Summer 8-2012

Jackknife Emperical Likelihood Method and its Applications

Hanfang Yang

JACKKNIFE EMPIRICAL LIKELIHOOD METHOD AND ITS APPLICATIONS

by

HANFANG YANG

Under the Direction of Dr. Yichuan Zhao

ABSTRACT

In this dissertation, we investigate jackknife empirical likelihood methods motivated by recent statistics research and other related fields. Computational intensity of empirical likelihood can be significantly reduced by using jackknife empirical likelihood methods without losing computational accuracy and stability. We demonstrate that proposed jackknife empirical likelihood methods are able to handle several challenging and open problems in terms of elegant asymptotic properties and accurate simulation result in finite samples. These interesting problems include ROC curves with missing data, the difference of two ROC curves in two dimensional correlated data, a novel inference for the partial AUC and the difference of two quantiles with one or two samples. In addition, empirical likelihood methodology can be successfully applied to the linear transformation model using adjusted estimation equations. The comprehensive simulation studies on coverage probabilities and average lengths for those topics demonstrate the proposed jackknife empirical likelihood methods have a good performance in finite samples under various settings. Moreover, some related and attractive real problems are studied to support our conclusions. In the end, we provide an extensive discussion about some interesting and feasible ideas based on our jackknife EL procedures for future studies.

INDEX WORDS: Empirical likelihood, Transformation model, U-statistics, Jackknife, Partial AUC, Smoothed empirical likelihood, Missing data, ROC curves, Difference of two quantiles.

JACKKNIFE EMPIRICAL LIKELIHOOD METHOD AND ITS APPLICATIONS

by

HANFANG YANG

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2012

JACKKNIFE EMPIRICAL LIKELIHOOD METHOD AND ITS APPLICATIONS

by

HANFANG YANG

Committee Chair: Dr. Yichuan Zhao

Committee: Dr. Yuanhui Xiao

Dr. Xin Qi

Dr. Yanqing Zhang

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2012

DEDICATION

For my parents, with love, and those who try to achieve the American Dream from China.

ACKNOWLEDGEMENTS

I would like to express the most sincere gratitude to my advisor, Dr. Yichuan Zhao, who first introduced me to enter the statistical academic community and taught me how to conduct a rigorous research. I will definitely miss his generous advice and provoking discussions on statistical philosophy. During the last five years, I was so lucky under his supervision with the endless patience and encouragement. It is so memorable five-years study and reserach experience without any compromise of my academic integrity. Keeping on this track, I wish to become a respectful and knowledge professor as he is.

Special thanks to other members in my dissertation committee, Dr. Yuanhui Xiao, Dr. Xin Qi and Dr. Yanqing Zhang for contributing their expertise and time to assist me to accomplish this dissertation. Also, during my study in the Georgia State university in last five years, I owe my gratitude to some current and former professors who have advised the thesis or dissertation research, Dr. Yu-sheng Hsu, Dr. Pulak Ghosh and Dr. Xu Zhang. Furthermore, I could not forget to show the deep appreciation to my classmates and their genuine friendship, including but not limited Jianhui Yu, Yinghua Lu, Ye Cui, Shan Luo, Crystal Yau, Phillip Kamau and Kokou Akolly. Our happy hours never fade away through my entire life time. I am also grateful to my pretty girl friend, Xue Qi for her selfless assistance.

In the end, I would like to acknowledge the support of my parents who have always been next to me even if Beijing is on the other side of earth.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

- ROC - Receiver operating characteristic
- EL - Empirical likelihood

CHAPTER 1

INTRODUCTION

Empirical likelihood (EL) is an nonparametric likelihood method for statistical inference, which employs the maximum likelihood method without having to assume a known distribution family for the data. Empirical likelihood enables us to successfully incorporate the advantages of the likelihood methods. First, unnecessary assumption of a family of distributions can be avoided in the empirical likelihood inference in the sense that the object is driven by data. Secondly, empirical likelihood confidence regions also can be established without manual adjustment, which shows advantage over many other nonparametric methods. Empirical likelihood can integrate external information through deterministic constraint, etc. Finally, the empirical likelihood is Bartlett correctable and can improve the accuracy of inferences.

However, the computational intensity is challenging for the application of empirical likelihood methods in practice. For some nonlinear multi-variables estimation equations, people can not neglect time spent on optimizing likelihoods function by current scientific computing technology. "Jackknife, as a kind of re-sample method, is applied with empirical likelihood and named as jackknife empirical likelihood, which surprisingly transforms nonlinear estimation equations as linear's and multi-variable optimization problem as simple-variables" (Jing et al., 2009). Hence, the jackknife EL can greatly simplify the optimization procedure, more precisely, split the entire computational problem into two segments, re-sample and optimizing.

For complete data, automatic confidence interval is determined by EL as Wilk's theorem occurs under the traditional EL procedure. We consider ROC curves with missing data. After hot deck imputation dealing with completely random missing data, scaled chi-squared distribution can be obtained from the likelihood function.

Motivated by comparison of two diagnostics tests, we attempt to investigate the difference of two ROC curves. General empirical likelihood method need involve three linking variables which dramatically influence the optimization. Hence, jackknife empirical likelihood demonstrates its advantage by the reduction about the nuisance variables. In addition, two tests are correlated in practical sense. Wilk's theorem shows jackknife empirical likelihood confidence intervals can be determined automatically for complete data even if correlation exists. The standard chi-square distribution controls the asymptomatic property of jackknife empirical likelihood. In some cases, rather than focusing on the entire ROC curve, people are interested in the ROC curve on the a special range of thresholds. Partial AUC (area under the ROC curve) is designed to answer this concern. We also proposed the jackknife empirical likelihood for the partial AUC's and the difference of two partial AUC's and show Wilk's theorem for partial AUC still holds.

Due to the less sensitivity with extremely value, quantile is recognized as a crucial robust statistics. Considering comparison of two distributions at a fixed criterion value, we recommend the difference of quantiles to explore the distance of distributions when the data has some outliers are distributed far away according to an assumed distribution. However, the structure of difference of two quantiles is complicated for us to make a reliable inference in small sample problem. Especially for some tail behavior problems, the existing method could help much. Jackknife empirical likelihood methodology is expected to be able to contribute its advantages in small sample for the difference of two quantiles and some applications, such as the difference of two quantiles and low income proportion, etc.

The transformation model is a natural generalization of proportional hazard models and proportional odds models and provides many other potential choices in survival analysis. In order to construct applicable empirical likelihood method for linear transformation models, we need to overcome the difficulty in the estimation of weights in the limiting distribution (Zhao, 2010). Involving additional compensation terms in the estimation equation, a new empirical likelihood method is appealing to avoid estimating weights. The proposed jackknife EL method is discussed in the last part.

CHAPTER 2

JACKKNIFE EMPIRICAL LIKELIHOOD FOR ROC CURVES WITH MISSING DATA

2.1 Background

2.1.1 ROC curve

The ROC curve has received considerable attention over past decades, and has been widely used in epidemiology, medical research, industrial quality control and signal detection, diagnostic medicine and material testing. In medical studies, the sensitivity or true positive rate (TPR) of the diagnostic test is the proportion of the diseased patients who have positive tests among diseased patients. The specificity or true negative rate (TNR) of the test is the proportion of the healthy people who have negative test among non-diseased people. A plot of sensitivity (TPR) against 1-specificity (FPR) defines the ROC curve, which is a graphical summary of the discriminatory accuracy of diagnostic tests. Furthermore, the ROC curve function can be represented by $ROC(p) = 1 - F(G^{-1}(1-p))$, where F and G are continuous cumulative distribution functions of positive population and negative population, respectively. Recent interesting literatures include Swets and Pickett (1982), Tosteson and Begg (1988), Hsieh and Turnbull (1996), Zou et al. (1997), Lloyd (1998), Pepe (1997), Metz et al. (1998) and Lloyd and Yong (1999), among others. Moreover, Pepe (2003) provided an excellent summary for recent research work and useful applications of ROC curves. Claeskens et al. (2003) developed smoothed empirical likelihood confidence intervals for the continuous-scale ROC curve in the absence of censoring.

2.1.2 Empirical likelihood

Empirical likelihood (EL) is a nonparametric method for statistical inference, which employs the maximum likelihood method without having to assume a known distribution family of data. Owen (1988, 1990) introduced EL method to construct confidence regions for the mean vector. Some related literatures include the Bartlett-correctability (DiCiccio and Hall, 1991), general estimating equations (Qin and Lawless, 1994), the general plug-in EL (Hjort et al., 2009) and so on. For ROC curves and copulas, a natural way is to transform nonlinear constraints to linear constraints by introducing some link variables as in Claeskens et al. (2003) and Chen et al. (2009), etc. More recently, jackknife empirical likelihood method, based on jackknife pseudo-sample, becomes more attractive. Jing et al. (2009) proposed the jackknife empirical likelihood method for a U-statistic. Gong et al. (2010) demonstrated that the smoothed jackknife empirical likelihood method for the continuous-scale ROC curve can outperform EL methods with more accurate coverage probability in a smaller sample size.

The imputation-based procedure is one of the most common methods to deal with missing data problem. In this chapter, we assume that data are missing completely at random (MCAR), which indicates "the causality of missing data is not associated with other values of observed or unobserved variables" (Little and Rubin, 2002; Qin and Qian, 2009). Using imputation method, Wang and Rao (2002) addressed missing response questions based on empirical likelihood methods. By empirical likelihood method, missing data problem was also studied by Wang and Rao (2001), Qin and Zhang (2008) and Qin and Qian (2009) among others. In this chapter, we consider hot deck imputation, which is the procedure in which missing data are randomly substituted by values from the observed sample data. In addition, An (2010) derived smoothed empirical likelihood for the ROC curve with missing data. However, the selection of bandwidth is still disputable about kernel estimators, especially with regard to missing data.

To the best of our knowledge, no paper has addressed the problem on how to construct confidence intervals for the continuous-scale ROC curve with missing completely at random data by jackknife EL methods. In this chapter, we apply smoothed jackknife EL to construct confidence intervals for the ROC curve with missing data to avoid adding extra constraints.

This chapter is organized as follows. Major procedures for the jackknife empirical likelihood ratio are proposed in Section 2.2, including methods to develop smoothed empirical likelihood and asymptotic results of jackknife empirical likelihood ratio. In Section 2.3, we conduct simulation studies to evaluate smoothed jackknife empirical likelihood confidence intervals for continuous-scale ROC curves in small and moderate samples in terms of coverage probability and average length of confidence intervals. Furthermore, we illustrate our approach using a real data example. We make a brief discussion in Section 2.4. All proofs are given in the Appendix.

2.2 Inference Procedure

2.2.1 Missing data and hot deck imputation

Consider random samples of $x_i,\ i = 1, ..., m$ in distribution F and independent missing indicators $\delta_{xi},\ i = 1, ..., m$ in Bernoulli distribution with response rate P_1, which means $P_1 = P(\delta_{xi} = 1|x_i)$. Similarly, the random samples are denoted by $y_i,\ i = 1, ..., n$ in distribution G and missing indicators $\delta_{yi},\ i = 1, ..., n$ in Bernoulli distribution with response rate P_2. Thus, we have $P_2 = P(\delta_{yi} = 1|y_i)$. Combining x_i with δ_{xi}, we can define $x_{i,m} = x_i * \delta_{xi},\ i = 1, ..., m$ as completely random missing data. Also, we have $y_{i,m} = y_i * \delta_{yi},\ i = 1, ..., n$. Denote the observed set as $X_{obs} = \{x_i : \delta_{xi} = 1\ , i = 1, ..., m\}$ and $Y_{obs} = \{y_i : \delta_{yi} = 1\ , i = 1, ..., n\}$. Then, we adopt the procedure of the hot deck imputation, replacing the missing value with values from observed set X_{obs} and Y_{obs}. Denote $r_1 = \sum_{i=1}^{m} \delta_{x_i}$, $r_2 = \sum_{j=1}^{n} \delta_{y_j}$, $m_1 = m - r_1$ and $m_2 = n - r_2$. Let $S_{rx} = \{i : \delta_{x_i} = 1\}$, $S_{mx} = \{i : \delta_{x_i} = 0\}$, $S_{ry} = \{j : \delta_{y_j} = 1\}$ and $S_{my} = \{j : \delta_{y_j} = 0\}$. x_i^* are generated by the discrete uniform distribution from observed data set X_{obs}, and y_i^* are generated by the discrete uniform distribution from observed data set Y_{obs}. Finally, we obtain the data after hot deck imputation $x_{I,i} = x_{i,m} + x_i^* * (1 - \delta_{xi}),\ i = 1, ..., m$ and $y_{I,i} = y_{i,m} + y_i^* * (1 - \delta_{yi}),\ i = 1, ..., n$.

