to 0 by conditioning on X_1 and X_2

Case 3d $i = 1, j = 2, i' = 3, j' = 2$, then the expected values of (6.27) are all equal to

$$\Sigma_{AB} = E\Bigg\{ \frac{Q(X_1, \beta_0)}{\bar{k}(X_1, \beta_0)} \left\{ C(X_2) k(X_1, X_2, \beta_0) - \overline{k_c}(X_1, \beta_0) \right\} \tag{6.31}$$
$$\times \frac{Q(X_3, \beta_0)\overline{k_c}(X_3, \beta_0)}{\bar{k}(X_3, \beta_0)^2} \left\{ k(X_3, X_2, \beta_0) - \bar{k}(X_3, \beta_0) \right\} \Bigg\}$$

Combining (6.18.a), (6.24) and (6.26), we know that (6.18) converges to $N(0, \Sigma_2)$, where

$$\Sigma_2 = Var\left\{ q_1(Y, \beta_0) C(Y) \right\} + Var\left\{ q_2(Y, \beta_0) \right\} - 2\Sigma_{AB} \tag{6.32}$$

As for the asymptotic covariance of (6.17) and (6.18), consider

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{i'=1}^{n} E\Bigg\{ Q(X_i, \beta_0) \left\{ C(X_i) - \frac{\overline{k_c}(X_i, \beta_0)}{\bar{k}(X_i, \beta_0)} \right\} \tag{6.33}$$
$$\times Q(X_{i'}, \beta_0) \left\{ \frac{n^{-1} \sum_{j \neq i}^{n} C(X_j) k(X_{i'}, X_j, \beta_0)}{n^{-1} \sum_{j \neq i}^{n} k(X_{i'}, X_j, \beta_0)} - \frac{\overline{k_c}(X_{i'}, \beta_0)}{\bar{k}(X_{i'}, \beta_0)} \right\} \Bigg\} \tag{6.34}$$

When $i \neq i'$, then the terms in (6.33) and (6.34) are independent, hence the expected value of the product of (6.33) and (6.34) equals to the product of their expected value. Furthermore, notice that (6.33) are centered at zero, therefore when $i \neq i'$, the expectation of (6.33)-(6.34) is zero. Next, we consider the case when $i = i'$. Since there are n identical terms, we can rewrite the expectation of the (6.33)-(6.34) as

$$E\Bigg\{ Q(X_1, \beta_0) \left\{ C(X_1) - \frac{\overline{k_c}(X_1, \beta_0)}{\bar{k}(X_1, \beta_0)} \right\}$$
$$\times Q(X_1, \beta_0) \left\{ \frac{n^{-1} \sum_{j \neq i}^{n} C(X_j) k(X_1, X_j, \beta_0)}{n^{-1} \sum_{j \neq i}^{n} k(X_1, X_j, \beta_0)} - \frac{\overline{k_c}(X_1, \beta_0)}{\bar{k}(X_1, \beta_0)} \right\} \Bigg\}$$

Since by strong law of large numbers, as n goes to infinity, $n^{-1}\sum_{j=1}^{n} C(X_j)k(X_1, X_j, \beta_0)$ and $n^{-1}\sum_{j=1}^{n} k(X_1, X_j, \beta_0)$ goes to $\overline{k_c}(X_1, \beta_0)$ and $\bar{k}(X_1, \beta_0)$, respectively, then by dominated convergence theorem, the expected value in the expression above also approach zero as n goes to infinity. Therefore (6.17) and (6.18) are asymptotically independent, and $\sqrt{n}\bar{m}_n(\boldsymbol{X}, \beta_0) \to N(0, \Sigma)$ with $\Sigma = \Sigma_1 + \Sigma_2$. □

Proof of Lemma 6.2.1 Let β be on the surface of the ball centered at β_0 and with radius $n^{-1/3}$. Hence for a unit vector u,

$$\beta = \beta_0 + un^{-1/3}.$$

Next, we give a lower bound for $l(\beta)$ on the surface of the ball. Let $\upsilon \in \mathbb{R}^p$ be a unit vector and $t = \|t\|\upsilon$. Owen in [32] showed that the (6.35) holds uniformly for values of $\beta \in \left\{\beta : \|\beta - \beta_0\| \leq n^{-1/3}\right\}$ under the CEE setting, i.e., when the criterion function is given by $m(X_i, \beta)$

$$t(\beta) = \left[\frac{1}{n}\sum_{i=1}^{n} m(X_i, \beta)^{\otimes 2}\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} m(X_i, \beta)\right] + o(n^{-1/3}), \tag{6.35}$$

uniformly about $\beta \in \left\{\beta : \|\beta - \beta_0\| \leq n^{-1/3}\right\}$.

It is essential that (6.35) is still true in the EEE setting, therefore we prove it following the steps in Qin and Lawless [33] and Owen [32]. Since we have shown the parallel preceding lemmas in Chapter 4, the steps are not much different than that in [33] or [32]. Let $Y_i(X_i, \boldsymbol{X}, \beta) = t^{tr} m_n(X_i, \boldsymbol{X}, \beta)$ and substitute

$$\frac{1}{1 + Y_i(X_i, \boldsymbol{X}, \beta)} = 1 - \frac{Y_i(X_i, \boldsymbol{X}, \beta)}{1 + Y_i(X_i, \boldsymbol{X}, \beta)} \tag{6.36}$$

Note that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{m_n(X_i,\boldsymbol{X},\beta)}{1+t^{tr}m_n(X_i,\boldsymbol{X},\beta)}=0 \tag{6.37}$$

can be re-written as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{m_n(X_i,\boldsymbol{X},\beta)}{1+Y_i(X_i,\boldsymbol{X},\beta)}=0 \tag{6.38}$$

Multiply (6.38) by υ, then plug in (6.36), then we get

$$\begin{aligned}0&=\frac{1}{n}\sum_{i=1}^{n}\upsilon^{tr}m_n(X_i,\boldsymbol{X},\beta)\left(1-\frac{Y_i(X_i,\boldsymbol{X},\beta)}{1+Y_i(X_i,\boldsymbol{X},\beta)}\right)\\&=\upsilon^{tr}\sum_{i=1}^{n}m_n(X_i,\boldsymbol{X},\beta)-\frac{1}{n}\sum_{i=1}^{n}\frac{\upsilon^{tr}m_n(X_i,\boldsymbol{X},\beta)m_n(X_i,\boldsymbol{X},\beta)^{tr}t}{1+Y_i(X_i,\boldsymbol{X},\beta)}.\end{aligned} \tag{6.39}$$

Define

$$\begin{aligned}\tilde{S}(\boldsymbol{X},\beta)&=\frac{1}{n}\sum_{i=1}^{n}\frac{m_n(X_i,\boldsymbol{X},\beta)^{\otimes 2}}{1+Y_i(X_i,\boldsymbol{X},\beta)}\\S(\boldsymbol{X},\beta)&=\frac{1}{n}\sum_{i=1}^{n}m_n(X_i,\boldsymbol{X},\beta)^{\otimes 2},\end{aligned} \tag{6.40}$$

Plug (6.40) into (6.39), yielding

$$\upsilon^{tr}\bar{m}_n(\boldsymbol{X},\beta)=\|t\|\upsilon^{tr}\tilde{S}(\boldsymbol{X},\beta)\upsilon \tag{6.41}$$

By the definition of $\tilde{S}$ in (6.40),

$$\tilde{S}(\boldsymbol{X},\beta)\geq\frac{1}{n}\sum_{i=1}^{n}\frac{m_n(X_i,\boldsymbol{X},\beta)^{\otimes 2}}{1+\max_i Y_i(X_i,\boldsymbol{X},\beta)}=S(\boldsymbol{X},\beta)\cdot\frac{1}{1+\max_i Y_i(X_i,\boldsymbol{X},\beta)},$$

implying

$$S(\boldsymbol{X},\beta)\leq\tilde{S}(\boldsymbol{X},\beta)(1+\max_i Y_i(X_i,\boldsymbol{X},\beta)),$$

where the notation that $A\geq B$ for matrix A and B means that $A-B$ is positive definite. We know that .

$$\|t\|\upsilon^{tr}S(\boldsymbol{X},\beta)\upsilon\leq\|t\|\upsilon^{tr}\tilde{S}(\boldsymbol{X},\beta)\upsilon(1+\max_i Y_i(X_i,\boldsymbol{X},\beta)). \tag{6.42}$$

Let $Z_n^*(\boldsymbol{X}, \beta) = \max_i \|m_n(X_i, \boldsymbol{X}, \beta)\|$ in (4.7), so

$$\|t\| \upsilon^{tr} \tilde{S}(\boldsymbol{X}, \beta) \upsilon (1 + \max_i Y_i(X_i, \boldsymbol{X}, \beta)) \leq \|t\| \upsilon^{tr} \tilde{S}(\boldsymbol{X}, \beta) \upsilon (1 + \|t\| Z_n^*(\boldsymbol{X}, \beta)) \quad (6.43)$$

Now, by (6.41), $\|t\| \upsilon^{tr} \tilde{S}(\boldsymbol{X}, \beta) \upsilon$ on the right hand side of (6.43) can be substituted by $\upsilon^{tr} \bar{m}_n(\boldsymbol{X}, \beta)$, hence

$$\|t\| \upsilon^{tr} \tilde{S}(\boldsymbol{X}, \beta) \upsilon (1 + \max_i Y_i(X_i, \boldsymbol{X}, \beta)) \leq \upsilon^{tr} \bar{m}_n(\boldsymbol{X}, \beta)(1 + \|t\| Z_n^*(\boldsymbol{X}, \beta)), \quad (6.44)$$

where

$$\bar{m}_n(\boldsymbol{X}, \beta) = \frac{1}{n} \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta). \quad (6.45)$$

Combining (6.42) and (6.44), we attain

$$\|t\| \upsilon^{tr} S(\boldsymbol{X}, \beta) \upsilon \leq \upsilon^{tr} \bar{m}_n(\boldsymbol{X}, \beta)(1 + \|t\| Z_n^*(\boldsymbol{X}, \beta)), \quad (6.46)$$

which result in

$$\|t\| (\upsilon^{tr} S(\boldsymbol{X}, \beta) \upsilon - Z_n^*(\boldsymbol{X}, \beta) \upsilon^{tr} \bar{m}_n(\boldsymbol{X}, \beta)) \leq \upsilon^{tr} \bar{m}_n(\boldsymbol{X}, \beta). \quad (6.47)$$

Note that we assume $t = \|t\| \upsilon$ and υ is a unit vector. By Proposition 4.2.1 and Lemma 4.2.2, we know that

$$Z_n^*(\boldsymbol{X}, \beta) \upsilon^{tr} \bar{m}_n(\boldsymbol{X}, \beta) = o(n^{1/2}) O(n^{\delta - 1/2}) = o(n^{\delta}). \quad (6.48)$$

Plug (6.48) into (6.47), yielding

$$\|t\| (\upsilon^{tr} S(\boldsymbol{X}, \beta) \upsilon + o(n^{\delta})) \leq O(n^{\delta - 1/2}) \quad (6.49)$$

Since $\upsilon^{tr} S \upsilon$ is bounded by the minimum and maximum eigen value of $Var(m_n(X_i, \boldsymbol{X}, \beta))$, (6.49) indicates that

$$\|t\| = O(n^{-1/2}). \quad (6.50)$$

From (6.38) we can see that

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)\left(1 - Y_i(X_i, \boldsymbol{X}, \beta) + \frac{Y_i(X_i, \boldsymbol{X}, \beta)^2}{1 + Y_i(X_i, \boldsymbol{X}, \beta)}\right) \\
&= \bar{m}_n(\boldsymbol{X}, \beta) - S(\boldsymbol{X}, \beta)t + \frac{1}{n}\sum_{i=1}^{n} \frac{m_n(X_i, \boldsymbol{X}, \beta)}{1 + Y_i(X_i, \boldsymbol{X}, \beta)} Y_i(X_i, \boldsymbol{X}, \beta)^2. \qquad (6.51)
\end{aligned}
$$

Now, let us discuss the norm of the last term in (6.51). Recall that $Y_i(X_i, \boldsymbol{X}, \beta) = t^{tr} m_n(X_i, ,\beta)$, therefore

$$
\begin{aligned}
\left\| \frac{1}{n}\sum_{i=1}^{n} \frac{m_n(X_i, \boldsymbol{X}, \beta)}{1 + Y_i(X_i, \boldsymbol{X}, \beta)} Y_i(X_i, \boldsymbol{X}, \beta)^2 \right\| &= \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{m_n(X_i, \boldsymbol{X}, \beta)}{1 + Y_i(X_i, \boldsymbol{X}, \beta)} \left(t^{tr} m_n(X_i, ,\beta)\right)^2 \right\| \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \|m_n(X_i, \boldsymbol{X}, \beta_0)\|^3 \|t\|^2 \frac{1}{1 + Y_i(X_i, \boldsymbol{X}, \beta)}.
\end{aligned} \qquad (6.52)
$$

Plug $Z_n^*(\boldsymbol{X}, \beta) = \max_i m_n(X_i, \boldsymbol{X}, \beta)$ into (6.52), yielding

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \|m_n(X_i, \boldsymbol{X}, \beta_0)\|^3 \|t\|^2 \frac{1}{1 + Y_i(X_i, \boldsymbol{X}, \beta)} \qquad (6.53) \\
&\qquad \leq Z_n^*(\boldsymbol{X}, \beta) \cdot \frac{1}{n}\sum_{i=1}^{n} \|m_n(X_i, \boldsymbol{X}, \beta)\|^2 \cdot \|t\|^2 \cdot \frac{1}{1 + Y_i(X_i, \boldsymbol{X}, \beta)}.
\end{aligned}
$$

Let δ be a number between 0 and 1/2, by Proposition 4.2.1, Lemma 4.2.2 and (6.50), we know that

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} \|m_n(X_i, \boldsymbol{X}, \beta_0)\|^3 \|t\|^2 \frac{1}{1 + Y_i(X_i, \boldsymbol{X}, \beta)} &\leq o(n^{1/2}) O(n^{\delta}) O(n^{-1}) \\
&= o(n^{-(1/2-\delta)})
\end{aligned} \qquad (6.54)
$$

Equations (6.52)-(6.54) implies that the norm of the last term in (6.51) is of order $n^{-(1/2-\delta)}$, therefore from (6.51), we know that

$$
t = S^{-1}(\boldsymbol{X}, \beta)\bar{m}_n(\boldsymbol{X}, \beta) + \gamma, \qquad \text{where } \gamma = o(n^{-1/3}). \qquad (6.55)
$$

Let $l(\beta)$ be the negative logarithm of the profile empirical likelihood function

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{n} \ln\left\{1 + t^{tr}(\beta) m_n(X_i, \boldsymbol{X}, \beta)\right\} \\ &= \sum_{i=1}^{n} t^{tr}(\beta) m_n(X_i, \boldsymbol{X}, \beta) - \frac{1}{2}\sum_{i=1}^{n}\left[t^{tr}(\beta) m_n(X_i, \boldsymbol{X}, \beta)\right]^2 + o(n^{1/3}) \qquad \text{a.s.} \end{aligned} \tag{6.56}$$

By (6.55), we can rewrite the right hand side of (6.56) as

$$\frac{n}{2}\left[\frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)\right]^{tr}\left[\frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)^{\otimes 2}\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)\right] + o(n^{1/3}) \tag{6.57}$$

Take Taylor expansion of (6.57) around the true parameter value β_0, attaining

$$\begin{aligned} &\frac{n}{2}\left[\frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0) + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial m_n(X_i, \boldsymbol{X}, \beta_0)}{\partial \beta} u n^{-1/3}\right]^{tr}\left[\frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)^{\otimes 2}\right]^{-1} \\ &\times\left[\frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0) + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial m_n(X_i, \boldsymbol{X}, \beta_0)}{\partial \beta} u n^{-1/3}\right] + o(n^{1/3}) \end{aligned} \tag{6.58}$$

Combining (6.56)-(6.58), by Lemma 4.2.2, we know that

$$\begin{aligned} l(\beta) = \frac{n}{2}&\left[O(n^{-1/2+\delta}) + E\left(\frac{\partial m(X, Y, \beta_0)}{\partial \beta}\right) u n^{-1/3}\right]^{tr} \\ &\times \Sigma_1^{-1} \\ &\times\left[O(n^{-1/2+\delta}) + E\left(\frac{\partial m(X, Y, \beta_0)}{\partial \beta}\right) u n^{-1/3}\right] + o(n^{1/3}) \\ &\geq (c - \varepsilon) n^{1/3} \qquad \text{a.s.,} \end{aligned}$$

where $0 < \delta < 1/6$, and

$$\Sigma_1 = E\left\{Q(Y_1, \beta_0)^2\left[C(Y_1) - \frac{E\left\{C(Y_2) k(Y_1, Y_2, \beta_0) | Y_1\right\}}{E\left\{k(Y_1, Y_2, \beta_0) | Y_1\right\}}\right]^{\otimes 2}\right\}. \tag{6.59}$$

with Y_1, Y_2 being i.i.d. replica of X_1 by Proposition 5.2.1. Similarly,

$$l(\beta_0) = \frac{n}{2}\left[\frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0)\right]^{tr}\left[\frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0)^{\otimes 2}\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0)\right] \qquad \text{a.s.}$$

$$= O(n^{\delta}) \qquad \text{a.s.}$$

Since $l(\beta)$ is continuous for $\beta \in \left\{\beta \| \|\beta - \beta_0\| \le n^{-1/3}\right\}$, there exists a minimum for $l(\beta)$ in the interior of the ball, and $\hat{\beta}_n$ satisfies

$$\begin{aligned}\left.\frac{\partial l(\beta)}{\partial \beta}\right|_{\beta=\hat{\beta}_n} &= \sum_{i=1}^{n} \left.\frac{(\partial t^{tr}(\beta)/\partial\beta)\, m_n(X_i, \boldsymbol{X}, \beta) + t^{tr}(\beta)\,(\partial m_n(X_i, \boldsymbol{X}, \beta)/\partial\beta)}{1 + t^{tr}(\beta) m_n(X_i, \boldsymbol{X}, \beta)}\right|_{\beta=\hat{\beta}_n} \\ &= \sum_{i=1}^{n} \frac{1}{1 + t^{tr}(\beta) m_n(X_i, \boldsymbol{X}, \beta)} \left.\left(\frac{\partial m_n(X_i, \boldsymbol{X}, \beta)}{\partial \beta}\right)^{tr} t(\beta)\right|_{\beta=\hat{\beta}_n} = 0\end{aligned}$$

□

Proof of Theorem 6.2.2 Let us take the derivative of Q_{1n} and Q_{2n} in Lemma 6.2.1 with respect to β and t and get

$$\frac{\partial Q_{1n}(\beta, 0)}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial m_n(X_i, \boldsymbol{X}, \beta)}{\partial \beta}, \qquad \frac{\partial Q_{1n}(\beta, 0)}{\partial t^{tr}} = \frac{1}{n}\sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta)^{\otimes 2},$$

$$\frac{\partial Q_{2n}(\beta, 0)}{\partial \beta} = 0, \qquad \frac{\partial Q_{2n}(\beta, 0)}{\partial t^{tr}} = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{\partial m_n(X_i, \boldsymbol{X}, \beta)}{\partial \beta}\right)^{tr}.$$

Then take the Taylor expansion of $Q_{1n}(\hat{\beta}_n, \hat{t}_n)$ and $Q_{2n}(\hat{\beta}_n, \hat{t}_n)$ at $(\beta_0, 0)$,

$$\begin{aligned}0 &= Q_{1n}(\hat{\beta}_n, \hat{t}_n) \\ &= Q_{1n}(\beta_0, 0) + \frac{\partial Q_{1n}(\beta_0, 0)}{\partial \beta}(\hat{\beta}_n - \beta_0) + \frac{\partial Q_{1n}(\beta_0, 0)}{\partial t^{tr}}(\hat{t}_n - 0) + o_P(\delta_n),\end{aligned}$$

$$\begin{aligned}0 &= Q_{2n}(\hat{\beta}_n, \hat{t}_n) \\ &= Q_{2n}(\beta_0, 0) + \frac{\partial Q_{2n}(\beta_0, 0)}{\partial \beta}(\hat{\beta}_n - \beta_0) + \frac{\partial Q_{2n}(\beta_0, 0)}{\partial t^{tr}}(\hat{t}_n - 0) + o_P(\delta_n),\end{aligned}$$

where $\delta_n = \|\hat{\beta}_n - \beta_0\| + \|\hat{t}_n\|$.

$$\begin{pmatrix} \hat{t}_n \\ \hat{\beta}_n - \beta_0 \end{pmatrix} = \hat{S}_n^{-1} \begin{pmatrix} -Q_{1n}(\beta_0, 0) + o_P(\delta_n) \\ o_P(\delta_n) \end{pmatrix},$$

where

$$\hat{S}_n = \begin{pmatrix} \dfrac{\partial Q_{1n}}{\partial t^{tr}} & \dfrac{\partial Q_{1n}}{\partial \beta} \\ \dfrac{\partial Q_{2n}}{\partial t^{tr}} & 0 \end{pmatrix} \xrightarrow[(\beta=\beta_0, t=0)]{P} \begin{pmatrix} \hat{S}_{11} & \hat{S}_{12} \\ \hat{S}_{21} & 0 \end{pmatrix} = \begin{pmatrix} -E\left\{m_n m_n^{tr}\right\} & E\left\{\dfrac{\partial m_n}{\partial \beta}\right\} \\ E\left\{\dfrac{\partial m_n}{\partial \beta}\right\} & 0, \end{pmatrix}, \tag{6.60}$$

where in (6.60), $m_n = m_n(X_i, \boldsymbol{X}, \beta_0)$.

$$Q_{1n}(\beta_0, 0) = \frac{1}{n} \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta_0) = O_P(n^{-1/2})$$

implies that $\delta_n = O_P(n^{-1/2})$. Therefore

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \hat{S}_{22.1}^{-1} \hat{S}_{21} \hat{S}_{11}^{-1} \sqrt{n} Q_{1n}(\beta_0, 0) + o_P(1),$$

where

$$S_{22.1} = \left\{ E\left(\frac{\partial m_n}{\partial \beta}\right)^{tr} (E m_n m_n^{tr})^{-1} E\left(\frac{\partial m_n}{\partial \beta}\right) \right\}.$$

Furthermore, by Lemma 6.2.1, $\sqrt{n} Q_{1n}(\beta_0, 0) \to N(0, \tilde{\Sigma})$, hence

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \to N(0, V), \quad \text{where } V = \hat{S}_{22.1}^{-1} \hat{S}_{21} \hat{S}_{11}^{-1} \tilde{\Sigma} \left\{ \hat{S}_{22.1}^{-1} \hat{S}_{21} \hat{S}_{11}^{-1} \right\}^{tr}. \tag{6.61}$$

□

Chapter 7: Computational Results

This chapter is devoted to showing some simulation results under the right censored *Accelerated Lifetime Model* (AFT) using *Extended Estimating Equation* (EEE) and *Profile Empirical Likelihood* (pEL) with R (3.4.1). Let T be the lifetime, and consider the model

$$Y = \beta_0^{tr} Z + \varepsilon, \tag{7.1}$$

where $Y = \ln(T)$, $\beta_0 \in \mathbb{R}^p$ is the structural parameter, Z is the p-dimensional covariate and ε is the error term centered at zero with unknown hazard rate function. We assume that T may be subject to right censoring C with unknown distribution function, therefore the data we actually observe are triplets (V, Δ, Z), where

$$V = \min(\ln(T), \ln(C)), \text{ and } \Delta = I\left\{T \leq C\right\}. \tag{7.2}$$

In previous chapters, we proposed two ways to construct an estimator for β_0, which were $\tilde{\beta}_n$ via solving *Extended Estimating Equation* (EEE), and $\hat{\beta}_n$ via minimizing the negative logarithm of the pEL, i.e.:

$$\text{EEE: } \tilde{\beta}_n \text{ such that } S_n(\tilde{\beta}) = \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \tilde{\beta}_n) = 0; \tag{7.3}$$

$$\text{pEL: } \hat{\beta}_n = \arg\min_{\beta} l(\beta), \quad \text{where } l(\beta) = \sum_{i=1}^{n} \ln\left(1 + t^{tr}(\beta) m_n(X_i, \boldsymbol{X}, \beta)\right) \tag{7.4}$$

where $t(\beta) = t(\beta, \boldsymbol{X})$ is the solution to the gradient of the Lagrangian function defined in (5.50), and under model assumption (7.1),

$$m_n(X_i, \boldsymbol{X}, \beta) = \sum_{i=1}^{n} \Delta_i \left\{ Z_i - \frac{\sum_{j=1}^{n} Z_j I\left\{V_j - \beta^{tr} Z_j \geq V_i - \beta^{tr} Z_i\right\}}{\sum_{j=1}^{n} I\left\{V_j - \beta^{tr} Z_j \geq V_i - \beta^{tr} Z_i\right\}} \right\}. \tag{7.5}$$

The purpose of this chapter is to provide algorithms to numerically compute $\tilde{\beta}_n$ and $\hat{\beta}_n$. Since both $S_n(\beta)$ and $l(\beta)$ involve indicator functions, usual root-finding and optimization methods may not be directly applied here. Then we compare (7.3) and (7.4) with respect to the asymptotic behavior of $\tilde{\beta}_n$ and $\hat{\beta}_n$ under different parameter settings, the time-efficiency of the two methods, and the local continuity of $S_n(\beta)$ and $l(\beta)$. We also compare the empirical variance-covarince matrices of $\tilde{\beta}_n$ and $\hat{\beta}_n$ with the corresponding theoretical ones, and with each other.