2.2.2 Smoothed empirical likelihood ratio

Let F and G be the distribution functions of the diseased and non-diseased populations, respectively. The ROC curve can be written as $ROC(p) = 1 - F(G^{-1}(1-p))$, where $0 < p < 1$

and G^{-1} denotes the quantile function of G. Denote $F_m(x) = 1/m \sum_{i=1}^{m} I(x_{I,i} \le x)$ and $G_n(y) = 1/n \sum_{j=1}^{n} I(y_{I,j} \le y)$. Let

$$K(p) = \int_{u \le p} w(u) du,$$

where w is a the smooth symmetric kernel function with support $[-1, 1]$. Define the smooth estimator of ROC(p) as

$$\hat{R}_{m,n}(p) \quad = \quad 1 - \frac{1}{m} \sum_{j=1}^{m} K\left(\frac{1 - p - G_n(x_{I,j})}{h}\right),$$

where $h = h(n) > 0$ is a bandwidth. Define

$$\begin{aligned} \hat{R}_{m,n,i}(p) &= 1 - \frac{1}{m-1} \sum_{1 \le j \le m, j \ne i} K\left(\frac{1 - p - G_n(x_{I,j})}{h}\right), 1 \le i \le m, \\ \hat{R}_{m,n,i}(p) &= 1 - \frac{1}{m} \sum_{j=1}^{m} K\left(\frac{1 - p - G_{n,m-i}(x_{I,j})}{h}\right), m+1 \le i \le m+n, \end{aligned}$$

where

$$G_{n,-k}(y) \quad = \quad \frac{1}{n-1} \sum_{1 \le i \le n, i \ne k} I(y_{I,i} \le y), k = 1, \ldots, n.$$

The jackknife pseudo-sample is defined as

$$\hat{V}_i(p) \quad = \quad (m+n)\hat{R}_{m,n}(p) - (m+n-1)\hat{R}_{m,n,i}(p), i = 1, \ldots, m+n. \tag{2.1}$$

The empirical likelihood ratio at $\tilde{R}$, based on the $\hat{V}_i(p)$, is

$$L(\tilde{R}, p) \quad = \quad \frac{\sup\{\prod_{i=1}^{m+n}\{p_i\} : \sum_{i=1}^{m+n} p_i = 1, \sum_{i=1}^{m+n} p_i \hat{V}_i(p) = \tilde{R}, p_i > 0, i = 1, \ldots, m+n\}}{\sup\{\prod_{i=1}^{m+n}\{p_i\}, \sum_{i=1}^{m+n} p_i = 1, p_i > 0, i = 1, \ldots, m+n\}}.$$

By using the Lagrange multiplier method, we have

$$l(\tilde{R}, p) = -2\log L(\tilde{R}, p) \quad = \quad 2\sum_{i=1}^{n} \log\{1 + \lambda(\hat{V}_i(p) - \tilde{R})\}, \tag{2.2}$$

where λ satisfies the equation

$$\sum_{i=1}^{m+n} \frac{\hat{V}_i(p) - \tilde{R}}{1 + \lambda\{\hat{V}_i(p) - \tilde{R}\}} \quad = \quad 0.$$

Define

$$v_{m,n}(p) = \frac{1}{m+n} \sum_{i=1}^{m+n} \left\{ \hat{V}_i(p) - \frac{1}{m+n} \sum_{i=1}^{m+n} \hat{V}_i(p) \right\}^2 . \tag{2.3}$$

We will develop the asymptotic properties of empirical variance $v_{m,n}(p)$ and the empirical likelihood ratio statistic for the true value $R(p)$ of the ROC curve at point p based on $x_{I,i}$, $i = 1, ..., m$, $y_{I,j}$, $j = 1, ..., n$. These results are used to construct an asymptotic confidence interval for $R(p)$. We give the following regularity conditions.

A.1. $p \in (a, b)$ for any subset $(a, b) \subset (0, 1)$;

A.2. $F(x)$ and $G(y)$ are continuous functions;

A.3. $\sup|f(x)| < \infty$ and $\sup|g(y)| < \infty$, where $f(x) = dF(x)/dx$ and $g(y) = dG(y)/dy$;

A.4. Assume that $m/n \to \gamma$, as $m + n \to \infty$. $0 < P_1 < 1$ and $0 < P_2 < 1$;

A.5. $w^{'}(u)$ is bounded by $M < \infty$ for $u \in (-1, 1)$;

A.6. The distribution function $F(x) \in \mathcal{F}$, where $\mathcal{F}$ and $\mathcal{G}$ are Donsker classes, i.e., $\mathcal{F} \in CLT(P_F)$ and $\mathcal{G} \in CLT(P_G)$, where $\mathcal{F} \in CLT(P_F)$ means $\sqrt{n}(P_{F_n} - P_F)$ converges weakly to P_F-Brownian bridge B_p which has bounded uniformly continuous sample paths almost surely.

Theorem 2.1. *Under assumptions A.1-A.6, assume conditions* $h = h(n) \to 0$, $nh^2/\log n \to$

∞ and $nh^4 \to 0$ as $n \to \infty$. Then, for $p \in (a,b)$,

$$v_{m,n}(p) \xrightarrow{\mathcal{P}} \sigma_1^2(p),$$

where

$$\sigma_1^2(p) = (1 - P_1 + P_1^{-1}) \left(1 + \frac{1}{r}\right) R(p)\{1 - R(p)\} + (1 - P_2 + P_2^{-1})(1 + r)R^{'2}(p)p(1 - p).$$

Theorem 2.2. *Under the conditions of Theorem 2.1, for $p \in (a,b)$, we have*

$$l(R(p), p) \xrightarrow{\mathcal{D}} c(p)\chi_1^2,$$

where $R(p)$ is the true ROC curve at p,

$$c(p) = \frac{\sigma_1^2(p)}{\sigma_2^2(p)},$$

$$\sigma_2^2(p) = \left(1 + \frac{1}{r}\right) R(p)(1 - R(p)) + (1 + r)R^{'2}(p)p(1 - p).$$

Remark: In our setting, the limiting distribution is the scaled chi-squared distribution because of missing mechanism. When the response rate $P_1 = P_2 = 1$, the limiting distribution is a standard chi-squared distribution.

We may use a consistent estimator $\hat{c}$ of $c(p)$ to construct our confidence intervals of $R(p)$. Thus, the asymptotic $100(1-\alpha)\%$ smoothed jackknife EL confidence interval for $R(p)$ is given by

$$I(p) = \left\{\tilde{R} : l(\tilde{R}, p) \leq \hat{c}\chi_1^2(\alpha)\right\},$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2.

2.3 Numerical Studies

In this section, we conduct simulation studies to compare the performance of jackknife empirical likelihood (JEL) method and smoothed empirical likelihood (SEL) proposed by An (2010) for the ROC curve in terms of coverage accuracy and average length of confidence intervals with various distributions, response rates and sample sizes. In the simulation studies, distributions of the diseased population (X) and the non-diseased population (Y) are represented by $F(x)$ and $G(y)$. We consider three scenarios, which are $(\mathbf{A}) F \sim N(0.2, 0.5), G \sim N(0, 0.5)$, $(\mathbf{B}) F \sim Exp(1), G \sim N(0, 0.5)$ and $(\mathbf{C}) F \sim Exp(1), G \sim Exp(1)$. Random samples x and y are independently drawn from populations X and Y. The response rates for data x and y are chosen as, $(P_1, P_2) = (0.7, 0.6)$ or $(0.9, 0.8)$. The sample sizes for x and y are $(m, n) = (50, 50)$, $(100, 100)$ and $(200, 150)$. For certain response rate and sample size, we generate 1000 independent random samples of missing data. Without the loss of generality, we use both methods to construct confidence intervals for ROC curves at p= $0.2, 0.3$. The nominal level of the confidence intervals is $1 - \alpha = 0.95$. Then, the Epanechnikov kernel

$$w(u) = \begin{cases} \frac{3}{4}(1 - u^2) & \text{if } |u| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

is used for both JEL method and SEL method (see An, 2010) and the smoothing parameter is chosen to be $h = n^{-1/3}$ for JEL method and $h_1 = m^{-1/3}$ and $h_2 = n^{-1/3}$ for SEL method. The simulation results of coverage probability are illustrated in Table 2.1. From Table 2.1, we find out that JEL method has much better performance than SEL method in the most simulation settings.

Next, we investigate the performance of average length of ROC curves using JEL method and SEL method. We arrange the same simulation settings as before. To obtain the average length, we applied the bisection method to find solutions. The process does not involve high computation costs because jackknife method can simplify the complexity of equations significantly. This is one of main advantages of the smoothed jackknife EL method. Table 2.2

Table 2.1 Coverage probability of 95% confidence intervals for $ROC(p)$.

n1	n2	p	P1	P2	JEL (A)	SEL (A)	JEL (B)	SEL (B)	JEL (C)	SEL (C)
50	50	0.1	0.7	0.6	0.918	0.942	0.951	0.787	0.932	0.773
100	100	0.1	0.7	0.6	0.923	0.940	0.953	0.813	0.917	0.789
200	150	0.1	0.7	0.6	0.937	0.952	0.946	0.804	0.927	0.814
50	50	0.3	0.7	0.6	0.928	0.950	0.968	0.812	0.946	0.856
100	100	0.3	0.7	0.6	0.928	0.948	0.955	0.848	0.925	0.903
200	150	0.3	0.7	0.6	0.942	0.953	0.955	0.853	0.932	0.908
50	50	0.1	0.9	0.8	0.908	0.960	0.951	0.822	0.919	0.821
100	100	0.1	0.9	0.8	0.923	0.953	0.951	0.859	0.929	0.847
200	150	0.1	0.9	0.8	0.933	0.962	0.958	0.869	0.938	0.877
50	50	0.3	0.9	0.8	0.926	0.962	0.941	0.832	0.930	0.894
100	100	0.3	0.9	0.8	0.929	0.960	0.955	0.865	0.940	0.918
200	150	0.3	0.9	0.8	0.932	0.941	0.947	0.905	0.933	0.928

shows comparable results between JEL method and SEL method based on average length. In summary, the JEL method has better coverage probability and similar average length in small samples, compared with the traditional SEL method.

In addition, we study empirical likelihood confidence intervals for ROC curves at different specificities generated by simulated data. Data A is employed in this case. We choose two sample sizes $(50, 50)$ and $(300, 300)$ with different response rates $(1, 1)$ and $(0.8, 0.9)$. We select 100 points on the ROC curve evenly to obtain JEL confidence intervals respectively. As Figure 2.1 shows, it is clear that the confidence intervals of ROC curves are located above diagonal line, which indicates two distributions can be distinguished by ROC curves. In addition, the ROC curve has the shorter confidence intervals when the sample sizes are larger.

2.4 Real Application

Moreover, using the real example, we illustrate our proposed method. Data set can be accessed publicly from the website of the Center for Machine Learning and Intelligent Systems at University of California, Irvine and are originated from Hewlett-Packard Labs. It contains 4601 observations with 57 attributes and one indicator variable for spam e-mails,

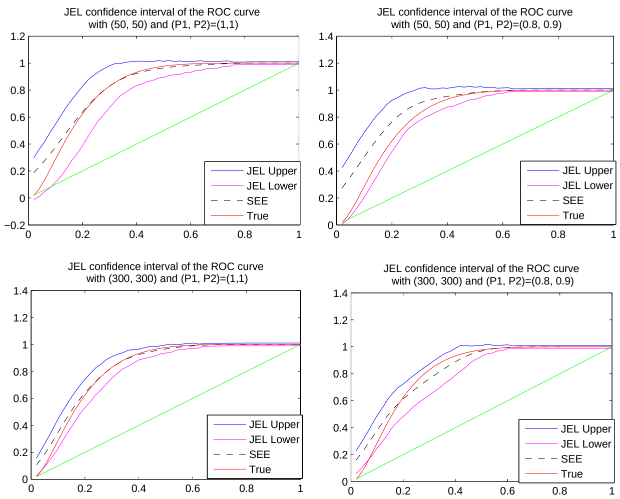

Figure 2.1 95% point-wise jackknife empirical likelihood confidence interval for ROC curves, where JEL Upper indicates the upper bound of jackknife empirical likelihood confidence interval, JEL Lower indicates the lower bound of jackknife empirical likelihood confidence interval, SEE means smoothed empirical estimator and True means the true value of ROC curve.

which are considered as the advertisements for products or web sites, make money fast schemes and pornography. Most of those attributes are valued by percentages of certain words appearing in the e-mail. In this chapter, we split the 24th attribute into two groups based on the spam indicator variable in order to construct the ROC curve with missing completely at random (MCAR) at 20% missing rate. Figure 2.2 shows the confidence interval for the ROC curve. The confidence intervals of ROC curves are above the diagonal line. Thus, the spam observation in 24th attribute can be clearly distinguished from the non-spam observation.

2.5 Discussion

In this chapter, we apply jackknife empirical likelihood method to construct confidence intervals for the continuous-scale ROC curve with missing data. The theoretical results provide asymptotic properties, including asymptotic variance and limiting distribution of the empirical likelihood ratio statistics. The simulation results demonstrate that coverage

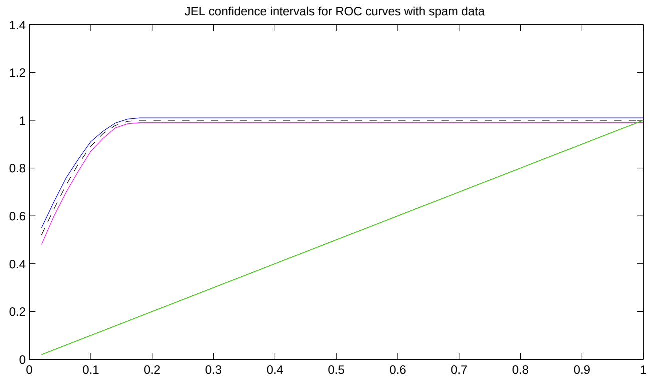

Figure 2.2 95% point-wise jackknife empirical likelihood confidence interval for ROC curves from spam data A, where JEL Upper indicates the upper bound of jackknife empirical likelihood confidence interval, JEL Lower indicates the lower bound of jackknife empirical likelihood confidence interval, SEE means smoothed empirical estimator and True means the true value of ROC curve.

probability of EL confidence interval can be close to nominal level at various high response rates and in the different locations of the ROC curve. Comparing with traditional SEL methods, JEL methods have less computational cost and a more precise coverage probability and similar average length.

There are other topics, which should be studied in the future. For instance, combining jackknife empirical likelihood method, imputation methods could be applied to solve other missing data problems. Moreover, we may consider to develop smoothed jackknife empirical likelihood method for ROC curves with other kinds of incomplete data, such as right censoring data and current status data.

Table 2.2 Average length of 95% confidence intervals for $ROC(p)$.

m	n	p	P1	P2	JEL (A)	SEL (A)	JEL (B)	SEL (B)	JEL (C)	SEL (C)
50	50	0.1	0.7	0.6	0.42492	0.40518	0.32284	0.21781	0.35455	0.22543
100	100	0.1	0.7	0.6	0.31733	0.31807	0.23647	0.18237	0.26238	0.19084
200	150	0.1	0.7	0.6	0.25668	0.27594	0.19612	0.15152	0.20800	0.16759
50	50	0.3	0.7	0.6	0.45558	0.44693	0.35349	0.24882	0.42148	0.33445
100	100	0.3	0.7	0.6	0.33850	0.36403	0.25595	0.21040	0.30840	0.28424
200	150	0.3	0.7	0.6	0.27029	0.31737	0.21029	0.17784	0.24542	0.24410
50	50	0.1	0.9	0.8	0.35404	0.36128	0.27093	0.20495	0.29431	0.22238
100	100	0.1	0.9	0.8	0.26807	0.28567	0.20001	0.17431	0.21982	0.18420
200	150	0.1	0.9	0.8	0.21693	0.25042	0.16592	0.15093	0.17548	0.16502
50	50	0.3	0.9	0.8	0.38486	0.41028	0.29694	0.22772	0.35025	0.31367
100	100	0.3	0.9	0.8	0.28426	0.33791	0.21581	0.19633	0.25848	0.26212
200	150	0.3	0.9	0.8	0.22810	0.28021	0.17743	0.17743	0.20477	0.22474

CHAPTER 3

SMOOTHED JACKKNIFE EMPIRICAL LIKELIHOOD INFERENCE FOR THE DIFFERENCE OF ROC CURVES

3.1 Background

3.1.1 The difference of two ROC curves

The ROC curve is a popular technique used to measure the performance of a classification, which is broadly applied in medical studies, machine learning and decision making, etc. It enables us to comprehensively visualize the discrimination ability of a decision rule at various thresholds. For instance, in a medical study, it can evaluate how well a diagnostic test distinguishes diseased people from non-diseased people. Many researchers have made great contributions to ROC curves studies, such as Metz et al.(1978), Tosteson and Begg (1988), Hsieh and Turnbull (1996), Pepe (1997), Lloyd (1998), Lloyd and Yong (1999) and Claeskens et al. (2003).

In practice, people have a great opportunity to encounter with bivariate correlated data (x_1, x_2) from diseased group and correlated data (y_1, y_2) from non-diseased group. A criterion is appealing to choose a better diagnostic test which is based on the data x_1 and y_1 or the alternative test from the data x_2 and y_2 with respect to the discriminant ability. In order to select a more powerful diagnostic test in the sense of the ROC curve, people consider to study the difference of two correlated ROC curves. Hanley et al. (1983) established the parametric model for the difference of two ROC curves. Delong et al. (1988) proposed a nonparametric approach for the difference of two correlated ROC curves. Moreover, the comparison of two diagnostic tests was studied in the following papers, such as Linnet (1987), Wieand et al. (1989) and Venkatraman and Begg (1996).

3.1.2 Empirical likelihood

Empirical likelihood was first introduced by Owen (1988, 1990). Later, other researchers expanded empirical likelihood methodology to many statistical fields, including some papers closely related to our topics, such as general estimating equations (Qin and Lawless, 1994) and ROC curves (Claeskens et al., 2003). However, due to involving nonlinear systems with many nuisance variables in some applications, such as the ROC curve and copulas, the application of empirical likelihood method is hindered by intensive computational burdens. Recently, jackknife empirical likelihood method has received more attention because it improves the computational efficiency successfully by reducing nuisance parameters. Jing et al. (2009) proposed jackknife empirical likelihood method for U-statistic. Gong et al. (2010) applied smoothed jackknife empirical likelihood method for ROC curves, and Peng and Qi (2010) developed tail copulas by jackknife empirical likelihood method.

In this chapter, for the p, we make an inference for two correlated continuous-scale ROC curves $\Delta(p) = ROC_1(p) - ROC_2(p)$ by smoothed jackknife empirical likelihood method.

The rest of this chapter is organized as follows. In Section 3.2, we prove that the smoothed jackknife empirical log-likelihood ratio for the difference of two correlated ROC curves converges to a chi-squared distribution. Furthermore, the simulation studies in terms of coverage probability and average length of confidence intervals are carried out in Section 3.3. We make a discussion about the future work in Section 3.4. The proofs are given in the Appendix.

3.2 Inference Procedure

3.2.1 Preliminaries

Let $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ be two-dimensional random variables. X and Y are independent. $F(x_1, x_2)$ and $G(y_1, y_2)$ are corresponding continuous bivariate distribution functions of the diseased and non-diseased populations, respectively. Denote marginal distributions $F_1(x_1)$, $F_2(x_2)$, $G_1(y_1)$ and $G_2(y_2)$. Consider a diagnostic test on X_1 and Y_1.

The continuous-scale ROC curve can be written as $R_1(p) = 1 - F_1(G_1^{-1}(1-p))$, where $0 < p < 1$, and G_1^{-1} denotes a quantile function of Y_1. Similarly, we can define the continuous-scale ROC curve, $R_2(p) = 1 - F_2(G_2^{-1}(1-p))$, where G_2^{-1} is a quantile function of Y_2. Thus, the difference of two correlated ROC curves at a fixed specificity p can be written as $\Delta(p) = F_2(G_2^{-1}(1-p)) - F_1(G_1^{-1}(1-p))$.

Consider two dimensional data $(X_{1i}, X_{2i}), i = 1, ..., m$, associated with diseased population and $(Y_{1j}, Y_{2j}), j = 1, ..., n$, associated with non-diseased population, where (X_{1i}, X_{2i}), $i = 1, ..., m$ are i.i.d., and $(Y_{1j}, Y_{2j}), j = 1, ..., n$ are i.i.d. Denote empirical estimators of bivariate distribution functions as $F_m(x_1, x_2) = 1/m \sum_{j=1}^{m} I(X_{1,j} \leq x_1, X_{2,j} \leq x_2)$ and the $G_n(y_1, y_2) = 1/n \sum_{j=1}^{n} I(Y_{1,j} \leq y_1, Y_{2,j} \leq y_2)$. The empirical estimators of marginal distributions are $F_{m,1}(x_1) = 1/m \sum_{i=1}^{m} I(X_{1,i} \leq x_1)$, $F_{m,2}(x_2) = 1/m \sum_{i=1}^{m} I(X_{2,i} \leq x_2)$, $G_{n,1}(y_1) = 1/n \sum_{i=1}^{n} I(Y_{1,j} \leq y_1)$ and $G_{n,2}(y_2) = 1/n \sum_{i=1}^{n} I(Y_{2,j} \leq y_2)$.