7.1 Description of the Algorithm Associated with the EEE

In this section, we outline the algorithms for computing $\hat{\beta}_n$ and $\tilde{\beta}_n$. To summarize, we compute $\hat{\beta}_n$ in two steps, namely, first to construct a function $t(\beta)$ that expresses the Lagrange multiplier in terms of β, and second to calculate $\hat{\beta}_n$ by maximizing the pEL. We compute $\tilde{\beta}_n$ by minimizing the Euclidean norm of $S_n(\beta)$ because, due to the discontinuity issue of $S_n(\beta)$, a root may not exist.

7.1.1 Algorithm for Computing the pEL Estimator

Given the AFT model described in (7.1), for fixed β, to find the probability vector that maximizes the empirical likelihood, we solve the following problem,

$$\begin{cases} \arg\max_{\boldsymbol{p}} \prod p_i, \quad \text{where } \boldsymbol{p} = (p_1, \ldots, p_n), \\ \text{subject to } \sum_{i=1}^{n} p_i = 1, p_i \in (0,1), \sum_{i=1}^{n} p_i m_n(X_i, \boldsymbol{X}, \beta) = 0. \end{cases} \tag{7.6}$$

We have proved in Chapter 5 that with probability approaching 1, there exists a neighbourhood of β_0 in which (7.6) has a unique solution given by

$$\hat{p}_i = \frac{1}{n} \cdot \frac{1}{1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta)}, \tag{7.7}$$

$$\text{where } t \text{ is the solution to } \sum_{i=1}^{n} \frac{m_n(X_i, \boldsymbol{X}, \beta)}{1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta)} = 0. \tag{7.8}$$

Therefore the algorithm to compute $\hat{\beta}_n$ is divided into two steps. First, write t as a function of β and $\boldsymbol{X}$ according to (7.8); then after combining (7.1) and (7.7), $\hat{\beta}_n$ is given by

$$\hat{\beta}_n = \arg\min_{\beta} \sum_{i=1}^{n} \ln\left(1 + t(\beta)^{tr} m_n(X_i, \boldsymbol{X}, \beta)\right) \tag{7.9}$$

In the following paragraph, we explain how these two steps are performed using software R (3.4.1).

Step1: Construct $t(\beta)$ To construct $t(\beta)$ we convert the root solving problem of equation (7.8) into an optimization problem. For any fixed β, let

$$t^* = \arg\max_{t} f(t), \quad \text{where } f(t) = \sum_{i=1}^{n} \ln(1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta)) \tag{7.10}$$

Since we seek the calculus maximum of (7.10), the solution t^* must satisfy

$$\nabla_t f(t^*) = 0, \quad \text{where } \nabla_t f(t) = \sum_{i=1}^{n} \frac{m_n(X_i, \boldsymbol{X}, \beta)}{1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta)} \tag{7.11}$$

In other words, for any fixed β, we can construct $t(\beta)$ via maximizing in (7.10) using the `nlm` function built in R. For any fixed β and $\boldsymbol{X}$, it is easy to verify that $\nabla_t^{\otimes 2} f(t)$ is always negative definite, therefore the solution of (7.11) is guaranteed to be a calculus maximum, if it exists. Since $t_{\text{initial}} = (0, \ldots, 0)$ always provides a legitimate probability vector (7.7), we use t_{initial} as the initial value for the `nlm` function in this step.

We also need to pay attention to the domain of $f(t)$, denoted by $\mathcal{D}_f$. For the univariate case, it is easy to show that

$$\mathcal{D}_f = \left(-(\max_i (m_n(X_i, \boldsymbol{X}, \beta)))^{-1}, -(\min_i (m_n(X_i, \boldsymbol{X}, \beta)))^{-1} \right).$$

However, for the vector valued β, the analytical form of $\mathcal{D}_f$ is no longer simple. In cases when $1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta)$ has negative components, $f(t)$ is no longer well-defined. Therefore we replace $\ln(\cdot)$ by a monotone function $h(\cdot)$ defined everywhere but extremely negative at feasible values t. This penalty function should guarantee that the interactive root-finding method like Newton-Raphson converges. Conditions on $h(z)$ include,

1. $h(z) = \ln(z)$ for any $z > \varepsilon$, where ε is a positive constant that is close to 0;

2. For any $z_1 \in \mathbb{R}^- \cup \{0\}$ and $z_2 \in \mathbb{R}^+$, $h(z_2) > h(z_1)$. This way, replacing the function $\ln(\cdot)$ by the penalty function will not change the solution to the maximization problem (7.10);

3. $h(z)$ is continuous and differentiable for $z \in \mathbb{R}$;

4. When seeking a root with Newton-Ralphson, if in the kth step $z_k < 0$, then $h'(z_k)$ should point to the direction such that $z_{k+1} > 0$. In other words, if $z < 0$, then $h'(z)$ should always guarantee $z - h(z)/h'(z) > 0$.

A feasible choice of such a function is

$$h(z) = \begin{cases} \ln(z) & \text{when } z > \varepsilon \\ \ln(\varepsilon) + \dfrac{z-\varepsilon}{\varepsilon^2} & \text{when } z \leq \varepsilon \end{cases} \tag{7.12}$$

Using $h(z)$, for any fixed β, we define $t = t(\beta)$ as

$$t(\beta) \equiv \arg\max_t f_h(t), \quad \text{where } f_h(\beta) = \sum_{i=1}^{n} h(1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta)). \tag{7.13}$$

Step 2: Estimate β_0 After writing the Lagrange multiplier t as a function of β, we can compute $\hat{\beta}_n$ by minimizing the negative logarithm of the pEL, namely, $\hat{\beta}_n = \arg\min_\beta l(\beta)$, where

$$l(\beta) = \sum_{i=1}^{n} \ln\left(1 + t^{tr}(\beta) m_n(X_i, \boldsymbol{X}, \beta)\right) \tag{7.14}$$

The quasi-Newton-Raphson methods do not work for (7.14) under the AFT model assumption due to the appearance of indicator functions in $m_n(X_i, \boldsymbol{X}, \beta)$. Therefore we use the default Nelder-Mead method in the `optim` function of R, which is a topological method that can be applied to nonlinear optimization problems for which derivatives may not exist. We use the least- square estimator on the uncensored data

denoted by b_0 as the initial value, where

$$\begin{aligned}
&\text{Uncensored Index Set: } \mathcal{R} = \{i : \Delta_i = 1\}, \quad \text{where } \#\mathcal{R} = u \\
&\text{Vector of Uncensored Lifetime: } O = (V_k; k \in \mathcal{R}) \in \mathbb{R}^u; \\
&\text{Matrix of Covariates with Uncensored Observations: } H = (Z_k : k \in \mathcal{R}) \in \mathbb{R}^{p \times u}
\end{aligned} \tag{7.15}$$

then $b_0 = (H^{tr}H)^{-1}H^{tr}O$ and $\hat{\beta}_n$ is computed by `optim(b0,l)$par`. This choice of b_0 is good for uncensored AFT, but in the right-censored case, a more reasonable initial choice is the estimator of Koul, Sursarla and van Ryzin in [25].

Restarting Improvement Ideally, with a reasonable initial value, Nelder-Mead method should give the solution to the minimization problem with a signle application of `optim`. However, the optimization procedure with respect to β turns out to require multiple restarts. More specifically, let b_0 be the least square estimator calculated on set $\mathcal{S}$, and

$$b_{r+1} = \arg\min_{\beta} l(\beta) \quad \text{with initial value } b_r,\ r = 0, 1, \ldots \tag{7.16}$$

Let $\|\cdot\|_1$ be the L^1 norm on vectors. The sequence of restarts stops when $\|b_r - b_{r+1}\|_1 \leq 0.01$. Ideally, the sequence (7.16) should stop at $r = 0$. However, we found that this ideal case only happens when the initial value is good enough, which occurs especially when the data are uncensored, or using the Koul-Susarla-van Ryzin estimator in [25] as b_0 for right censored data. On the contrary, when the simple least square estimator is far away from β_0, then the restarting improvement is necessary.

For example, when

$$\begin{aligned} &Z_1 \sim \text{Bernoulli}(0.5), \quad Z_2 \sim N(0,1), \quad Z_3 \sim \ln(F_{3,5}), \\ &Z_4 \sim \ln(\text{Beta}(5,3)), \quad \varepsilon = \text{Weibull}(1) - 1,\ C = \text{Exp}(5.8); \end{aligned} \tag{7.17}$$

for sample size $n = 200$ and batches of simulations of size $m = 1000$, the number of iterations r are 1, 2 and 3, with frequencies 780, 209 and 11.

7.1.2 Algorithm for Computing the Z-Estimator

As we discussed in previous chapters, we can construct an estimator for β_0 by solving the extended estimating equation $S_n(\beta) = 0$. Due to the discontinuity of $S_n(\beta)$, instead of directly solving the equation, we define

$$\tilde{\beta}_n \equiv \arg\min_{\beta} \|S_n(\beta)\|, \tag{7.18}$$

where $\|\cdot\|$ denotes the Euclidean norm. Again, we minimize using `optim` with initial value b_0, i.e., the least square estimator on set $\mathcal{R}$.

As for the computation of $\hat{\beta}_n$, the Nelder-Mead method is not guaranteed to give the local solution to the minimization problem (7.18) in one step. We use the same restarting strategy described on page 146. In the following section, we will show that the negative profile log-likelihood function $l(\beta)$ is much smoother than $S_n(\beta)$, therefore we should expect to require more restarts in the calculation of $\tilde{\beta}_n$ than $\hat{\beta}_n$. This conjecture will be verified numerically in Section 7.4

7.2 Local Continuity of EEE and pEL

In this section, we compare the local continuity of the EEE $S_n(\beta)$ with the negative logarithm of the pEL function $l(\beta)$. From the plots of $l(\beta)$ and $S_n(\beta)$, we will see clearly that the pEL is much smoother than the EEE. At the end of this section, we define a quantity tot.dif that measures the continuity pattern in the neighborhood of $\hat{\beta}_n$, β_{LS}, and β_0. From the histograms tot.dif, we will see that continuity behavior in the neighborhoods of $\hat{\beta}_n$ and β_0 are very similar.

7.2.1 Plots of Projection of EE and EL

Since both $S_n(\beta)$ and $l(\beta)$ are defined in $\mathbb{R}^p$, it is not easy to provide a plot directly. Therefore we plot only one "slice" at a time, namely, project $S_n(\beta)$ and $l(\beta)$ with respect to β^i, where in general, for $v \in \mathbb{R}^p$, v^i is the ith component of v.

More specifically, for EEE

$$S_n(\beta) = \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta), \quad \beta \in U_{\beta_0}, \tag{7.19}$$

consider function $d_S(s, i, j)$

$$d_S(s; \beta, i, j) = S_n(\beta + se_j) \cdot e_i, \tag{7.20}$$

where $\beta \in U_{\beta_0}$, $s \in (-\delta, \delta)$ with δ being a small positive number, $i, j = 1, \ldots, p$, and e_i is the ith column of the $p \times p$ identity matrix. The function $d_S(s; \beta, i, j)$ allows β to change only in the direction of e_j, then records the value of of the ith component in $S_n(\beta + s \cdot e_j)$ as output.

As for the negative log empirical likelihood function, consider

$$d_l(s;\beta,u,i) = l(\beta + s\cdot u^i), \quad i = 1,\ldots,p, \tag{7.21}$$

where $\beta \in U_{\beta_0}$, and $s \in (-\delta,\delta)$ for a small positive number δ, and u is a randomly generated unit direction. In the simulation, we use `rnorm(p)` to generate u, then divide it by its Euclidean norm.

We run the simulation under two parameter settings as follows, for $\beta_0 = (4,3,2,1)$ with normally distributed covariates in Set 1, and none-normally distributed covariates in Set 2,

$$\begin{aligned} \text{Set 1: } & Z_1 \sim \text{Bernoulli}(0.5),\ Z_2 \sim N(0,1),\ Z_3 \sim N(3,25), \\ & Z_4 \sim N(5,9),\ \varepsilon \sim \text{Weibull}(1) - 1,\ C \sim \text{Exp}(0.015) \end{aligned} \tag{7.22}$$

and

$$\begin{aligned} \text{Set 2: } & Z_1 \sim \text{Bernoulli}(0.5),\ Z_2 \sim N(0,1),\ Z_3 \sim F_{3,5} \\ & Z_4 \sim \text{Beta}(5,3),\ \varepsilon \sim \text{Weibull}(1) - 1, C \sim \text{Exp}(.08) \end{aligned} \tag{7.23}$$

With $n = 1000$ for $m = 1$ replica. For this Set 1 and Set 2, censoring rates are given by 18.6% and 33.2%. $\hat{\beta}_n$ and $\tilde{\beta}_n$ in both sets are identical to the second decimal place, and are $(4.01, 3.00, 1.99, 0.99)$ and $(3.86, 3.09, 2.05, 0.80)$, respectively. We plot d_S for both Set 1 and Set 2. As we can see from Figure 7.1 and 7.2, even for a large sample size $n = 1000$, within a small neighbourhood of true β_0, the extended estimating equation has a lot of jumps.