3.2.2 Methodology

Let $K(p)$ be the smooth distribution function which is

$$K(p) = \int_{u \leq p} w(u) \mathrm{d}u,$$

where $w(u)$ is a symmetric density function with support $[-1, 1]$. Because of Remark 1 of Gong et al. (2010), we also consider the smooth estimators of ROC curves and the difference of two continuous-scale ROC curves as,

$$\begin{aligned}
\hat{R}_{m,n,1}(p) &= 1 - \frac{1}{m} \sum_{j=1}^{m} K \left\{ \frac{1 - p - G_{n,1}(X_{1,j})}{h} \right\}, \\
\hat{R}_{m,n,2}(p) &= 1 - \frac{1}{m} \sum_{j=1}^{m} K \left\{ \frac{1 - p - G_{n,2}(X_{2,j})}{h} \right\}, \\
\hat{\Delta}_{m,n}(p) &= \hat{R}_{m,n,1}(p) - \hat{R}_{m,n,2}(p),
\end{aligned}$$

where $h = h(n) > 0$ is a bandwidth. Then, we introduce the procedure to generate jackknife pesudo-sample. We denote

$$\hat{\Delta}_{m,n,i}(p) = \tfrac{1}{m-1} \sum_{1 \le j \le m, j \ne i} \left\{ K\left(\tfrac{1-p-G_{n,2}(X_{2,j})}{h}\right) - K\left(\tfrac{1-p-G_{n,1}(X_{1,j})}{h}\right) \right\}, 1 \le i \le m,$$
$$\hat{\Delta}_{m,n,i}(p) = \tfrac{1}{m-1} \sum_{j=1}^{m} \left\{ K\left(\tfrac{1-p-G_{n,m-i,2}(X_{2,j})}{h}\right) - K\left(\tfrac{1-p-G_{n,m-i,1}(X_{1,j})}{h}\right) \right\}, m+1 \le i \le m+n,$$

where

$$G_{n,-i,k}(y) \quad = \quad \frac{1}{n} \sum_{1 \le j \le n, j \ne i} I(Y_{k,j} \le y), i = 1, \ldots, n, k = 1, 2.$$

The jackknife pseudo-sample is defined as

$$\hat{V}_i(p) \quad = \quad (m+n)\hat{\Delta}_{m,n}(p) - (m+n-1)\hat{\Delta}_{m,n,i}(p), i = 1, \ldots, m+n. \tag{3.1}$$

Then, the empirical likelihood log-ratio at p and general value $\tilde{\Delta}$ based on the pesudo-sample $\hat{V}_i(p)$ is

$$L(\tilde{\Delta}, p) \quad = \quad \frac{\sup\{\prod_{i=1}^{m+n} p_i : \sum_{i=1}^{m+n} p_i = 1, \sum_{i=1}^{m+n} p_i \hat{V}_i(p) = \tilde{\Delta}, p_i > 0, i = 1, \ldots, m+n\}}{\sup\{\prod_{i=1}^{m+n} p_i, \sum_{i=1}^{m+n} p_i = 1, p_i > 0, i = 1, \ldots, m+n\}}.$$

Using the Lagrange method, we have

$$l(\tilde{\Delta}, p) = -2 \log L(\tilde{\Delta}, p) \quad = \quad 2 \sum_{i=1}^{n} \log\{1 + \lambda(\hat{V}_i(p) - \tilde{\Delta})\}, \tag{3.2}$$

where Lagrange multiplier λ satisfies the equation

$$\sum_{i=1}^{m+n} \frac{\hat{V}_i(p) - \tilde{\Delta}}{1 + \lambda(\hat{V}_i(p) - \tilde{\Delta})} \quad = \quad 0.$$

Define the pseudo-sample variance

$$v_{m,n}(p) = \frac{1}{m+n} \sum_{i=1}^{m+n} \left\{ \hat{V}_i(p) - \frac{1}{m+n} \sum_{i=1}^{m+n} \hat{V}_i(p) \right\}^2 . \tag{3.3}$$

In order to establish the main theorem, we assume the following regularity conditions similar to Gong et al. (2010):

A.1. $F_1(x_1)$, $F_2(x_2)$, $G_1(y_1)$, $G_2(y_2)$, $F(x_1, x_2)$ and $G(y_1, y_2)$ are continuous functions and have continuous, bounded first derivatives;

A.2. ROC curves $R_1(p)$ and $R_2(p)$, and their first derivatives $R_1^{'}(p)$ and $R_2^{'}(p)$, are bounded and continuous in $p \in (0, 1)$;

A.3. $w(u)$ is a symmetric density function with support $[-1, 1]$ and $w^{'}(u)$ is bounded, continuous for $u \in [-1, 1]$;

A.4. $h = h(n) \to 0$, $nh^2/\log n \to \infty$, $nh^4 \to 0$ as $n \to \infty$;

A.5. $p \in (a, b)$ for any subset $(a, b) \subset (0, 1)$;

A.6. $m/n \to r$, where $r > 0$.

Remark: A.1-A.2 provide mathematical descriptions for the continuity of ROC curves, which allow us to accomplish the proof of Theorem 2.1 and Theorem 2.2. Hence, in real application, the continuous-scaled diagnostic tests are appropriate for our results. A.3-A.4 specify the regular properties of the kernel function and its bandwidth. A.5 makes us avoid discussing the boundary issue which is negligible in practice. A.6 guarantees the two sample sizes are comparable.

Theorem 3.1. *Under assumptions* $A.1 - A.6$, *for* $p \in (a, b)$, *the pseudo sample variance has the asymptotic property*

$$v_{m,n}(p) \xrightarrow{\mathcal{P}} \sigma^2(p),$$

where

$$\sigma^2(p) = \sigma_1^2(p) + \sigma_2^2(p) + 2\sigma_{12}^2(p),$$

$$\sigma_i^2(p) = \frac{1+r}{r} R_i(p)\{1 - R_i(p)\} + (1+r)(1-p)p\{R_i^{'}(p)\}^2, i = 1, 2,$$

$$\sigma_{12}^2(p) = \frac{1+r}{r}[F\{G_1^{-1}(p), G_2^{-1}(p)\} - \{1 - R_1(p)\}\{1 - R_2(p)\}]$$

$$+(1+r)[G\{G_1^{-1}(p), G_2^{-1}(p)\} - p^2]R_1^{'}(p)R_2^{'}(p).$$

Thus, the asymptotic $100(1-\alpha)\%$ smoothed jackknife EL confidence interval for $\Delta(p)$ is given by

$$I(p) = \left\{\tilde{\Delta} : l(\tilde{\Delta}, p) \leq \chi_1^2(\alpha)\right\},$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2.

3.3 Numerical Studies

In order to examine the finite sample performance of Theorem 2.2, we conduct extensive simulation studies in terms of coverage probability and average lengths of confidence intervals under various data settings. For data set A, $F(x_1, x_2)$ is generated from a multi-normal distribution with mean $(1, 2)$ and covariance matrix $\begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$, and $G(y_1, y_2)$ is a multi-normal distribution with mean $(0, 1)$ and covariance matrix $\begin{pmatrix} 2 & -0.8 \\ -0.8 & 2 \end{pmatrix}$. For data set B, we select a multi-normal distribution with mean $(0, 1)$ and covariance matrix $\begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$ as $F(x_1, x_2)$ and a multi-normal distribution with mean $(0, 1)$ and covariance matrix $\begin{pmatrix} 2 & -0.8 \\ -0.8 & 2 \end{pmatrix}$ as $G(y_1, y_2)$. For data set C, we select a log-normal distribution as $F(x_1, x_2)$, which is created by a normal distribution with mean $(1, 2)$ and covariance matrix $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, and we choose a log-normal distribution $G(y_1, y_2)$ transformed from a normal distribution with mean $(0, 1)$ and covariance matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. For data set D, we select bi-variate exponential distribution as $F(x_1, x_2)$ which has two independent marginal standard exponential distribution, and choose a Gumbels Type I bivariate exponential distribution as $G(y_1, y_2)$ introduced by Gumbel (1960), i.e.,

$$G(y_1, y_2) = 1 - e^{-y_1} - e^{-y_2} + e^{-(y_1+y_2+\phi y_1 y_2)}, \quad 0 \le \phi \le 1,$$

where ϕ is the parameter which relates the correlation coefficient ρ. In fact, if $\phi = 1$, the correlation coefficient $\rho = -0.404$. To generate random data from Gumbel Type I bivariate distribution, we employ a density mixture method. See details in Balakrishnan and Lai (2009).

In our simulation studies, we focus on two points lying on ROC curves, $p = 0.4, 0.6$. Moreover, three pairs of sample sizes are chosen, i.e., (50, 50), (100, 100) and (200,150). We clarify the kernel functions, which is one crucial factor in our JEL procedure. We use the Epanechnikov kernel function

$$w(u) = \begin{cases} \frac{3}{4}(1-u^2) & \text{if } |u| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

for JEL in the simulation study. The bandwidth is defined as $h = n^{-1/3}$, which is determined automatically from the sample size. Furthermore, in order to find average lengths of confidence intervals, the bisection method with small jump 0.001 is applied for seeking the upper bound and lower bound. The nominal level is fixed at 95% and data sets are simulated with 1000 repetitions.

In Table 3.1, we report coverage probabilities. We can observe that all results are close to nominal level 95%. The simulation results of average lengths are illustrated in Table 3.2. It is clear that average length becomes shorter as the sample size becomes larger.

In addition, we plot the jackknife empirical likelihood confidence interval for the difference of two correlated ROC curves. Data set E is simulated from a multi-normal distribution with mean $(2, 1)$ and covariance matrix $\begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$, and $G(y_1, y_2)$ is a multi-normal distribution with mean $(0,1)$ and covariance matrix $\begin{pmatrix} 2 & -0.8 \\ -0.8 & 2 \end{pmatrix}$. Data set B and Data set E are employed to demonstrate its performance at various specificities. Choosing two sample sizes $(50,50)$ and $(200,200)$, we plot the difference of two correlated ROC curves as Figure 3.1 including empirical estimators and smoothed JEL confidence intervals at 100 points, respectively. The two true ROC curves from the distributions about data B are identically same. From Figure 1, we find the 95% jackknife EL confidence intervals of Data set B include the x axis at most levels p. Hence, it is not significant that one ROC curve is different from the other. For data set E, we can clearly distinguish the test with a dominant discrimination ability from another one.

3.4 Real Application

Furthermore, we utilized the Pancreatic Cancer Serum Biomarkers data (see Wieand et al. 1989) to illustrate the proposed JEL method. Mayo Clinic's case cohort study collected those dataset originally and the data were investigated by Wieand et al. (1989) using non-parametric and semi-parametric methods. The dataset including two biomarkers, CA-125 (V1), a cancer antigen, and CA-19-9 (V2), a carbohydrate antigen, were split according to indicators, which distinguish pancreatic cancer (90 patients) and pancreatitis (51 patients), respectively. To compare the efficiency of diagnostics evaluations based on each biomarker, we calculate the estimate of the difference of two ROC curves. Figure 2 shows the jackknife

Table 3.1 Coverage probability of 95% confidence interval for the difference of two ROC curves $\Delta(p)$.

m	n	p	(A)	(B)	(C)	(D)
50	50	0.4	0.945	0.930	0.935	0.941
100	100	0.4	0.941	0.940	0.947	0.950
200	150	0.4	0.927	0.955	0.943	0.946
50	50	0.6	0.955	0.936	0.948	0.953
100	100	0.6	0.955	0.920	0.949	0.945
200	150	0.6	0.942	0.935	0.943	0.937

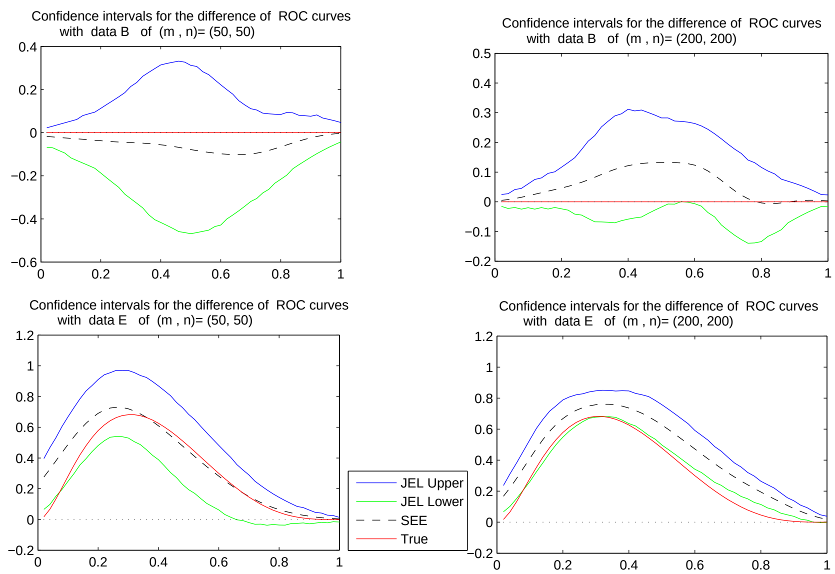

Figure 3.1 95% point-wise jackknife empirical likelihood confidence interval for the difference of two ROC curves from data B, where JEL Upper indicates the upper bound of jackknife empirical likelihood confidence interval, JEL Lower indicates the lower bound of jackknife empirical likelihood confidence interval and SEE means smoothed empirical estimator.

empirical likelihood confidence interval and smoothed empirical estimators for the difference of two ROC curves. It is clear that the 95% confidence interval is located above than 0 at 1-specificities from 0 to 0.7. Hence, the first biomarker, CA-125 (V1), a cancer antigen, has better capability to distinguish pancreatic cancer and pancreatitis.

3.5 Discussion

In this chapter, we study the difference of two correlated ROC curves based on smoothed jackknife empirical likelihood method. Jackknife pseudo-sample makes the computation less intensive because it can avoid solving nonlinear system with link variables as the standard empirical likelihood does. From the simulation studies, we demonstrate that smoothed jackknife empirical likelihood method works very well in finite sample sizes, and the coverage probabilities are close to the nominal level. The key contribution of this chapter is that we extend the application of jackknife empirical likelihood to two dimensional correlated data and save the computational intensity. In the future, we will report our result for the difference of two ROC curves, Δ_p with incomplete data, such as missing at random.

Table 3.2 Average length of 95% confidence interval for the difference of two ROC curves $\Delta(p)$.

m	n	p	(A)	(B)	(C)	(D)
50	50	0.4	0.5193	0.5735	0.3668	0.4683
100	100	0.4	0.3853	0.4260	0.2706	0.3430
200	150	0.4	0.3173	0.3533	0.2132	0.2505
50	50	0.6	0.3174	0.5720	0.2591	0.5460
100	100	0.6	0.2237	0.4243	0.1877	0.4022
200	150	0.6	0.1772	0.3526	0.1476	0.2947

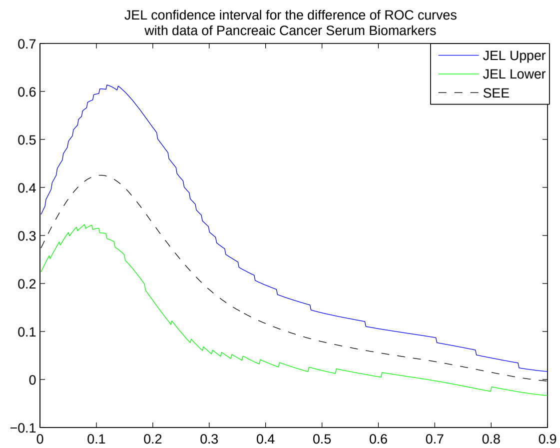

Figure 3.2 95% point-wise jackknife empirical likelihood confidence interval for the difference of two ROC curves from Pancreatic Cancer Serum Biomarkers, where JEL Upper indicates the upper bound of jackknife empirical likelihood confidence interval, JEL Lower indicates the lower bound of jackknife empirical likelihood confidence interval and SEE means smoothed empirical estimator.

CHAPTER 4

JACKKNIFE EMPIRICAL LIKELIHOOD INFERENCE FOR THE PARTIAL AUC

4.1 Background

For comparing two diagnostic tests, a ROC curve is a fundamental tool which has been extensively studied. However, even if ROC curve provides a graphical measurement, it can not be easily used to make an indisputable judgement which curve indicates a better test generally. For instance, two ROC curves maybe meet together at some thresholds points and no one curve dominates the other absolutely. As a summary of the whole ROC curve, AUC (area under ROC curve) is regarded as an integration of sensitivity over specificity. In most cases, rather than focusing on the entire ROC curve, People's interest can be based on the a special threshold range, partial AUC. By aggregating ROC curve on the partial range of threshold, the partial AUC takes the advantages of both ROC curve and AUC. The underlying range of partial AUC is able to be adjusted by people with their own interest. As an index, the partial AUC is convenient for people to make an evaluation, comparison and inference, etc.

Nonparametric method about partial AUC has been developed by Hsieh and Turnbull (1996). Under empirical likelihood framework, Qin and Zhou (2006) proposed the inference procedure about the AUC. Further, after Jing et al. (2009) introduce the jackknife empirical likelihood, Adimari and Chiogna (2011) applied this methodology to the partial AUC using standard nonparametric estimation. However, this estimation equation involves many arguable and disputable discussions, such as Remark mentioned by Gong et al. (2010). On the other hand, Wang and Chang (2011) seminally developed a novel estimation for the partial AUC based on the integration and smoothing technique. In this chapter, we construct the jacknife empirical likelihood method for partial AUC with the smoothed estimation introduced by Wang and Chang (2011).

The rest of the chapter is organized as follows. In Section 4.2, we provide the outline of inference procedure for partial AUC using jackknife empirical likelihood method. In Section 4.3, we conduct simulation studies in terms of coverage probability and average length of confidence intervals. Furthermore, we make a discussion about the difference of two pAUC's and our future work in Section 4.4.

4.2 Inference Procedure

4.2.1 Preliminaries

First, we clarify settings in this chapter. Let X and Y be two independent random variables with distribution functions $F(x)$ and $G(x)$. We define the partial AUC from 0 to p, $pAUC(p) = \int_0^p ROC(t)dt$, where $ROC(t) = 1 - F(G^{-1}(1-t))$. We can simply obtain the partial AUC with two boundary points from p_1 to p_2 as $pAUC(p_2) - pAUC(p_1)$, if $0 < p_1 < p_2 < 1$.

Let $\mathbf{x} = \{x_i, i = 1, ..., m\}$ and $\mathbf{y} = \{y_i, i = 1, ..., m\}$ be random samples from the distribution function $F(x)$ and $G(y)$, respectively. A straightforward discrete estimator of partial AUC is provided by Wang and Chang (2011) as follows,

$$\hat{A}(p) = \frac{1}{n}\sum_{i=1}^{n}\left[p - \min\left\{\frac{1}{m}\sum_{j=1}^{m} I(X_j > Y_i), p\right\}\right].$$

A consistent smoothed estimator of partial AUC is proposed by Wang and Chang (2011) as follows,

$$\tilde{A}(p) = \frac{1}{n}\sum_{i=1}^{n}\left[p - h\log\frac{1+\exp(p/h)}{1+\exp\{[p - \frac{1}{m}\sum_{j=1}^{m} K\{(X_j - Y_i)\}/h]/h\}}\right],$$

where $K(t) = 1/[1 + exp(-t)]$.

4.2.2 Methodology

Starting from the discrete estimator, we define other estimators from partial samples,

$$\hat{A}_{-k}(p) = \frac{1}{n}\sum_{i=1}^{n}\left[p - \min\left\{\frac{1}{m}\sum_{j=1,j\neq k}^{m} I(X_j > Y_i), p\right\}\right], 1 \le k \le m,$$

$$\hat{A}_{-k}(p) = \frac{1}{n}\sum_{i=1,i\neq k-m}^{n}\left[p - \min\left\{\frac{1}{m}\sum_{j=1}^{m} I(X_j > Y_i), p\right\}\right], m+1 \le k \le m+n.$$

The corresponding jackknife pseudo-samples are obtain by

$$\hat{Q}_k(p) = (m+n)\hat{A}(p) - (m+n-1)\hat{A}_{-k}(p), 1 \le k \le m+n.$$

Similarly, based on the smoothed estimator, we construct the jackknife pseudo-sample as follows.

$$\tilde{A}_{-k}(p) = \frac{1}{n}\sum_{i=1}^{n}\left[p - h\log\frac{1+\exp(p/h)}{1+\exp\{[p - \frac{1}{m-1}\sum_{j=1,j\neq k}^{m} K\{(X_j - Y_i)\}/h]/h\}}\right], 1 \le k \le m.$$

$$\tilde{A}_{-k}(p) = \frac{1}{n-1}\sum_{i=1,i\neq k-m}^{n-1}\left[p - h\log\frac{1+\exp(p/h)}{1+\exp\{[p - \frac{1}{m}\sum_{j=1}^{m} K\{(X_j - Y_i)\}/h]/h\}}\right], m+1 \le k \le m+n.$$

Then, we have

$$\tilde{Q}_k(p) = (m+n)\tilde{A}(p) - (m+n-1)\tilde{A}_{-k}(p), 1 \le k \le m+n.$$

We treat $\tilde{Q}_k(p), k = 1, ..., m+n$ as pseudo-sample of partial AUC from 0 to p.

Based on jackknife pesudo-sample, the empirical likelihood log-ratio at p, $\hat{L}(p, pAUC(p))$, is

$$\frac{\sup\{\prod_{i=1}^{m+n} p_i : \sum_{i=1}^{m+n} p_i = 1, \sum_{i=1}^{m+n} p_i\hat{Q}_i(p) = pAUC(p), p_i > 0, i = 1, \ldots, m+n\}}{\sup\{\prod_{i=1}^{m+n} p_i, \sum_{i=1}^{m+n} p_i = 1, p_i > 0, i = 1, \ldots, m+n\}}.$$

Using standard Lagrange multiplier method, we obtain a log-empirical likelihood ratio routinely,

$$\hat{l}(p, pAUC(p)) = -2\log \hat{L}(p, pAUC(p)) \;=\; 2\sum_{i=1}^{n} \log\{1 + \lambda(\hat{Q}_i(p) - pAUC(p))\}, \quad (4.1)$$

where Lagrange multiplier λ satisfies the equation

$$\sum_{i=1}^{m+n} \frac{(\hat{Q}_i(p) - pAUC(p))}{1 + \lambda(\hat{Q}_i(p) - pAUC(p))} \;=\; 0. \quad (4.2)$$

Similarly, we have smoothed log-empirical likelihood ratio,

$$\tilde{l}(p, pAUC(p)) \;=\; 2\sum_{i=1}^{n} \log\{1 + \lambda(\tilde{Q}_i(p) - pAUC(p))\}, \quad (4.3)$$

where Lagrange multiplier λ satisfies the equation

$$\sum_{i=1}^{m+n} \frac{(\tilde{Q}_i(p) - pAUC(p))}{1 + \lambda(\tilde{Q}_i(p) - pAUC(p))} \;=\; 0. \quad (4.4)$$

To build up the theorems of this chapter, we make these assumptions as Wang and Chang (2011):

D.1 Random variables X and Y have uniformly continuous density function;

D.2 m and n are comparable, i.e., $m/n \to r$, $r > 0$;

D.3 $h = O(n^{-1/4})$;

D.4 $p \in (0, 1)$.

Theorem 4.1. *Under the regularity conditions D.1-D.4, we have*

$$\hat{l}(p, pAUC(p)) \xrightarrow{\mathcal{D}} \chi_1^2. \quad (4.5)$$

Theorem 4.2. *Under the regularity conditions D.1-D.4, we have*

$$\tilde{l}(p, pAUC(p)) \xrightarrow{\mathcal{D}} \chi_1^2. \tag{4.6}$$

The proof of Theorem 4.1 and 4.2 is very similar to the proof of Jing et al. (2009). It is a natural extension from the U-statistics to our setting.