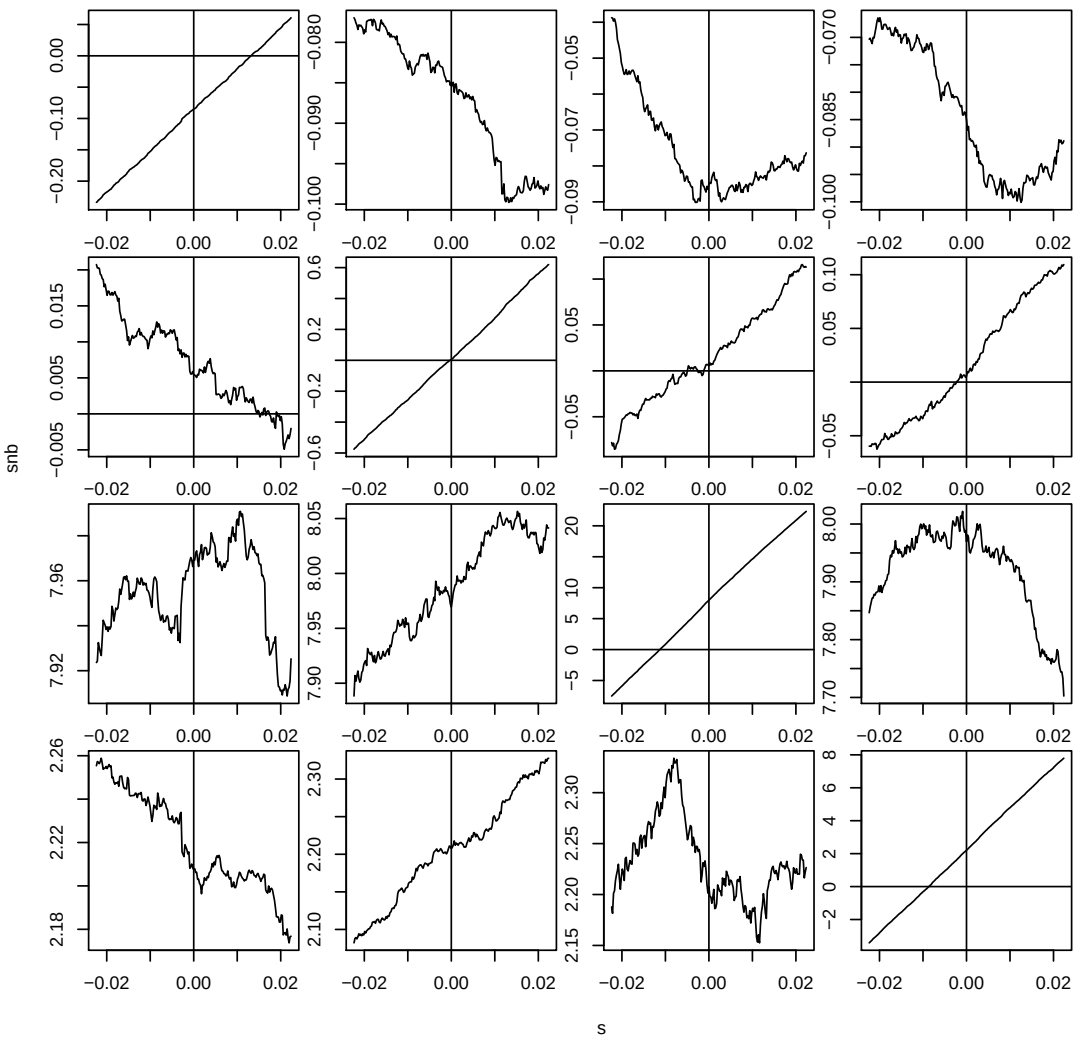

Figure 7.1: Plot of $d_S(s; \beta_0, i, j)$, normal covariates

Figure 7.1 shows a plot of $d_S(s; \beta_0, i, j) = S_n(\beta_0 + se_j)^{tr} e_i$ for $s \in (-0.02, 0.02)$, where $i, j = 1, \ldots, p$ index horizontal and vertical plots. Horizontal lines indicate level 0. It implies that the function is only linear and smooth on the diagonal. The magnitude of the y-coordinates indicates $S_n(\beta)$ is close to a diagonal matrix times β.

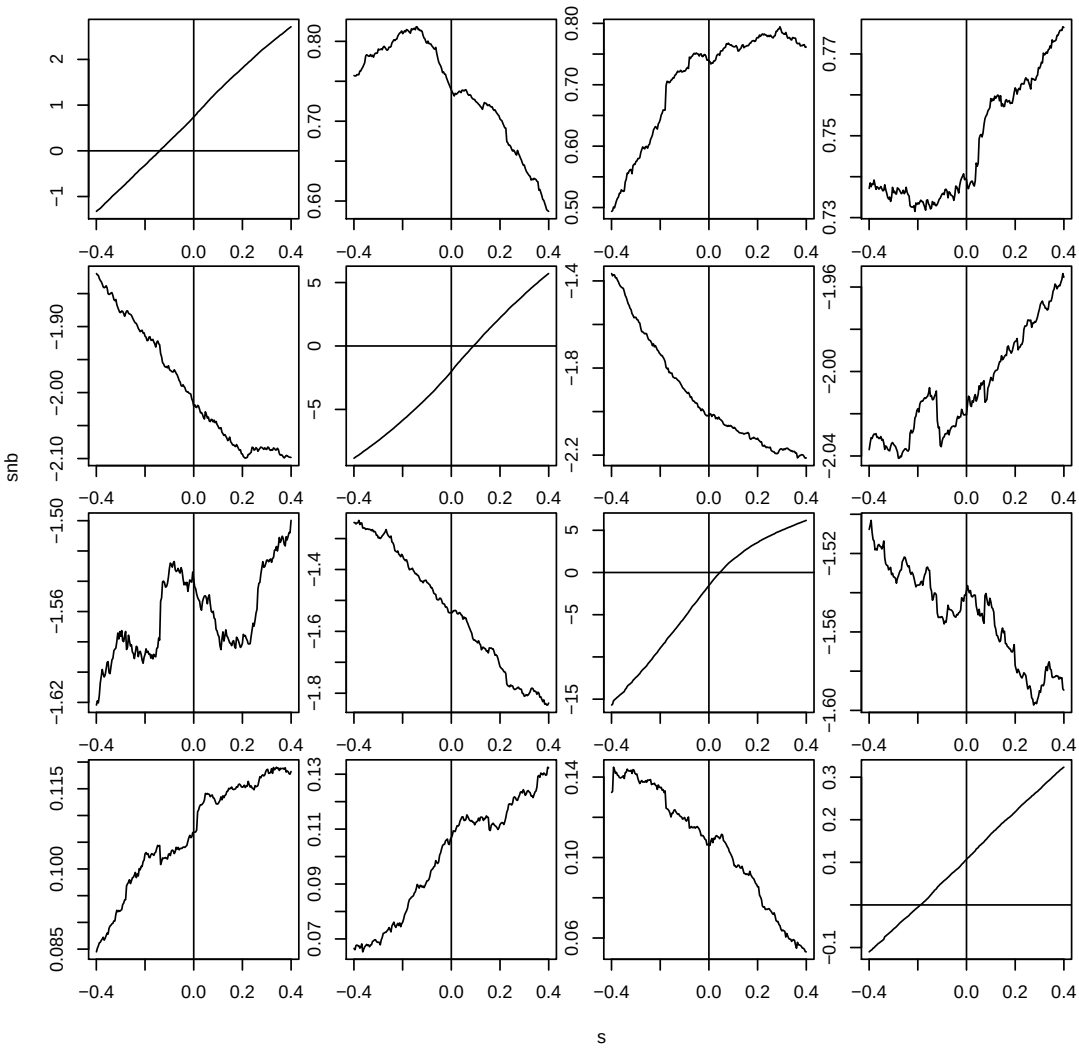

Figure 7.2: Plot of $d_S(s; \beta_0, i, j)$, non-normal covariates

Figure 7.2 shows a plot of $d_S(s; \beta_0, i, j) = S_n(\beta_0 + se_j)^{tr} e_i$ for $s \in (-0.4, 0.4)$, where $i, j = 1, \ldots, p$ index horizontal and vertical plots. Horizontal lines indicate level 0. It implies that the function is only linear and smooth on the diagonal. The magnitude of the y-coordinates indicates $S_n(\beta)$ is close to a diagonal matrix times β.

Then we check the smoothness of the negative profile log likelihood $l(\beta)$ for β is in a neighbourhood of $\hat{\beta}_n$ by plotting $d_l(s; \beta_n, u, i)$. Figure 7.3 and 7.4 are calculated under Set 1 (normal covariates) and Set 2 (non-normal covariates), respectively. The solid vertical line in each picture denotes $\hat{\beta}_n$, and the dotted vertical line denotes the true β_0. As we can see from Figure 7.3 and 7.4, the distance between the estimated value and the true parameter value are small, the profile likelihood function is very smooth within the neighborhood of $\hat{\beta}_n$, and $\hat{\beta}_n$ is the calculus maximum of $l(\beta)$ in each randomly generated directions for both normal and non-normal covariates.

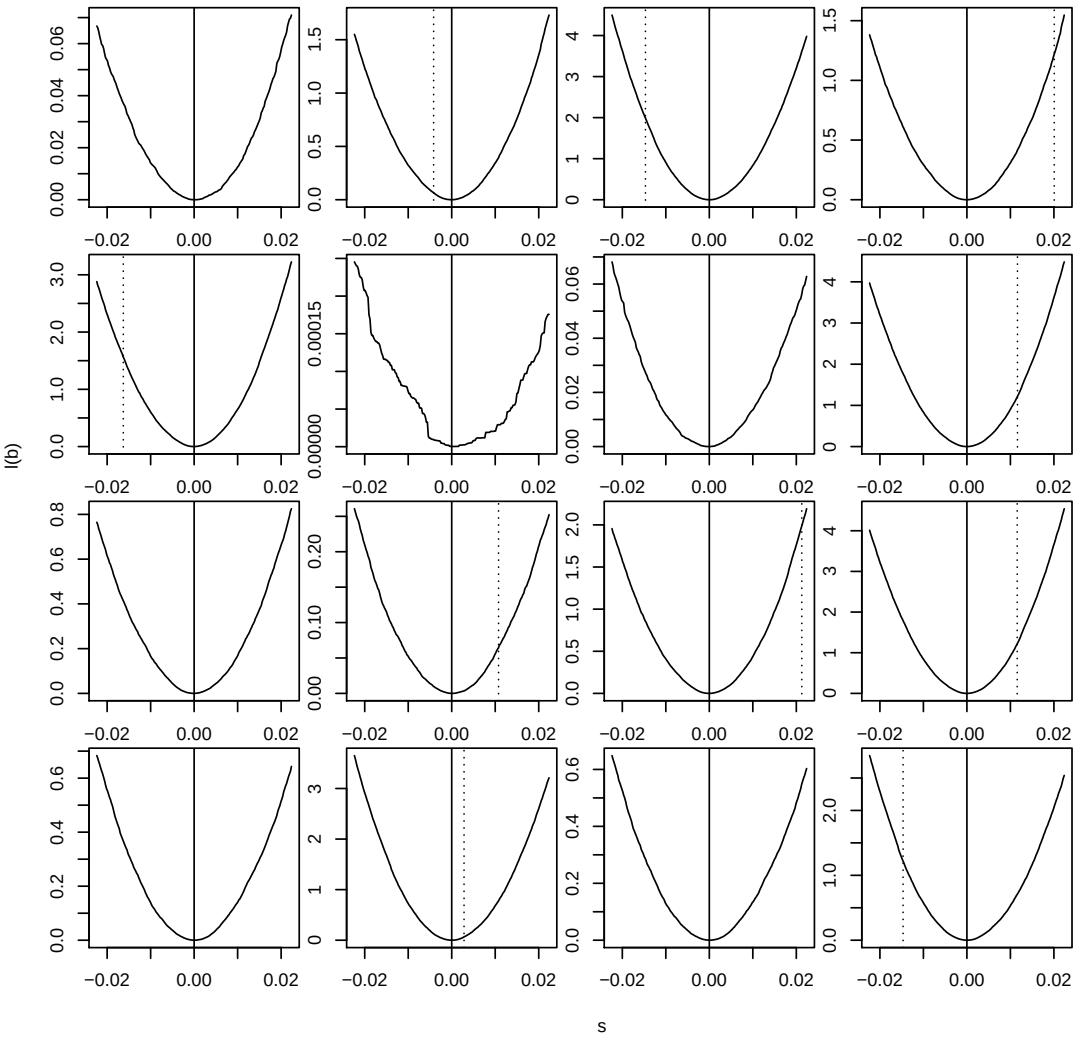

Figure 7.3: Plot of $d_l(s; \hat{\beta}_n, u, i)$, normal covariates

Figure 7.3 shows a plot of $d_l(s; \hat{\beta}_n, u, i) = l(\hat{\beta}_n + su^i)$ for $s \in (-0.02, 0.02)$, where $i, j = 1, \ldots, p$ index horizontal and vertical plots. The solid vertical line in each picture denotes $\hat{\beta}_n$, and the dotted vertical line denotes the true β_0. The picture indicates on any random direction, $l(\beta)$ is smooth and has a parabolic form, and $\hat{\beta}_n$ is the calculus minimizer of $l(\beta)$.