4.3 Numerical Studies

In order to investigate the performance of our proposed method in small samples, we conduct comprehensive simulation studies in this section. Focusing on two indexes, the coverage probability and average lengths of confidence intervals, we expect the coverage probability is close to 95% nominal level and average lengths get narrower as sample sizes become larger. In this simulation, we only illustrate partial AUC from 0 to varied p and can easily extend to the case with two flexible p_1 and p_2 and even the difference of two partial AUC's with our methodology. First, we assume random variable X follows the normal distribution with mean 0.2 and standard deviation 0.5 and Y follows a normal distribution with mean 0 and standard deviation 0.5. Using the built-in function in Matlab, we generate the data set A under above assumption. Similarly, we obtain the data set B from the exponential distribution with parameter 1 for X and a normal distribution with mean 1 and standard deviation 0.5 for Y. In data set C, we apply the exponential distribution with parameter 1 for both distributions. Two possible p's in our simulation studies are selected as 0.4 or 0.6. A group of sample sizes m and n are chosen as $(50, 50)$, $(80, 80)$ and $(100, 100)$. The bandwidth is selected as $h = m^{-1/4}$. Table 4.1 and Table 4.2 show the simulation result for the discrete version. Table 4.3 and Table 4.4 demonstrate the performance of jackknife empirical likelihood based on the smoothed estimation equation. From Table 4.1 and Table 4.3, coverage probabilities of 95% confidence interval for the partial AUC at different scenarios are close to 95%. Table 4.2 and Table 4.4 demonstrates the average length of 95% confidence interval for the partial AUC at different scenarios become smaller when the sample size become larger. We make a comparison about two types of interval estimations and realize that jackknife empirical likelihood method with the discrete estimator can provide the comparable average lengths with jackknife EL with smoothed estimator. In a special setting, the discrete estimator has a little under-coverage problem, comparing with jackknife EL in the smoothed estimator. However, it is the trade-off or drawback of smooth jackknife empirical likelihood that people need choose appropriate bandwidth h and argue how to make an optimal selection even if it is not crucial to obtain accurate results based on our simulations.

4.4 Discussion

Inspired by Wang and Chang (2011) and Jing et al. (2009), we proposed jackknife empirical likelihood inference for partial AUC with an elegant estimation which is different from Adimari and Chiogna (2011). Jackknife pseudo-sample avoids many link variables and reduces the computational intensity. Wilk's theorem in our case is proved under regularity conditions. A simulation study shows a great performance for our methodology in small and moderate sample size.

Moreover, we develop jacknife empirical likelihood method for the difference of two partial AUC's. As in Chapter 3, we introduce two-dimensional random variables. Let $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ be two-dimensional random variables. X and Y are independent with bivariate distribution functions $F(x_1, x_2)$ and $G(y_1, y_2)$. Consider two

Table 4.1 Coverage probability of 95% confidence interval for the partial AUC from 0 to p based on discrete estimators.

m	n	p	(A)	(B)	(C)
50	50	0.4	0.907	0.902	0.900
80	80	0.4	0.927	0.912	0.926
100	100	0.4	0.928	0.925	0.926
50	50	0.6	0.949	0.941	0.949
80	80	0.6	0.951	0.961	0.943
100	100	0.6	0.948	0.942	0.946

dimensional data $(X_{1i}, X_{2i}), i = 1, ..., m$, and $(Y_{1j}, Y_{2j}), j = 1, ..., n$. Denote marginal distributions $F_1(x_1)$, $F_2(x_2)$, $G_1(y_1)$ and $G_2(y_2)$. We define the partial AUC from 0 to p, $pAUC_j(p) = \int_0^p ROC_j(t)dt$, where $ROC_j(t) = 1 - F_j(G_j^{-1}(1-t))$, $j = 1, 2$. We can simply get the difference of two partial AUC's as $D(p) = pAUC_1(p) - pAUC_2(p)$, if $0 < p < 1$.

A discrete estimator of the difference of two partial AUC's is easily obtain from Wang and Chang (2011) as follows, $\hat{D}(p) = \hat{A}_1(p) - \hat{A}_2(p)$, where

$$\hat{A}_1(p) = \frac{1}{n}\sum_{i=1}^{n}\left[p - \min\left\{\frac{1}{m}\sum_{j=1}^{m} I(X_{1j} > Y_{1i}), p\right\}\right],$$

$$\hat{A}_2(p) = \frac{1}{n}\sum_{i=1}^{n}\left[p - \min\left\{\frac{1}{m}\sum_{j=1}^{m} I(X_{2j} > Y_{2i}), p\right\}\right].$$

The difference of two partial AUC's is estimated smoothly as follows, $\tilde{D}(p) = \tilde{A}_1(p) - \tilde{A}_2(p)$,

$$\tilde{A}_1(p) = \frac{1}{n}\sum_{i=1}^{n}\left[p - h\log\frac{1+\exp(p/h)}{1+\exp\{[p - \frac{1}{m}\sum_{j=1}^{m} K\{(X_{1j} - Y_{1i})\}/h]/h\}}\right]$$

and

$$\tilde{A}_2(p) = \frac{1}{n}\sum_{i=1}^{n}\left[p - h\log\frac{1+\exp(p/h)}{1+\exp\{[p - \frac{1}{m}\sum_{j=1}^{m} K\{(X_{2j} - Y_{2i})\}/h]/h\}}\right],$$

where $K(t) = 1/[1 + exp(-t)]$.

Following the routine process, we construct two kinds of jackknife pseudo samples for

Table 4.2 Average length of 95% confidence interval for the partial AUC from 0 to p based on discrete estimators.

m	n	p	(A)	(B)	(C)
50	50	0.4	0.1286	0.1436	0.1088
80	80	0.4	0.1036	0.1144	0.0877
100	100	0.4	0.0930	0.1037	0.0785
50	50	0.6	0.1819	0.2085	0.1678
80	80	0.6	0.1455	0.1674	0.1343
100	100	0.6	0.1305	0.1505	0.1209

the difference of two partial AUC's at p. For discrete conditions, $\hat{D}_{-k}(p) = \hat{A}_{-k,1}(p) - \hat{A}_{-k,2}(p), 1 \le k \le m+n$, where

$$\hat{A}_{-k,l}(p) = \frac{1}{n}\sum_{i=1}^{n}\left[p - \min\left\{\frac{1}{m}\sum_{j=1, j\neq k}^{m} I(X_{lj} > Y_{li}), p\right\}\right], 1 \le k \le m, l = 1, 2,$$

$$\hat{A}_{-k,l}(p) = \frac{1}{n}\sum_{i=1, i\neq k-m}^{n}\left[p - \min\left\{\frac{1}{m}\sum_{j=1}^{m} I(X_{lj} > Y_{li}), p\right\}\right], m+1 \le k \le m+n, l = 1, 2.$$

The jackknife pseudo-samples are obtain by

$$\hat{H}_k(p) = (m+n)\hat{D}(p) - (m+n-1)\hat{D}_{-k}(p), 1 \le k \le m+n.$$

Similarly, for the smoothed estimator, the jackknife pseudo-samples are obtained as follows.

$\tilde{D}_{-k}(p) = \tilde{A}_{-k,1}(p) - \tilde{A}_{-k,2}(p), 1 \le k \le m+n$, where

$$\tilde{A}_{-k,1}(p) = \frac{1}{n}\sum_{i=1}^{n}\left[p - h\log\frac{1+\exp(p/h)}{1+\exp\{[p - \frac{1}{m-1}\sum_{j=1,j\neq k}^{m} K\{(X_{lj} - Y_{li})\}/h]/h\}}\right], 1 \le k \le m, l = 1, 2,$$

$$\tilde{A}_{-k,2}(p) = \tfrac{1}{n-1}\textstyle\sum_{i=1,i\neq k-m}^{n-1}\left[p - h\log\frac{1+\exp(p/h)}{1+\exp\{[p-\frac{1}{m}\sum_{j=1}^{m} K\{(X_{lj}-Y_{li})\}/h]/h\}}\right],$$
$$m+1 \le k \le m+n, l = 1, 2.$$

Table 4.3 Coverage probability of 95% confidence interval for the partial AUC from 0 to p based on smoothed estimators.

m	n	p	(A)	(B)	(C)
50	50	0.4	0.947	0.923	0.935
80	80	0.4	0.946	0.938	0.956
100	100	0.4	0.939	0.940	0.950
50	50	0.6	0.955	0.939	0.943
80	80	0.6	0.945	0.944	0.942
100	100	0.6	0.943	0.946	0.948

Hence,

$$\tilde{H}_k(p) = (m+n)\tilde{D}(p) - (m+n-1)\tilde{D}_{-k}(p), 1 \le k \le m+n.$$

From the regular jackknife empirical likelihood method, we obtain log-empirical likelihood ratio for the difference of two partial AUC's under both discrete and smooth estimates,

$$\hat{l}_D(p, D(p)) \;=\; 2\sum_{i=1}^{n} \log\{1 + \lambda(\hat{H}_i(p) - D(p))\}, \tag{4.7}$$

where Lagrange multiplier λ satisfies the equation

$$\sum_{i=1}^{m+n} \frac{(\hat{H}_i(p) - D(p))}{1 + \lambda(\hat{H}_i(p) - D(p))} \;=\; 0. \tag{4.8}$$

Similarly,

$$\tilde{l}_D(p, D(p)) \;=\; 2\sum_{i=1}^{n} \log\{1 + \lambda(\tilde{H}_i(p) - D(p))\}, \tag{4.9}$$

where Lagrange multiplier λ satisfies the equation

$$\sum_{i=1}^{m+n} \frac{(\tilde{H}_i(p) - D(p))}{1 + \lambda(\tilde{H}_i(p) - D(p))} \;=\; 0. \tag{4.10}$$

Table 4.4 Average length of 95% confidence interval for the partial AUC from 0 to p based on smoothed estimators.

m	n	p	(A)	(B)	(C)
50	50	0.4	0.1238	0.1334	0.1068
80	80	0.4	0.1016	0.1106	0.0859
100	100	0.4	0.0914	0.1011	0.0785
50	50	0.6	0.1763	0.2019	0.1635
80	80	0.6	0.1432	0.1643	0.1331
100	100	0.6	0.1291	0.1485	0.1195

Theorem 4.3. *Under the regularity conditions D.1-D.4, we have*

$$\hat{l}_D(p, D(p)) \xrightarrow{\mathcal{D}} \chi_1^2. \tag{4.11}$$

Theorem 4.4. *Under the regularity conditions D.1-D.4, we have*

$$\tilde{l}_D(p, D(p)) \xrightarrow{\mathcal{D}} \chi_1^2. \tag{4.12}$$

The proofs of Theorem 4.3 and 4.4 are similar to those of Theorem 4.1 and 4.2.

Using the same distributions to generate the data sets as we did in Chapter 3, the correlated two dimensional data sets (A), (B) and (C) are applied in this simulation environment. To compare the two partial AUC's from 0 to $p = 0.4$ or from 0 to $p = 0.6$, Table 4.5 and Table 4.6 with discrete estimations and Table 4.7 and Table 4.8 with the smoothed estimations demonstrate as follows. It can be found the coverage probabilities under different settings are close to nominal level 95% and the average lengths became narrow when sample sizes increase. From the comparison between the discrete estimator and smoothed estimator, it is another evidence to support the rule we found in the last section that the jackknife empirical likelihood method with smoothed estimator has better performance in terms of coverage probabilities of confidence interval. However, people need to discuss the selection of bandwidth h for their theoretical rigourousness.

Moreover, we conduct the real data study for the difference of two partial AUC's. Using

Table 4.5 Coverage probability of 95% confidence interval for the partial AUC from 0 to p based on discrete estimators.

m	n	p	(A)	(B)	(C)
50	50	0.4	0.954	0.954	0.918
80	80	0.4	0.925	0.942	0.944
100	100	0.4	0.949	0.946	0.929
50	50	0.6	0.935	0.946	0.937
80	80	0.6	0.940	0.944	0.957
100	100	0.6	0.937	0.949	0.960

Pancreatic Cancer Serum Biomarkers data, the same data set as we used in Chapter 3, we calculate jackknife empirical likelihood confidence intervals for the difference of two partial AUC's at changing criterion level p from 0 to 1. From Figure 4.1, we prefer the first biomarker CA-125 (V1) rather than CA-19-9 (V2) as the index to distinguish pancreatic cancer and pancreatitis.

Furthermore, we may consider the difference of two partial AUC's with covariates. Missing data would be another interesting idea which can be applied to extend JEL for the difference of two partial AUC's.