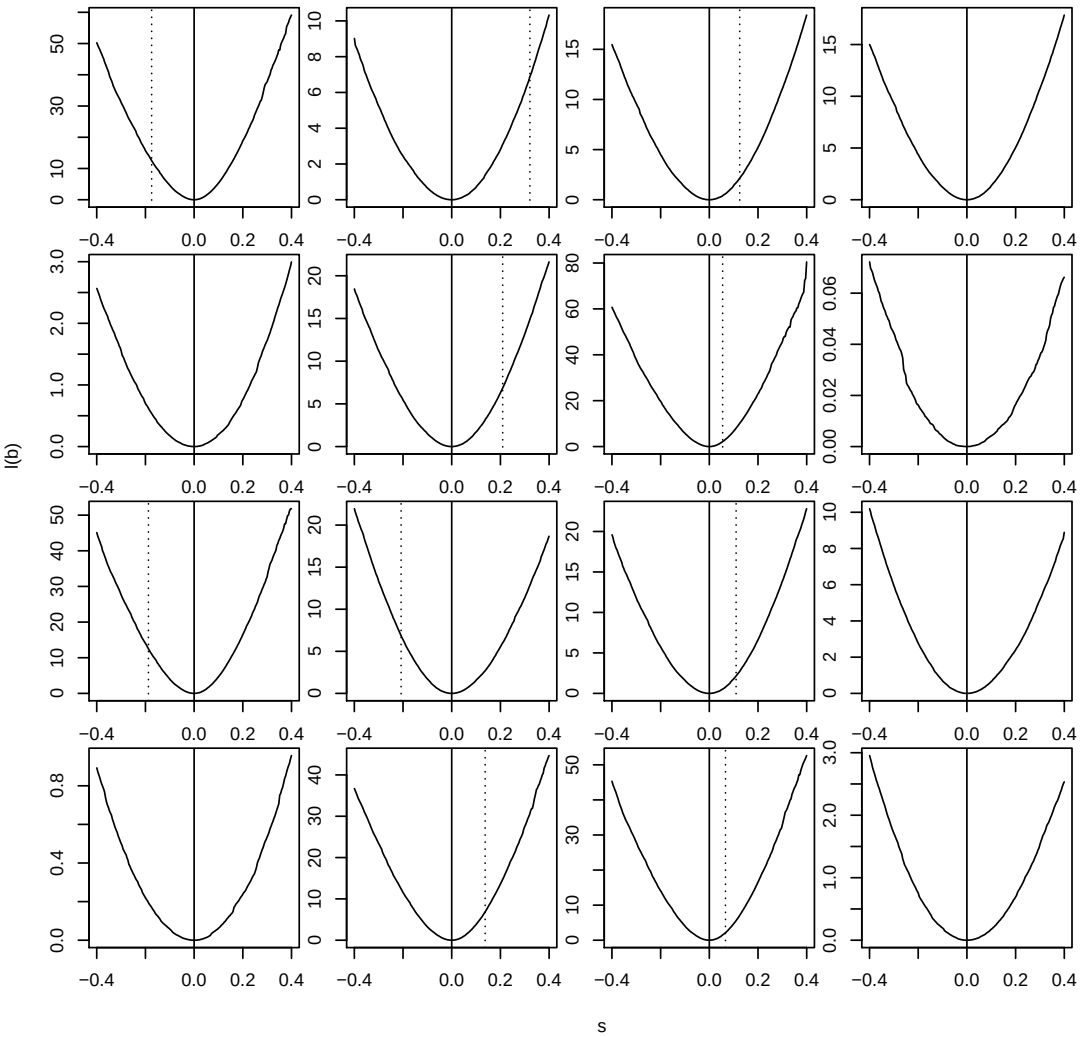

Figure 7.4: Plot of $d_l(s; \hat{\beta}_n, u, i)$, non-normal covariates

Figure 7.4 shows a plot of $d_l(s; \hat{\beta}_n, u, i) = l(\hat{\beta}_n + su^i)$ for $s \in (-0.4, 0.4)$, where $i, j = 1, \ldots, p$ index horizontal and vertical plots. The solid vertical line in each picture denotes $\hat{\beta}_n$, and the dotted vertical line denotes the true β_0. The picture indicates on any random direction, $l(\beta)$ is smooth and has a parabolic form, and $\hat{\beta}_n$ is the calculus minimizer of $l(\beta)$.

7.2.2 Quantitative Measurement of Local Continuity of $l(\beta)$

In fact, the continuity of the negative profile log-likelihood function $l(\beta)$, the local continuity around β_0 can be quantified using the total difference of the approximated derivative for function $d_l(s;\beta,i)$ in the following steps,

1. For fixed, $\beta \in U_{\beta_0}$, $s \in (-\delta, \delta)$ with δ being a small positive number,

$$d_l(s;\beta,i) = l(\beta + s \cdot e_i), \quad , i = 1, \ldots, p \tag{7.24}$$

   where e_i is the ith row of $p \times p$ identity matrix.

2. For $i = 1, \ldots, p$, approximate the derivative of $d_l(s)$ using

$$der(s;\beta,i) = \frac{1}{2\varepsilon_1}(d_l(s+\varepsilon_1;\beta,i) - d_l(s-\varepsilon_1;\beta,i)). \tag{7.25}$$

3. Generate a grid of points over interval $(-a, a)$ with grid length ε_0. Calculate $der(s;\beta,i)$ at each grid point. Record the output in a $p \times l$ matrix *out*, where l is the number of grid points. $out_{i,j}$ approximates the derivative of $l(\beta)$ in the direction of β^i at the jth grid point, where $i = 1, \ldots p$ and $j = 1, \ldots, l$.

4. For each row of the matrix *out* Step 3, calculate the absolute value of the difference between the consecutive components,

$$abs.d_{i,j} = |out_{i,j} - out_{i,j+1}|, \quad i = 1, \ldots, p, j = 1, \ldots, l-1, \tag{7.26}$$

   then record the row sum these differences, i.e.,

$$tot.d_i = \sum_{j=1}^{l-1} abs.d_{i,j}, \quad \text{where } i = 1, \ldots, p. \tag{7.27}$$

Remark 7.2.1 *We do not have a universal criterion to choose a, ε_0 and ε_1 that is guaranteed to work for every parametric setting. Since we are only interested in the local continuity around a fixed β, ε_0 should not be too large. The choice of ε_1 depends on ε_0. In general, the magnitude of $\varepsilon_1/\varepsilon_0$ should not be too large, in which case, the total difference cannot reflect the subtle differences for different choices of β; on the other hand, if the ratio is too small, then due to the appearance of the indicator functions, the total difference would be large for any choice of β.*

The quantity $tot.d_i$ should be much smaller for smooth functions than non-smooth functions. We also expect the total difference around $\hat{\beta}_{n,m}$ to be similar to the total difference of around β_0, where m denotes the number of simulations. To test these conjecture, we use the same simulated data sets in (7.22) in Section 7.1. The parameters for grids and difference quotients are $a = .5$, $\varepsilon_0 = .01$, $\varepsilon_1 = .001$. We ran the simulation for $m = 1000$ times with sample size 200. The censoring rate ranges from 22% to 42% and the simulation means are (3.954,2.970,1.919,0.974).

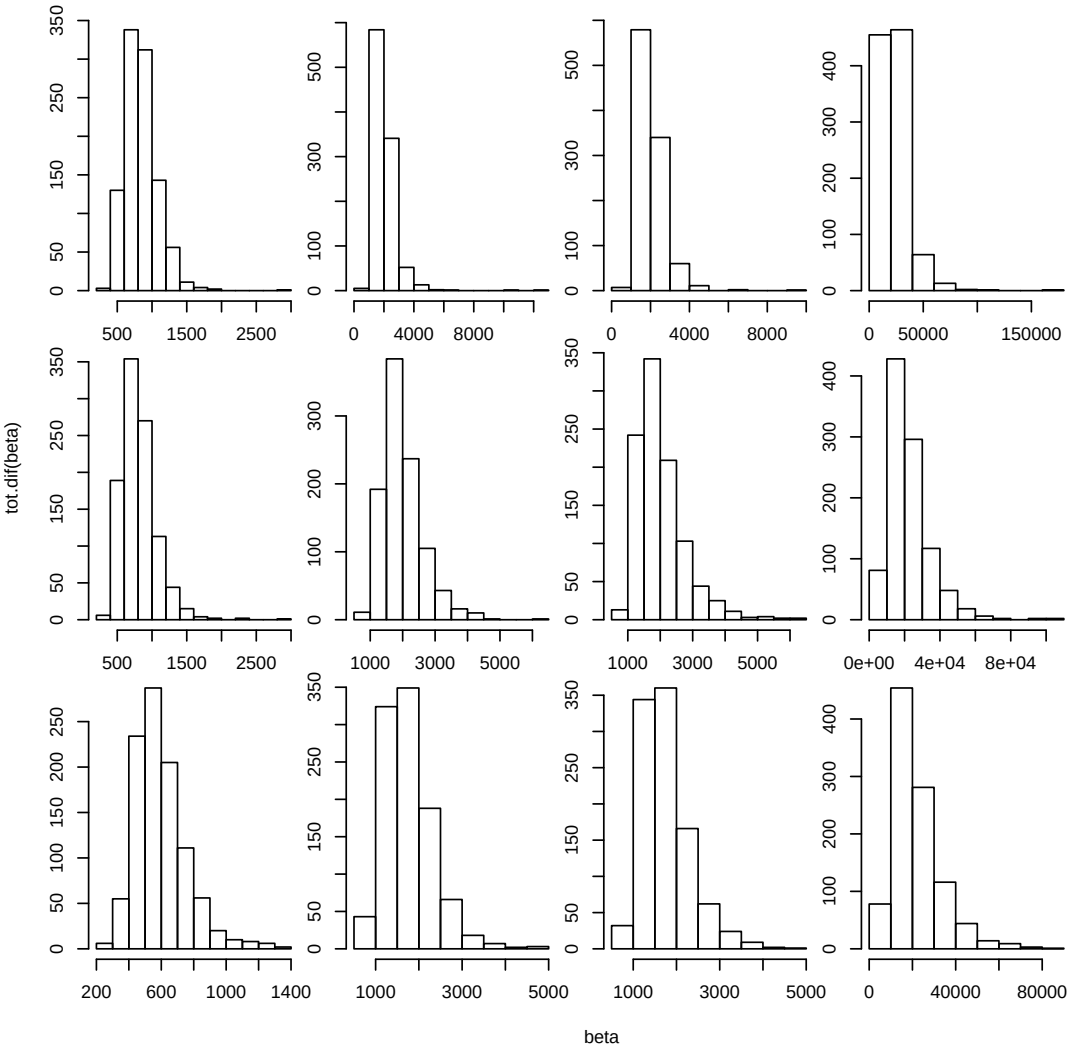

Figure 7.5: Plot of $tot.diff$, normal covariates

This picture shows the histogram of total differences defined in (7.27), where the first row is for β_0, second for $\hat{\beta}_n$, and third for β_{LS}. The shapes and ranges of the histogram indicates that the $\hat{\beta}_n$ has a similar continuity pattern compared with β_0, and such pattern cannot be preserved when β is far from β_0 evidenced by the histogram of $\hat{\beta}_{LS}$.

7.3 Convergence of the Estimator

Since $t^* = t(\beta)$ is the calculus maxima of $f(t) = \sum_{i=1}^{n} h(1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta))$ for each fixed β, the gradient $\nabla_t f(t)$ at t^* must be zero, i.e.,

$$g(t^*, \beta) = \sum_{i=1}^{n} \frac{m_n(X_i, \boldsymbol{X}, \beta)}{1 + t^{*tr} m_n(X_i, \boldsymbol{X}, \beta)}, \quad \text{where } t^* = t(\beta). \tag{7.28}$$

Therefore at $\hat{\beta}_n$, the gradient $g(t(\hat{\beta}_n), \hat{\beta}_n)$ should be close to zero regardless of parametric setting or sample size. To test this claim, we did experiments under $\beta_0 = (1, 2, 3, 4)$

$$\begin{aligned} \text{Set 3: } & Z_1 = \text{Bernoulli}(0.5),\ Z_2 = \ln(\Gamma(3)),\ Z_3 = \ln(F_{3,5}), \\ & Z_4 = \ln(\text{Beta}(5,3)),\ \varepsilon \sim \text{Weibull}(1) - 1;\ C \sim \text{Exp}(0.5) \end{aligned} \tag{7.29}$$

sample sizes $n = 30$, 50 and 100 for $m = 100$ simulations. We found the gradient is always bounded by a very small number to the magnitude of $1e - 5$.