Table 4.6 Average length of 95% confidence interval for the partial AUC from 0 to p based on the discrete estimators.

m	n	p	(A)	(B)	(C)
50	50	0.4	0.1319	0.1603	0.0674
80	80	0.4	0.1056	0.1272	0.0537
100	100	0.4	0.0960	0.1148	0.0483
50	50	0.6	0.1796	0.2404	0.1222
80	80	0.6	0.1436	0.1912	0.0963
100	100	0.6	0.1292	0.1722	0.0864

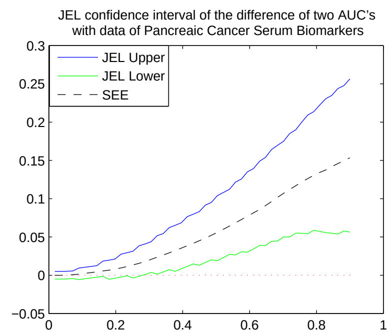

Figure 4.1 95% point-wise jackknife empirical likelihood confidence interval for the difference of two partial AUC's with Pancreatic Cancer Serum Biomarkers, where JEL Upper indicates the upper bound of jackknife empirical likelihood confidence interval, JEL Lower indicates the lower bound of jackknife empirical likelihood confidence interval and SEE means smoothed empirical estimator.

Table 4.7 Coverage probability of 95% confidence interval for the partial AUC from 0 to p based on smoothed estimators.

m	n	p	(A)	(B)	(C)
50	50	0.4	0.943	0.954	0.960
80	80	0.4	0.955	0.951	0.951
100	100	0.4	0.947	0.940	0.953
50	50	0.6	0.946	0.948	0.954
80	80	0.6	0.953	0.960	0.954
100	100	0.6	0.942	0.952	0.951

Table 4.8 Average length of 95% confidence interval for the partial AUC from 0 to p based on smoothed estimators.

m	n	p	(A)	(B)	(C)
50	50	0.4	0.1286	0.1564	0.0668
80	80	0.4	0.1039	0.1260	0.0533
100	100	0.4	0.0966	0.1150	0.0480
50	50	0.6	0.1771	0.2376	0.1210
80	80	0.6	0.1424	0.1898	0.0960
100	100	0.6	0.1283	0.1711	0.0866

CHAPTER 5

SMOOTHED JACKKNIFE EMPIRICAL LIKELIHOOD INFERENCE FOR THE DIFFERENCE OF TWO QUANTILES

5.1 Background

The quantile is an attractive statistics measure with robustness property against the extreme value. Motivated by this favorable property, the researchers from different fields developed the many applications of quantile as their fundamental tools, such as Value-at-Risk (VaR) in risk management, quantile regression in econometrics, etc. Generally, the quantile is defined by $F^{-1}(x) = \inf\{x : F(x) > p\}$, where $0 < p < 1$. Csörgo (1987) established the theoretical foundation of quantile estimation. To overcome the discreteness issue of empirical estimation of quantile, Sheather and Marron (1990) introduced the kernel quantile estimators and the bandwidth selection procedure.

The empirical likelihood (EL) is a popular methodology featured by nonparametric likelihood function. Owen (1988, 1990, 2001) built up the framework of EL as a new philosophy of statistics. Chen (1993) studied the quantile estimation using empirical likelihood method. Qin and Lawless (1994) proposed empirical likelihood for the general estimation equation incorporating side information. Claeskens et al. (2003) developed the EL confidence intervals for ROC curves. Recently, the jackknife empirical likelihood is recognized as a better method over traditional empirical likelihood for complicated nonlinear question due to its computational efficiency in small sample, see Jing et al. (2009) and Gong et al. (2010) for detailed discussions.

For comparison of two quantiles, some well-known methods include Q-Q plot and interquartile range, which are less vulnerable to extreme value and heavy-tail data. Kosorok (1999) developed two-sample quantile nonparametric tests for a variety of empirical distribution function for censored data and repeated measures data. Veraverbeke (2001) studied the asymptotical properties of the difference of quantiles for one sample. Zhou and Jing ($2003a, 2003b$) constructed the smoothed empirical likelihood confidence interval for quantiles and one sample difference of quantiles. Shen and He (2007) proposed the empirical likelihood method for one sample difference of quantiles with right censoring. Yau (2009) has proposed EL method for the difference of two quantiles with censoring. Baysal and Staum (2010) developed the empirical likelihood inference for the value-at-risk and expected shortfall.

In this chapter, we develop the novel method for the inference of the difference of two sample quantiles $\theta(p) = F_1^{-1}(p) - F_2^{-1}(p)$ and one sample difference of two quantiles $\eta(s,t) = F^{-1}(t) - F^{-1}(s)$, where the quantile functions are defined as $F_j^{-1}(x) = \inf\{x : F_j(x) > p\}$, $j = 1, 2$. In order to reduce heavy computational intensity of multiple estimation equations (see Zhou and Jing, 2003; Shen and He, 2007), motivated by Jing et al. (2009), we construct the jackknife pseudo samples and derive the empirical likelihood based on those pseudo samples. Moreover, we obtain the asymptotical result of the difference of quantiles with one sample and two samples using jackknife empirical likelihood and demonstrate the computational efficiency and accuracy in the small sample.

The rest of the chapter is organized as follows. In Section 5.2, we prove that it is an asymptotically chi-squared distribution for jackknife empirical log-likelihood ratio for the difference of two quantiles with one sample. In Section 5.3, we develop the same theorem with two samples. For a small sample simulation studies and real data application, coverage probability and average length of confidence intervals are reported in Section 5.4. Furthermore, we discuss some extensions of our method and future work in Section 5.5. The proofs are provided in the Appendix.

5.2 Main Results

5.2.1 Inference procedure for the difference of quantiles with two samples

Let X_1 and X_2 be independent random variables with distribution functions $F_1(x)$ and $F_2(x)$. The difference of quantiles at p can be written as $\theta(p) = F_1^{-1}(p) - F_2^{-1}(p)$, where $0 < p < 1$ and F_j^{-1} denotes the quantile function of $F_j(x)$, $j = 1, 2$. We can define $D(\theta, p) = F_1(\theta + F_2^{-1}(p))$ and rearrange the difference of quantiles as $D(\theta, p) = p$. $D^{'}(\theta, p)$ is the first derivative of $D(\theta, p)$ with respect to p.

Let $X_{1,i}$, $i = 1, ..., m$ and $X_{2,i}, i = 1, ..., n$ be i.i.d. random samples from the distribution functions $F_1(x)$ and $F_2(x)$, respectively. The empirical estimators of distributions function $F_1(x)$ and $F_2(x)$ are defined by $F_{m,1}(x) = 1/m \sum_{i=1}^{m} I(X_{1,i} \leq x)$, $F_{n,2}(x) = 1/n \sum_{i=1}^{n} I(X_{2,i} \leq x)$. To analyze this problem using continuous function, we use smoothed version of those non-parametric estimators. Let $K(p)$ be the smooth distribution function which satisfies

$$K(p) = \int_{u \leq p} w(u) \mathrm{d}u,$$

where $w(u)$ is a symmetric density function with support $[-1, 1]$. We propose the smooth estimation equation for the difference of two quantiles

$$\Pi_{m,n}(p, \theta) = \frac{1}{m} \sum_{j=1}^{m} K \left\{ \frac{p - F_{n,2}(X_{1,j} - \theta)}{h} \right\} - p,$$

where $h = h(n) > 0$ is a bandwidth. We assume the following regularity conditions,

A.1. $w(u)$ is a symmetric density function with support $[-1, 1]$ and $w^{'}(u)$ is bounded, continuous for $u \in [-1, 1]$;
A.2. Let $p \in (0, 1)$. We assume $m/n \to r$, where $r > 0$;
A.3. $h = h(n) \to 0$, $nh^2/\log n \to \infty$, $nh^4 \to 0$ as $n \to \infty$.

Theorem 5.1. *Under the regularity conditions A.1-A.3, we have*

$$\Pi_{m,n}(p,\theta) \xrightarrow{\mathcal{P}} 0, \tag{5.1}$$

where $\theta(p)$ is a true value of difference of two quantiles at p.

After obtaining a consistent kernel estimation equation of difference of quantiles, we develop the procedure to generate jackknife pesudo-sample. Denote

$$\Pi_{m,n,i}(p,\theta) = \tfrac{1}{m-1}\textstyle\sum_{1\le j\le m, j\ne i} K\left(\tfrac{p-F_{n,2}(X_{1,j}-\theta)}{h}\right) - p, \quad 1 \le i \le m,$$
$$\Pi_{m,n,i}(p,\theta) = \tfrac{1}{m}\textstyle\sum_{j=1}^{m} K\left(\tfrac{p-F_{n,2,m-i}(X_{1,j}-\theta)}{h}\right) - p, \quad m+1 \le i \le m+n,$$

where

$$F_{n,2,-i}(p,\theta) \;=\; \frac{1}{n-1}\sum_{1\le j\le n, j\ne i} I(X_{2,j}\le y), i = 1,\ldots,n.$$

The jackknife pseudo-sample is defined as

$$\hat{V}_i(p,\theta) \;=\; (m+n)\Pi_{m,n}(p,\theta) - (m+n-1)\Pi_{m,n,i}(p,\theta), i = 1,\ldots,m+n. \tag{5.2}$$

We consider the following conditions:

A.4. $F_1(x)$, $F_2(x)$ are continuous functions and have continuous, bounded first derivatives;

A.5. $D(\theta,p)$ and its first derivative $D^{'}(\theta,p)$ are bounded and continuous in $p \in [-1,1]$;

Theorem 5.2. *Under the regularity conditions A.1-A.5, we have*

$$\sqrt{m+n}\left\{\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i(p,\theta)\right\} \xrightarrow{\mathcal{D}} N(0,\sigma^2), \tag{5.3}$$

where

$$\sigma^2 = \frac{1+r}{r}(1-p)p + (1+r)(1-p)pD^{'}(\theta,p)^2.$$

Define the pseudo-sample variance

$$v_{m,n}(p,\theta) = \frac{1}{m+n}\sum_{i=1}^{m+n}\left\{\hat{V}_i(p,\theta) - \frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i(p,\theta)\right\}^2. \tag{5.4}$$

Theorem 5.3. *Under the conditions A.1-A.5, we have*

$$v_{m,n}(p,\theta) \xrightarrow{\mathcal{P}} \sigma^2.$$

Based on jackknife pesudo-sample, the empirical likelihood log-ratio at $\theta(p)$ is

$$L(p,\theta) = \frac{\sup\{\prod_{i=1}^{m+n} p_i : \sum_{i=1}^{m+n} p_i = 1, \sum_{i=1}^{m+n} p_i \hat{V}_i(p,\theta) = 0, p_i > 0, i = 1,\ldots,m+n\}}{\sup\{\prod_{i=1}^{m+n} p_i, \sum_{i=1}^{m+n} p_i = 1, p_i > 0, i = 1,\ldots,m+n\}}.$$

We follow the standard Lagrange multiplier method and have

$$l(p,\theta) = -2\log L(p,\theta) = 2\sum_{i=1}^{n}\log\{1+\lambda\hat{V}_i(p)\}, \tag{5.5}$$

where Lagrange multiplier λ satisfies the equation

$$\sum_{i=1}^{m+n}\frac{\hat{V}_i(p,\theta)}{1+\lambda\hat{V}_i(p,\theta)} = 0. \tag{5.6}$$

Assuming the regularity conditions A.1-A.5, we can establish the main theorem as follows.

Theorem 5.4. *Under the regularity conditions of Theorem 5.1, we have*

$$l(p,\theta) \xrightarrow{\mathcal{D}} \chi_1^2, \tag{5.7}$$

where θ is the true value of the difference of quantiles at $p \in (0,1)$.

Remark: Under our settings for the difference of quantiles with two samples, it is straightforward to extend two quantiles from fixed one point to different two points.

Thus, using smoothed jackknife empirical likelihood method, the asymptotic $100(1-\alpha)\%$ jackknife EL confidence interval for the difference of two quantiles is proposed as

$$I(p) = \left\{\tilde{\theta} : l(p, \tilde{\theta}) \leq \chi_1^2(\alpha)\right\},$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2.