7.4 Asymptotic Normality of the Z-Estimator and pEL Estimator

In this section, we compare $\tilde{\beta}_{n,m}$, the solution to the EEE, and $\hat{\beta}_{n,m}$ the value that minimizes the negative logarithm of the pEL. Under three different parametric settings, we compare the center of $\tilde{\beta}_{n,m}$ and $\hat{\beta}_{n,m}$ as well as the empirical variances. We seek proper sample sizes n_1 and n_2 such that $\tilde{\beta}_{n_1,m}$ and $\hat{\beta}_{n_2,m}$ reaches asymptotic normality. We also compare the computational efficiency in terms of time lapsed and iterations that is needed for restarting improvement described on page 147. Throughout this section, $m = 1000$ if not otherwise specified.

7.4.1 Covariates without Normality

Let us start with a parametric setting with covariates that are not normally distributed. Let $\beta_0 = (0.1, 0.2, 0.3, 0.4)$, and consider Set 3 described in (7.29),

$$\begin{aligned} \text{Set 3: } & Z_1 = \text{Bernoulli}(0.5),\ Z_2 = \ln\left(\Gamma(3)\right),\ Z_3 = \ln\left(F_{3,5}\right), \\ & Z_4 = \ln\left(\text{Beta}(5,3)\right),\ \varepsilon \sim \text{Weibull}(1) - 1;\ C \sim \text{Exp}(0.5) \end{aligned} \tag{7.29}$$

Normality We found that the smallest sample size that is needed to attain normality is affected by the symmetricity of covariates. Checking the histograms, Z_1 and Z_3 are appears much more symmetric, the sample size that for $\tilde{\beta}_n^1$, $\tilde{\beta}_n^3$, $\hat{\beta}_n^1$ and $\hat{\beta}_n^3$ to reach normality is significantly smaller than that for $\tilde{\beta}_n^2$ and $\tilde{\beta}_n^4$, and $\hat{\beta}_n^2$ and $\hat{\beta}_n^4$. With $n = 100$, censoring rate ranging from 16% to 43% and mean equal to 29.3%. Based on the Shapiro test, $\tilde{\beta}_{n,m}^i$ and $\hat{\beta}_{n,m}^i$ reaches normality only when $i = 1$ and $i = 3$.

When the sample size $n = 200$, both $\tilde{\beta}_{n,m}^i$ and $\hat{\beta}_{n,m}^i$ pass the Shapiro test for $i = 1, \ldots, 4$. However, if we use the Mardia's Test for multivariate normality, then the sample size needs to be increased to 450. A histograms for $\tilde{\beta}_{450,1000}^i$ and scattered plot of $\tilde{\beta}_{450,m}^i$ against $\tilde{\beta}_{450,m}^i$ can be found in Figure 7.6

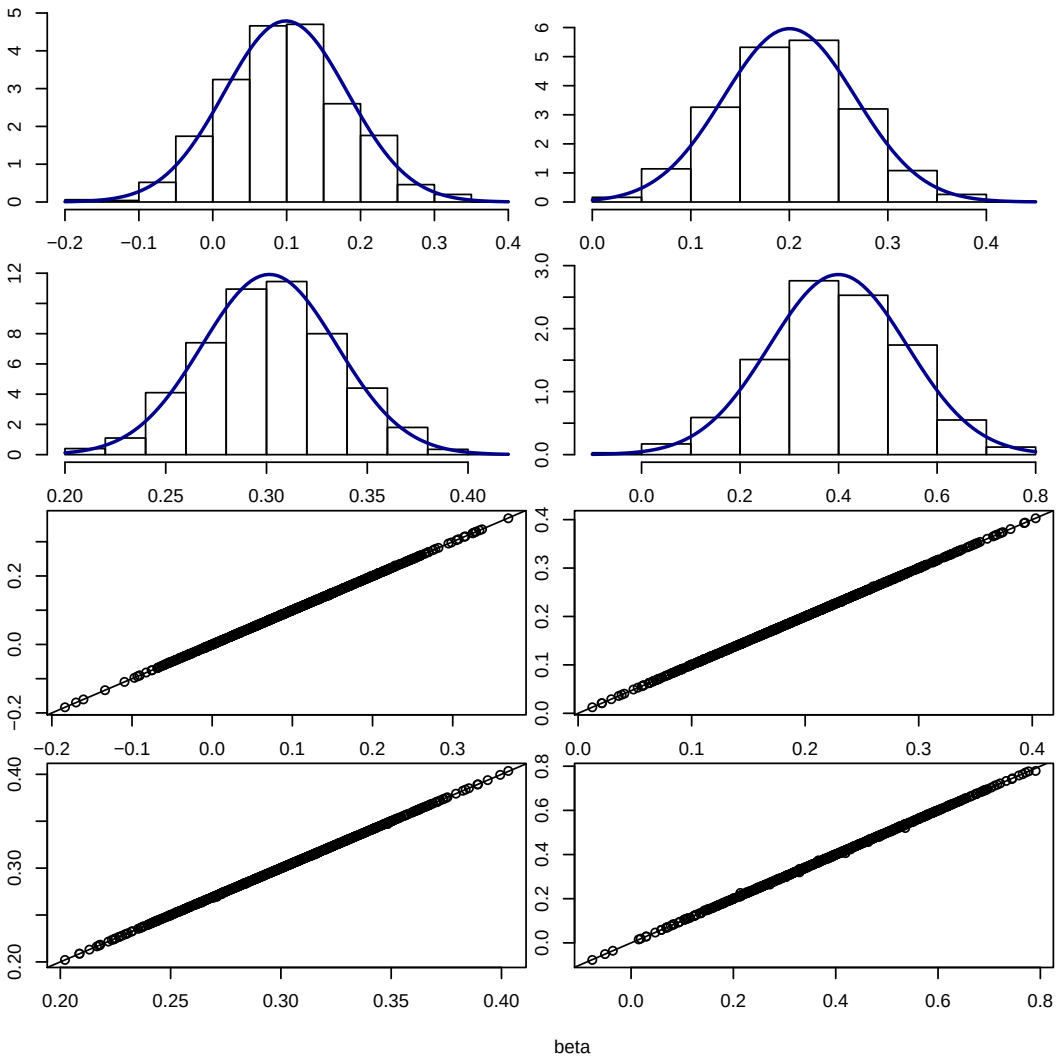

Figure 7.6: Histogram and QQ Plots for None-Normal Covariate Covariates

Accuracy and Asymptotic Variance Covariance The first thing that we compared was whether the two estimators $\hat{\beta}_{n,m}$ and $\tilde{\beta}_{n,m}$ differs a lot from replica to replica. In our simulation, we measure this difference using L^1 vector norm of $\tilde{\beta}_{n,m} - \hat{\beta}_{n,m}$, which should be very small theoretically for proper n. This is because when constructing $\hat{\beta}_{n,m}$, we first got $t = t(\beta)$ by

$$t^* = \arg\max_t f(t), \quad \text{where } f(t) = \sum_{i=1}^{n} \ln(1 + t^{tr} m_n(X_i, \boldsymbol{X}, \beta)) \tag{7.30}$$

for fixed β. Since t is a calculus maxima for $f(t)$,

$$\nabla_t f(t) = \sum_{i=1}^{n} \frac{m_n(X_i, \boldsymbol{X}, \beta)}{1 + t(\beta)^{tr} m_n(X_i, \boldsymbol{X}, \beta)} = 0 \tag{7.31}$$

On the other hand, $\tilde{\beta}_n$ is the solution to

$$S_n(\beta) = \sum_{i=1}^{n} m_n(X_i, \boldsymbol{X}, \beta) = 0 \tag{7.32}$$

Comparing (7.31) and (7.32) we see that $S_n(\beta)$ is exactly $\nabla f(t)$ when $t = 0$. On the other hand, theoretically, the solution to (7.30) has a very small magnitude relative to the order of $\|\beta - \beta_0\|$. Since for large n, $\|\hat{\beta}_n - \beta_0\|$ is small, therefore by the continuity of $\nabla_t f(t)$, $\tilde{\beta}_n$ and $\hat{\beta}_n$ should also be very close as well. We can observe this fact from the scattered plot of $\tilde{\beta}^i_{n,m}$ against $\hat{\beta}^i_{n,m}$ for $i = 1, \ldots, p$. For $n = 100$, the dots is distributed closely to the line $y = x$. As n grows to 150 and 200, the linear pattern is even more clear.

As for accuracy, even for $n = 100$, the mean and median for both $\tilde{\beta}_{n,m}$ and $\hat{\beta}_{n,m}$ are very close to the true β_0. We have also compared the empirical variance covariacne matrix of the two estimators denoted by cov_1 and cov_2. The largest eigenvalues of $cov_1 - cov_2$ for $n = 100$ is to the magnitude of n^{-3}.

Restarting Improvement and Timing We found that $\tilde{\beta}_{n,m}$ is always faster to compute than $\hat{\beta}_{n,m}$. This is because the latter one requires two steps of optimization, yet the $\tilde{\beta}_{n,m}$ requires only one step. As for the iterations that is needed for the restarting improvement, we found that $\tilde{\beta}_{n,m}$ always requires more rounds. This is not surprising because the negative logarithm of the pEL function $l(\beta)$ is much smoother than the extended estimating equation function $S_n(\beta)$. Consequently, it takes more restartings for to find the minimum value of $\|S_n(\beta)\|$.

7.4.2 Normally Distributed Covariates

We investigate similar aspects of $\tilde{\beta}_{n,m}$ and $\hat{\beta}_{n,m}$ under $\beta_0 = (0.51, 0.15, 1.18, 0.92)$

$$\begin{aligned} \text{Set 4: } & Z_1 \sim \text{Bernoulli}(0.5),\ Z_2 \sim N(0,1)\ Z_3 \sim N(0.3, 0.25); \\ & Z_4 \sim N(0.1, 0.09),\ \varepsilon \sim \text{Weibull}(1) - 1,\ C \sim \text{Exp}(0.5) \end{aligned} \tag{7.33}$$

When all the covariates are normally distributed, the n that for $\tilde{\beta}^i_{n,m}$ $\hat{\beta}^i_{n,m}$ to gain normality is much smaller, where $i = 1, \ldots, p$. For $n = 100$ with censoring rate ranging from 15% to 44% and centered at mean = 28%, all components of $\tilde{\beta}_{n,m}$ and $\hat{\beta}_{n,m}$ passed the Shapiro normality test. When $n = 400$, both estimators pass the Mardia test for multivariate normality. A histograms for $\tilde{\beta}^i_{400,1000}$ and scattered plot of $\tilde{\beta}^i_{400,m}$ against $\tilde{\beta}^i_{400,m}$ can be found in Figure 7.7

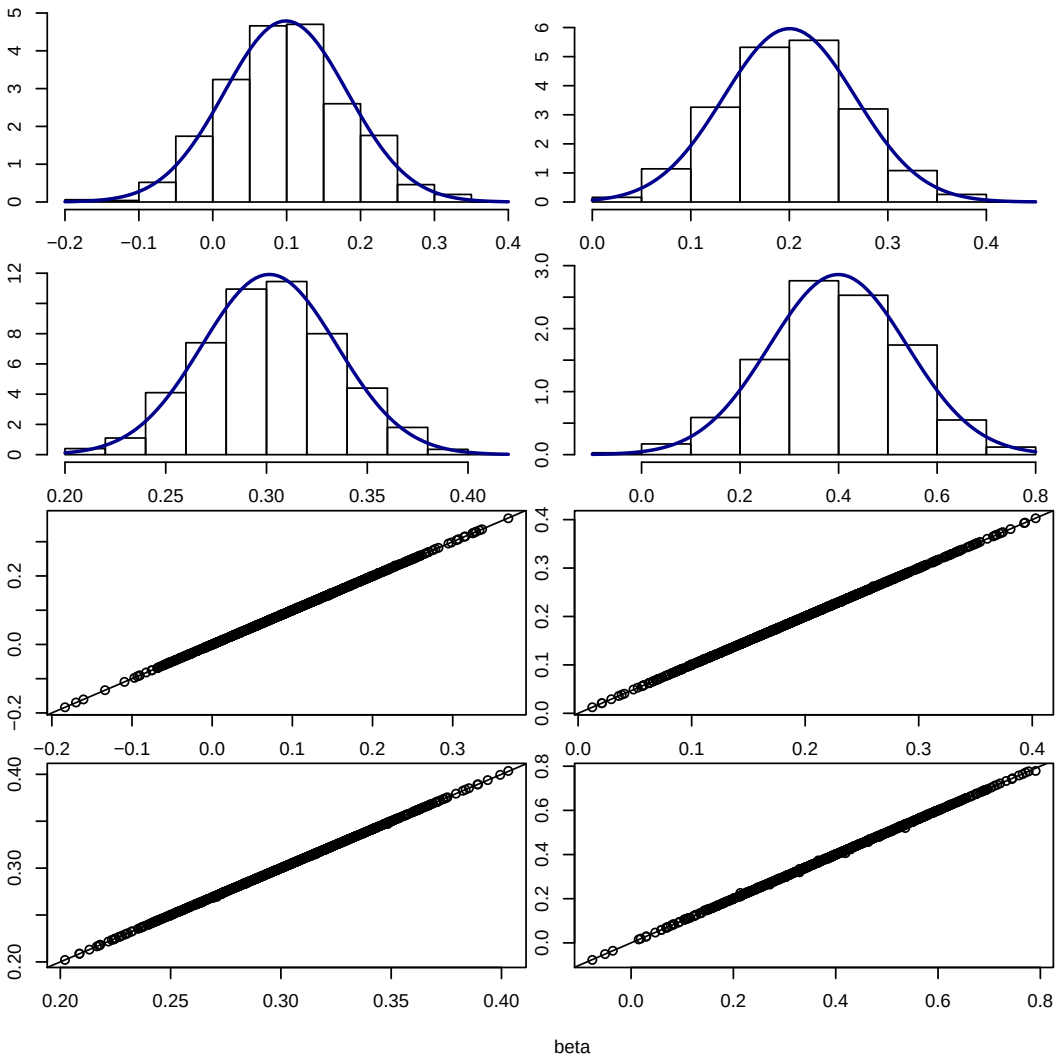

Figure 7.7: Histogram and QQ Plots for Normal Covariate Covariates

Similar to the none-normal setting, the L^1 norm of $\tilde{\beta}_{n,m} - \hat{\beta}_{n,m}$ is always small evidenced by the the scattered plot of $\tilde{\beta}^i_{n,m} - \hat{\beta}^i_{n,m}$ for $i = 1, \ldots, 4$, on which the points are distributed closely along the line $y = x$. The magnitude of the absolute value of eigen$\Big(cov(\tilde{\beta}_{n,m}) - cov(\hat{\beta}_{n,m})\Big)$ is also small. When $n = 450$, the eigenvalue with the largest absolute value is 1.66e-05

Under Set 4, the the EL method is still more costly in terms of time. However, like in Set 3, the number of iterations needed for restarting improvement is much smaller for the EL method. When $n = 400$, 938 replicas are finished by only 1 restarting and the rest 62 are finished by in 2 restarting iterations for $\hat{\beta}_{n,m}$; in comparison, when calculating $\tilde{\beta}_{n,m}$, only 834 replicas are finished within 1 restarting, and the maximum number of iterations is 4.

7.4.3 Extreme Cases

In this section, we consider an "extreme" parameter setting, under which the lifetime t is either very large or very small. This is rarely seen in practice, however, we would like to compare the behavior of $\tilde{\beta}_n$ and $\hat{\beta}_n$ out of theoretical interest. Consider $\beta_0 = (0.4, 0.3, 0.2, 0.1)$,

$$\begin{aligned} \text{Set 5: } & Z_1 \sim \text{Bernoulli}(0.5),\ Z_2 \sim N(0,1),\ Z_3 \sim F_{3,5}, \\ & Z_4 \sim \text{Beta}(5,3),\ \varepsilon \sim \text{Weibull}(1) - 1,\ C \sim \text{Exp}(0.8) \end{aligned} \tag{7.34}$$

Under parametric setting (7.34), the censoring rate ranges between 25.8% to 38.6% with mean around 31.3%. From the histogram, we can see that T and V are severely skewed to the right with extremely large outliers, which happens because the co-

variate Z_3 is severely skew to the right. Therefore we expect the estimator for β_0^3 to gain asymptotic normality with a much larger n.

For both $\tilde{\beta}_{n,m}$ and $\hat{\beta}_{n,m}$, a large sample size is needed in order to attain asymptotic normality. Unfortunately, we was not able to fine a proper n such that $\tilde{\beta}^3_{n,m}$ or $\hat{\beta}^3_{n,m}$ gains normality. For $n = 1000$, the p-values of Shapiro test on both $\tilde{\beta}^i_{n,m}$ and $\hat{\beta}^i_{n,m}$ for $i = 1, 2, 4$ are above 0.6, yet 0.0004 for $\tilde{\beta}^3_{n,m}$ and $\hat{\beta}^3_{n,m}$.

In comparison between $\tilde{\beta}_{n,m}$ and $\hat{\beta}_{n,m}$ they are both very closed evidenced by a small L^1 norm for $m = 1, \ldots, 1000$, and the absolute difference of variance covariance matrices with the magnitude of eigenvalue 1e-5 when $n = 1000$, and 1e-4 when $n = 500$.

Chapter 8: Contribution and Future Work

In this section, we summarize the major contributions of this thesis, and then give an outline of what future work can be done related to the *Extended Estimating Equations* (EEE), and how to make the extension even further so that the EEE can be applied to a broader class of semiparametric models.

8.1 Original Contribution

The most important concept we developed in this thesis is the EEE. Motivated by the lack of coverage of *Classic Estimating Equation* (CEE) in the regime of right censored semiparametric models including the widely used Cox model and *Accelerated Failure Time* (AFT) model, we see the necessity to extend the scope of CEE so that it can also serve as a tool for those semiparametric setting.

Inspired by the construction of the partial likelihood equation for Cox model in [11] and the linear rank equation for the AFT in [38], we establish the EEE by allowing the criterion function not only depend on data and structural parameter, but also on the nuisance parameter. Using the concept of Glivenko-Cantelli class and Donsker class, we are able to prove lemmas parallel to the EL under EEE setting in [32]

and [33], which lead to the asymptotic normality of the corresponding Z-estimator, i.e., the root to the EEE.

We also constructed the φ-transformation model, which to our knowledge, is the broadest class of semiparametric models of which the structural estimator can be defined as the root of EEE. The φ-transformation is a generalization of the *General Transformed Model* (GMT), and can represent a series of semiparametric models including linear model, AFT, location-scale model, and operational time model, and etc. The GMT model has been well research in statistical literature such as in [38] by Tsiatis. Similar to Tsiatis's work, we construct a martingale structure on the residual scale. This structure guarantees the zero mean assumption of EEE. Then we prove the local asymptotic linearity of the associated EEE around true structural parameter, which leads to the asymptotic normality of the structural parameter.

We can also use EEE as a constraint in the *Empirical Likelihood* (EL) maximization. We prove that with criterion functions that are continuous with respect to the structural parameter, or a martingale structure as described in the φ-transformation model and Cox model, there exists a non-shrinking neighbourhood of the true structural parameter such that for any fixed value in that neighbourhood, there exists a unique probability vector that maximizes the EL with probability approaching 1. Then a structural estimator can be calculated by maximizing the *Profile Empirical Likelihood* (pEL). After establishing lemmas that are parallel to those in [32] and [33]

using empirical process theory, we show that the pEL estimator is asymptotically normal. When the dimension of the EEE equals to that of the structural parameter, the asymptotic variance covariance matrix of the pEL is identical to the sandwich-formed variance covariance matrix of the Z-estimator via solving the corresponding EEE.

From simulation with R, we are able to visually compare the local continuity of EEE and the pEL, and we found the latter is much smoother than the former. We also propose a concept of "total difference" to quantify the local continuity of the pEL. Although we are not clear about the statistical behavior of this quantity, from the simulation result, we can see that the pEL estimator preserves a similar pattern of total difference to the true parameter value. Although strictly speaking, neither EEE or pEL is continuous with respect to the structural parameter, the pEL does appear to be more smooth. One benefit of this "continuity" is that it takes much less iterations to perform the restart improvement procedure using the Nelder-Mead optimization.

We also find that for a sufficiently large sample size, the Z-estimator and the pEL estimator are very close. Namely, the L^1 distance between the two is small, and the magnitude of the eigenvalue of the difference between the variance covariance matrix small as well. The smallest sample size that is needed for the pEL to reach asymptotic normality is affected by the skewness of the covariate. The more symmetric the covariance is, the smaller the threshold sample size is.

8.2 Future Work

Under the CEE setting, one of the important benefits of EL is constructing the Wilk's type confidence region and develop the EL ratio test. Compared with the Wald type statistics, the EL confidence region does not involve calculating an variance covariance matrix, therefore is more computationally efficient. Owen developed the EL confidence region in [32] for the case when the dimension of the criterion function r equals the dimension of the structural parameter p, and Qin and Lawless generalized his conclusion to the case when $r > p$. Following the thread of Qin and Lawless's work, we were able to establish the asymptotic normality of the pEL estimator associated with EEE, however, we did not establish the Wilk's type theorem for EEE. Since we have already proved the asymptotic normality of EEE evaluated at the true structural parameter, with the same Taylor expansion technique applied to the pEL as the proof of Theorem 2 in [33], we should be able to show the EL ratio statistic follows a chi squared distribution with degree of freedom equal to the dimension of the structural parameter, and a confidence region will follow after this result.

A major advantage of using CEE as a constraint in EL method rather than directly solving for an Z-estimator is that the former allows the dimension of the CEE r exceeds that of the structural parameter p. In this thesis, we restrict our discussion to the case when $r = p$ except for the proof of the lemmas that are parallel to Lemma 1 and Theorem 1 of [33]. It is appealing to allow $r > p$ under the EEE

setting. Related additional proof could involve establishing a neighbourhood of the true structural parameter on which there exists a unique probability vector maximizing the EL and etc.

The efficiency is of the Z-estimator and the pEL estimator associated with EEE is an untouched area in this thesis. Recall that the broadest class of semiparametric models associated with the EEE is the φ-transformation model, and the φ-transformation model is generalized from the *General Transformation Model* (GMT). In [38], Tsiatis proposesd a class of estimators using linear rank tests for the GMT, and constructed the efficient estimators within this class together with conditions when they are fully efficient. In Theorem 3 of [33], Qin and Lawless pointed out the pEL with $r > p$ is efficient in the sense of [40] and [5]. As for EEE, we are interested in the following two questions. First, when $r = p$, are the Z-estimator and pEL efficient in any sense? Second, in the case when $r > p$, the asymptotic variance covariance matrices are no longer identical. Therefore it is appealing to inspect whether the two matrices can become the same after further simplification; if not, then how does the efficiency compare with one another.

Up to now, the broadest class of semiparametric models that can we construct to serve as an example of EEE is the φ-transformation model. Therefore a natural question is whether it is possible to extend EEE even further to cover more models. One such model could be the *Frailty Model* first introduced by Vaupel in univariate survival models in [43], and later applied to multivariate situation on familial ten-

dency in chronic disease incidence by Clayton in [9]. Frailty model introduces an unobserved random effect in the exponential proportionality part of the Cox model, i.e.,

$$\lambda_{T|\tilde{Z}(t),W}\left\{t|\tilde{z}(t),w\right\} = \lambda_0(t)\exp\left\{\ln W + \beta^{tr}Z(t)\right\}, \tag{8.1}$$

where $\ln W$ is an unobserved continuous random variable unique to the linear predictor of each observation that is independent of Z, $Z = \{Z(t) : t \geq 0\}$ is a $p \times 1$ covariate that may be time dependent, $\lambda_{T|\tilde{Z}(t),W}\left\{t|\tilde{z}(t),w\right\}$ is the hazard function of T conditional on W and $\tilde{Z}(t) = \{Z(s), s \leq t\}$, $\lambda_0(t)$ is an unspecified base hazard function. Let $f(w,\gamma)$ be the density of function of W with unkown parameter $\gamma \in \mathbb{R}$. Let

$$\theta = (\beta,\gamma) \in \Theta, \quad \text{where } \Theta = U_{\beta_0} \times \mathbb{R},$$

then (8.1) yields a class of semiparametric models with parameters

$$\psi = (\theta,\lambda) \in \Theta \times \mathcal{H},$$

where $\mathcal{H} \equiv \left\{\lambda(s) : \lambda(u) > 0 \text{ for all } u \in \mathbb{R}^+; \int_0^\infty \lambda(u)du = \infty\right\}$. We did some preliminary calculation on the (8.1) assuming W follows a gamma distribution, and we found that an estimator of β can be defined via maximizing the pEL, which eventually breaks down to solving the an equation that involves quotients among three higher order summations. So the question is whether we could make a general definition out of the frailty model that is an extended EEE. On the one hand, this extended EEE should cover cases including like frailty model, φ-transformation model and Cox model; on the other hand, there should be examples beyond those three semiparametric models that can be investigated using the extended EEE.

Appendix A: Empirical Process Theory

In this appendix, we first show the that under suitable conditions, if $\mathcal{F}$ is a Donsker class with finite bracketing integral, then it is also a Glivenko-Cantelli class. Then we prove that under P_θ with $\theta = (\beta, \mu)$,

$$\frac{1}{n}\sum_{j=1}^{n} Z_j I(V_j > t) e^{\beta^{tr} Z_j} \xrightarrow{P} E\left\{Z_1 I(V_1 > t) e^{\beta^{tr} Z_1}\right\}, \tag{A.1}$$

where $\{(Z_j, V_j)\}_{j=1}^{\infty}$ are *independently identically distributed* (i.i.d.) samples defined in Chapter 1, and the convergence is uniform in parameter $(\beta, t) \subset U_{\beta_0} \times \mathbb{R}^+$, where U_{β_0} is a ball in $\mathbb{R}^p$ that contains the true parameter value. Definitions of Donsker, Glivenko-Cantelli class and bracketing integral $J_{[]}(1, \mathcal{F}, L_2(P)) < \infty$ can be found on page 269-270, Chapter 19 of [41].

Theorem A.0.1 *Let F be a class of measurable functions such that the bracketing integral $J_{[]}(1, \mathcal{F}, L_2(P)) < \infty$. Then $\mathcal{F}$ is a Glivenko-Cantelli class.*

Proof: Let $N_{[]}(\varepsilon, \mathcal{F}, L_p(P))$ be the ε-bracketing number in L_p defined on page 270 of [41]. By this definition, every ε-bracket in L_2 is also a ε-bracket in L_1, we know that

$$N_{[]}(\varepsilon, \mathcal{F}, L_1(P)) \leq N_{[]}(\varepsilon, \mathcal{F}, L_2(P)).$$

By the definition of the bracketing integral on page 270, Section 19.2 in [41],

$$J_{[]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{\ln N_{[]}(\varepsilon, \mathcal{F}, L_2(P))} d\varepsilon, \tag{A.2}$$

Therefore given $J_{[]}(1, \mathcal{F}, L_2(P)) < \infty$, then $\sqrt{\ln N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}$ is finite for almost every ε. Consequently, $N_{[]}(\varepsilon, \mathcal{F}, L_2(P)) < \infty$ a.e. By Theorem 19.4 in [41], $\mathcal{F}$ is also Glivenko-Cantelli. □

The following theorem is established as part of Example 19.20 on page 277 of [41]. It shows how to construct a new Donsker class from two existing Donsker classes via Lipschitz transformation.

Theorem A.0.2 (Lipschitz Transformation) *Let $\phi : \mathbb{R}^2 \mapsto \mathbb{R}$ be a fixed Lipschitz function. If f and g range over Donsker classes $\mathcal{F}$ and $\mathcal{G}$ with integrable envelope functions, then the set of functions $\phi(f, g)$ is Donsker.*

Now we are ready show that the convergence in (A.1) is uniform in parameter $\theta = (\beta, t)$.

Theorem A.0.3 *Let $x = (z, s) \in \mathcal{Z} \times \mathbb{R}^+$, $\theta = (\beta, t) \in U_{\beta_0} \times \mathbb{R}^+$, where $\mathcal{Z}$ is a compact and bounded set in $\mathbb{R}^p$ and $U_{\beta_0} \subset \mathbb{R}^p$ is a ball centered at β_0. Then*

$$\frac{1}{n} \sum_{j=1}^{n} Z_j I(V_j > t) e^{\beta^{tr} Z_j} \xrightarrow{P} E\left\{ Z_1 I(V_1 > t) e^{\beta^{tr} Z_1} \right\}, \tag{A.3}$$

and the convergence is uniform in θ.

Proof: For fixed $\theta = (\beta, t) \in U_{\beta_0} \times \mathbb{R}^+$, the convergence of (A.3) is guaranteed by the *Law of Large Numbers* (LLN). To prove this convergence is uniform in θ, for

any fixed θ, define function $f(x,\theta)$ with argument $x = (z, v) \in \mathcal{Z} \times \mathbb{R}^+$,

$$f(\cdot,\theta) : \mathcal{Z} \times \mathbb{R}^+ \mapsto \mathbb{R} : \qquad f(x,\theta) = ze^{\beta^{tr} z} I(v \geq t). \tag{A.4}$$

If we can show that

$$\mathcal{F} = \{f(x,\theta) = ze^{\beta^{tr} z} I(s \geq t), \text{ where } \theta = (\beta, t) \in U_{\beta_0} \times \mathbb{R}^+\}$$

is Glivenko-Cantelli, then the uniformity is proved. The idea is to consider $\mathcal{F}$ as a Lipschitz transformation of a parametric class and a bounded variation class, then apply Theorem A.0.2.

For any fixed β, let $g(x,\beta) : (\mathcal{Z} \times \mathbb{R}^+) \times \mathbb{R}^+ \to \mathbb{R}$, and

$$\mathcal{G} = \{g : g(x,\beta) = ze^{\beta^{tr} z}, \text{ where } \beta \in U_{\beta_0}\},$$

Let us show that $\mathcal{G}$ is a parametric class that satisfies conditions for the parametric class described in Example 19.7 of [41], where β corresponds to θ in Example 19.7. For any β_1 and β_2 in U_{β_0},

$$|ze^{\beta_1^{tr} z} - ze^{\beta_2^{tr} z}| = |z(e^{\beta_1^{tr} z} - e^{\beta_2^{tr} z})| \leq |z||e^{\beta_1^{tr} z} - e^{\beta_2^{tr} z}| \tag{A.5}$$

Since the exponential function e^x is Lipschitz continuous when x is bounded, there exists a constant C such that

$$|e^{\beta_1^{tr} z} - e^{\beta_2^{tr} z}| \leq C|\beta_1^{tr} z - \beta_2^{tr} z| = C|(\beta_1 - \beta_2)^{tr} z| \leq |\beta_1 - \beta_2||z|$$

Plug the equation above back into (A.5), yielding

$$|ze^{\beta_1^{tr} z} - ze^{\beta_2^{tr} z}| \leq C|z|^2|\beta_1 - \beta_2|.$$

Hence we have shown that for $m(X) = C|Z|^2$. Consequently, $\mathcal{G}$ is a Donsker class by Example 19.7 of [41].

For any fixed $t \in \mathbb{R}^+$, consider function $h(\cdot, t) : \mathbb{R} \mapsto \mathbb{R}$ of the form of

$$\mathcal{H} = \{h(x,t) = I(s \geq t), \text{ where } t \in \mathbb{R}^+\}.$$

For any $h \in \mathcal{H}$, the variation is 1. Then by the bounded variation class described in Example 19.11, $\mathcal{H}$ is also a Donsker class. Next, we consider the product of functions from $\mathcal{G}$ and $\mathcal{H}$. Let $\phi(fg) : \mathbb{R}^2 \mapsto \mathbb{R}$ be

$$\phi(fg) = fg.$$

This is a Lipschitz function since z is assumed to be bounded and therefore $\mathcal{G}$ is too. Hence by Theorem A.0.2, the new class of $\phi(g,h)$, where $g \in \mathcal{G}$ and $h \in \mathcal{H}$, is also Donsker, i.e., the class of functions $f : \mathcal{Z} \times \mathbb{R}^+ \times \mathbb{R}^+ \times U_{\beta_0} \to \mathbb{R}$ of the form of

$$\mathcal{F} = \{f : f(x,\theta) = ze^{\beta^{tr} Z} I(s \geq t), \text{ where } \theta = (\beta, t) \in U_{\beta_0} \times \mathbb{R}^+\}$$

is also a Donsker class. $\square$

Bibliography

[1] P. Andersen, O. Borgan, R. Gill and N. Keiding, *Statistical Models Based on Counting Processes*, (Springer, New York), Chap. 2.

[2] P. Andersen and R. Gill, " Cox's Regression Model for Counting Processes: a Large Sample Study", Ann. Statist. **10**, (1982).

[3] V. Bagdonavičius and M. Nikulin, "Semiparametric Models in Accelerated Life Testing," Queen's Papers in Pure and Applied Mathematics. **98**, (1995).

[4] B. Bibby and M. Sørensen, "Martingale Estimation Functions for Discretely Observed Diffusion Processes", Bernoulli **1**, (1995).

[5] P. Bickel, C. Klaassen, Y. Ritov and J. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, (Johns Hopkins Univ. Press).

[6] P. Billingsley, *Probability and Mearsure*, (Wiley, New York).

[7] S. Boyd, and L. Vandenberghe, *Convex Optimization*, (Cambridge Univ. Press).

[8] J. Buckely and I. James, "Linear Regression with Censored Data", Biometrika **66**, (1976).

[9] D. Clayton, "a Model for Association in Bivariate Life Tables and Its Application in Epidemological Studies of Familial Tendency in Chronic Diease Incidence", Biometrika **65**, (1978).

[10] E. Çinlar, and S.Özekici, "Reliability of Complex Devices in Random Environments", Probability in the Engineering and Informational Sciences. **1**, (1987).

[11] D. R. Cox, " Regression Models and Life-Tables", Journal of the Royal Statistical Society. Series B (Methodological). **34**, (1972).

[12] D. R. Cox, " Partial Likelihood", Biometrika. **62**, (1975).

[13] D. R. Cox and O. Oakes, *Analysis of Survival Data*, (Chapman and Hall/CRC).

[14] B. Efron, "The Efficiency of Cox's Likelihood Function for Censored Data", JASA **72**, (1977).

[15] R. Ellis and D. Gulick, *Calculus: w/ Concepts in Calculus*, (Cengage Learning, Ohio).

[16] V. Godambe, "an Optimum Property of Regular Maximum Likelihood Estimation", Ann. Math. Statist. **31**, (1960).

[17] G. Heller, "Smoothed Rank Regression With Censored Data," JASA. **102**, (2007).

[18] T. R. Hettmansperger, *Statistical Inference Based on Ranks*, (Wiley, Newyork).

[19] N. Hjort, I. Mckeague and I. van Keilegom , "Extending the scope of empirical likelihood," Ann. Statist. **37**, (2009).

[20] S. Y. Hwang and I. V. Basawa, "A Review Toward a Unifying Tool", Conference paper seen from S. Lahiri et al, *Contemporary Developments in Statistical Theory*, pp 9 - 28.

[21] J. Jacod, "Un théorème de représentation pour les martingales discontinues," Z. Warsh. verw Gebiete, **34**, (1976).

[22] I. James and P. Smith, "Consistency Results for Linear Regression iwth Censored Data", Ann. Statist. **12**, (1984).

[23] J. Jurečková, "Asymptotic linearity of a rank statistic in regression parameter," Ann. Statist. **40**, (1969).

[24] J. Jurečková, "Nonparametric estimate of regression coefficients," Ann. Statist. **42**, (1971).

[25] J. Koul, V. Susarla and J. Van Ryzin, "Regression analysis with randomly right censored data.," Ann. Statist. **9**, (1981).

[26] T. Lai and Z. Ying, "Large Sample Theory of a Modified Buckley-James Estimator for Regression Analysis with Censored Data", Ann. Statist. **19**, (1991).

[27] E. Lopez, I. van Keilegom and N. Veraverbeke, "Empirical likelihood for nonsmooth criterion functions," Scand. J. of Stat. **36**, (2009).

[28] T. A. Louis, "Nonparametric analysis on an accelerated failure time model," Biometrika. **68**, (1981).

[29] T. Merkouris, "Transform Martingale Estimating Functions", Ann. Statist **35**, (2007).

[30] R. Miller, "Least Squares Regression with Censored Data", Biometrika **63**, (1976).

[31] A. Owen, " Empirical Likelihood Ratio Confidence Intervals for a Single Functional," Biometrika. **75**, (1988).

[32] A. Owen, *Empirical Likelihood*, (CRC Press, Raton, Florida).

[33] J. Qin and J. Lawless, "Empirical Likelihood and General Estimating Equations," Ann. Statist. **22**, (1994).

[34] Y. Ritov, "Estimation in a linear regression model with censored data," Ann. Statist. **18**, (1990).

[35] Y. Ritov and J. Wellner, "Censoring, Martingales and the Cox Model", Contemp. Math. **80**, (1988).

[36] R. Serfling, *Approximation Theorems of Mathematical Statistics* (Wiley-Interscience, New York).

[37] J. Shao, *Mathematical Statistics* (Springer, New York).

[38] A.Tsiatis, "Estimating regression parameters using linear rank tests for censored data," Ann. Statist. **18**, (1990).

[39] A. Tsiatis, *Semiparametric Theory and Missing Data* (Springer, New York)

[40] A. Van der Vaart, *Statistical Estimation in Large Parameter Spaces* (Centrum voor Wiskunde en Informatica, Amesterdam).

[41] A. van der Vaart, *Asymptotic Statistics* (Cambridge University Press, New York).

[42] A. van der Vaart and J. Wellner, *Weak Convergence and Empirical Processes* (Springer, New York).

[43] J. Vaupel, K. Manton, and E. Stallard, "the Impact of Heterogeneity in Individual Frailty on the Dynamics of mortality," Demography. **16**, (1979).

[44] L.J. Wei and M.H. Gail, "Nonparametric estimation for a scale-change with censored observations," J. Amer. Statist. Assoc. **78**, (1983).

[45] L. Wei, Z. Ying and D. Lin, "Linear Regression Analysis of Censored Survival Data Based on Rank Tests", Biometrika **77**, (1990).

[46] M. Zhou, "Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model," Biometrika. **92**, (2005).

[47] M. Zhou, "Empirical likelihood ratio with arbitrarily censored/truncated data by EM algorithm," J. Computational and Graphical Statistics. **14**, (2005).

[48] M. Zhou, *Empirical Likelihood Method in Survival Analysis* (Chapman and Hall/CRC).