5.2.2 Inference procedure for the difference of quantiles with one sample

Suppose X is a random variable and $F(x)$ is its distribution function. The difference of quantiles with one sample between s and t is defined as $\eta(s,t) = F^{-1}(t) - F^{-1}(s)$, where $0 < s < t < 1$ and $F^{-1}(x)$ is the quantile function. Let X_i, $i = 1, ..., m$ be i.i.d. random sample from the distribution function $F(x)$. Denote $F_m(x) = 1/m \sum_{i=1}^m I(X_i \leq x)$ as the empirical estimators of distributions function. Consider the smooth estimation equation for the the difference of quantiles with one sample

$$\Phi_m(s, t, \eta) = \frac{1}{m} \sum_{j=1}^{m} K \left\{ \frac{s - F_m(X_j - \eta)}{h} \right\} - t,$$

where $K(x)$ is the smooth distribution function with bandwidth $h = h(n) > 0$ defined previously. Assume the following regularity conditions,

B.1. $w(u)$ is a symmetric density function with support $[-1, 1]$ and $w'(u)$ is bounded, continuous for $u \in [-1, 1]$;
B.2. $h = h(m) \to 0$, $mh^2/\log m \to \infty$, $mh^4 \to 0$ as $m \to \infty$.

Theorem 5.5. *Under the regularity conditions B.1, B.2, we have*

$$\Phi_m(s, t, \eta) \xrightarrow{\mathcal{P}} 0, \tag{5.8}$$

where η is a true difference of quantiles with one sample between at t and s.

Theorem 5.6. *Under the regularity conditions B.1-B.2, we have*

$$\sqrt{m}\Phi_m(s,t,\eta) \xrightarrow{\mathcal{D}} N(0,\sigma_1^2), \tag{5.9}$$

where

$$\sigma_1^2 = (1-s)sQ^{'}(s,t,\eta)^2 + 2(1-s)sQ^{'}(s,t,\eta) + t(1-t),$$

$Q(s,t,\eta) = F(\eta + F^{-1}(s)) - t$, *and* $Q^{'}(s,t,\eta)$ *the first derivative of* $Q(s,t,\eta)$ *with respect to* s.

Further, we propose jackknife pesudo-sample based on our estimation equation. Denote

$$\Phi_{m,-i}(s,t,\eta) = \tfrac{1}{m-1}\textstyle\sum_{j=1}^{m} K\left(\tfrac{s-F_{m,-i}(X_j-\eta)}{h}\right) - t, \quad 1 \leq i \leq m,$$

where

$$F_{m,-i}(y) \quad = \quad \frac{1}{m-1}\sum_{1\leq j\leq n, j\neq i} I(X_j \leq y), i = 1,\ldots,m.$$

The jackknife pseudo-sample is defined as

$$\hat{U}_i(s,t,\eta) \quad = \quad m\Phi_m(s,t,\eta) - (m-1)\Phi_{m,-i}(s,t,\eta), i = 1,\ldots,m. \tag{5.10}$$

We consider the following conditions:

B.3. $F(x)$ is continuous functions and has continuous, bounded first derivative;

B.4. $Q(s,t,\eta)$ and its first derivative $Q^{'}(s,t,\eta)$ are bounded and continuous;

Theorem 5.7. *Under the regularity conditions B.1-B.4, we have*

$$\sqrt{m}\left\{\frac{1}{m}\sum_{i=1}^{m}\hat{U}_i(s,t,\eta)\right\} \xrightarrow{\mathcal{D}} N(0,\sigma_1^2), \tag{5.11}$$

Define the pseudo-sample variance

$$v_m(s,t,\eta) = \frac{1}{m}\sum_{i=1}^{m}\left\{\hat{U}_i(s,t,\eta) - \frac{1}{m}\sum_{i=1}^{m}\hat{U}_i(s,t,\eta)\right\}^2. \tag{5.12}$$

Theorem 5.8. *Under the conditions B.1-B.4, we have*

$$v_m(s,t,\eta) \xrightarrow{\mathcal{P}} \sigma_1^2.$$

Based on jackknife pesudo-sample, the empirical likelihood log-ratio at $\eta(p)$ is

$$\tilde{L}(s,t,\eta) = \frac{\sup\{\prod_{i=1}^{m} p_i : \sum_{i=1}^{m} p_i = 1, \sum_{i=1}^{m} p_i\hat{U}_i(s,t,\eta) = 0, p_i > 0, i = 1,\ldots,m\}}{\sup\{\prod_{i=1}^{m} p_i, \sum_{i=1}^{m} p_i = 1, p_i > 0, i = 1,\ldots,m\}}.$$

We follow the standard Lagrange multiplier method and have

$$\tilde{l}(s,t,\eta) = -2\log L(s,t,\eta) = 2\sum_{i=1}^{n}\log\{1+\lambda\hat{U}_i(s,t,\eta)\}, \tag{5.13}$$

where Lagrange multiplier λ satisfies the equation

$$\sum_{i=1}^{m+n}\frac{\hat{U}_i(s,t,\eta)}{1+\lambda\hat{U}_i(s,t,\eta)} = 0. \tag{5.14}$$

Theorem 5.9. *Under the above regularity conditions B.1-B.4, we have*

$$\tilde{l}(s,t,\eta) \xrightarrow{\mathcal{D}} \chi_1^2, \tag{5.15}$$

where η is the true the difference of quantiles with one sample between s and t.

Using smoothed jackknife empirical likelihood method, the $100(1-\alpha)\%$ jackknife EL confidence interval for the difference of quantiles with one sample is constructed as,

$$I(t,s) = \left\{\tilde{\eta} : \tilde{l}((s,t,\tilde{\eta}) \le \chi_1^2(\alpha)\right\},$$

where $\chi_1^2(\alpha)$ is the upper α-quantile of χ_1^2.

5.3 Numerical Studies

5.3.1 Two sample simulation

We conduct a comprehensive simulation study to illustrate our method. It includes two benchmarks, the converge probability and average lengths of confidence intervals and involves various factors, such as sample size, the value of p and the distribution functions F_1 and F_2. For data set A, $F_1(x)$ is generated from a normal distribution with mean 0.2 and standard deviation 0.5. $F_2(x)$ is a normal distribution with mean 0 and standard deviation 0.5. For data set B, X_1 is simulated from the exponential distribution with parameter 1, and $F_2(x)$ is a normal distribution with mean 1 and standard deviation 0.5. For data set C, $F_1(x)$ and $F_2(x)$ have the same exponential distribution with parameter 1. In our simulation studies, two points on the difference of quantiles are selected to $p = 0.2, 0.6$, and sample size m and n are chosen as (50, 50), (100, 100) and (200,150). Furthermore, the kernel function is the Epanechnikov kernel function

$$w(u) = \begin{cases} \frac{3}{4}(1 - u^2) & \text{if } |u| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Without the secondary estimation, the bandwidth is determined automatically as $h = m^{-1/3}$. We utilized the f-solve function in Matlab to solve the λ in (2.6). The nominal level is fixed at 95% and data sets are simulated with 1000 repetitions.

The coverage probability in Table 5.1 can reach the nominal level 95% closely. The simulation results of average lengths are reported in Table 5.2.

Furthermore, the bisection method with small jump 0.001 is applied for seeking the upper bound and lower bound of the difference of quantiles at the level $p = 0.2$ or $p = 0.6$. In Table 5.2, we find that the results with larger sample size have shorter average length.

In addition, using the proposed method, we make a plot to illustrate the confidence interval for the difference of quantiles. In order to show the effect of sample sizes and specified points, we choose two sample sizes $(50, 50)$ and $(300, 300)$ and 100 points on x-axis. Figure 5.1 shows the smoothed jackknife empirical likelihood confidence interval for the difference of two quantiles and its smoothed empirical estimators. Under the setting A, the true value of the difference of two quantiles is constantly equal to 0.2, which is inside the jackknife empirical likelihood confidence intervals at almost every point except for some boundary points in Figure 5.1. Under the setting C, two identical exponential distributions, the jackknife empirical likelihood confidence intervals includes the x-axis, the true value of the difference of two quantiles. From Figure 5.1, we can also observe the narrower confidence interval in the larger samples.

5.3.2 One sample simulation

Inter-quartile range is the most widely used for the difference of one sample quantile, which specifics $s = 0.25$ and $t = 0.75$. We generate data from a chi-squared distribution with degree of freedom 2 (Distribution D), an exponential with parameter 2 (Distribution E) and a normal distribution with mean 0 and standard deviation 0.5 (Distribution F). The small sample sizes are selected as 50, 80 and 100. Two benchmarks, coverage probability and average length of confidence intervals, are considered in our simulation. The results are reported at Table 5.3 and Table 5.4. Coverage probabilities are close to nominal 95% level in most cases for moderate sample sizes, average lengths become smaller as the sample sizes

Table 5.1 Coverage probability of 95% confidence interval for the difference of quantiles with two samples at p.

m	n	p	(A)	(B)	(C)
50	50	0.2	0.941	0.907	0.929
100	100	0.2	0.948	0.895	0.954
200	150	0.2	0.943	0.903	0.944
50	50	0.6	0.941	0.951	0.942
100	100	0.6	0.956	0.939	0.939
200	150	0.6	0.931	0.944	0.942

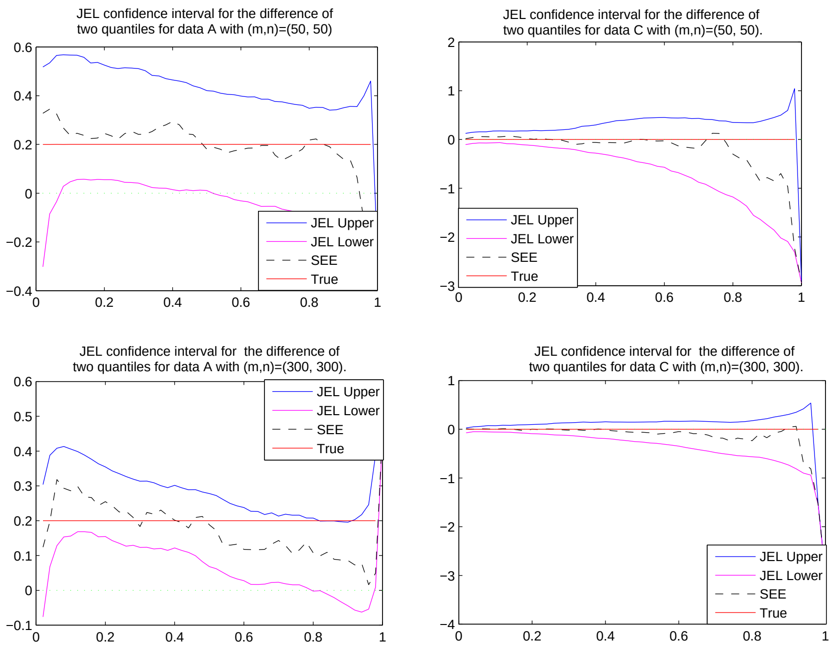

Figure 5.1 95% point-wise jackknife empirical likelihood confidence interval for the difference of quantiles, where JEL Upper indicates the upper bound of jackknife empirical likelihood confidence interval, JEL Lower indicates the lower bound of jackknife empirical likelihood confidence interval and SEE means smoothed empirical estimator.

get larger.

5.4 Real Application

Moreover, we apply the real data to illustrate our proposed method. Data set can be accessed publicly from the website of the Center for Machine Learning and Intelligent Systems at University of California, Irvine and are originated from Hewlett-Packard Labs. It contains 4601 observations with 57 attributes and one indicator variable for spam e-mails, which are considered as the advertisements for products or web sites, make money fast schemes and pornography. Most of those attributes are valued by percentages of certain words appearing in the e-mail. In this chapter, we split the 24th attribute into two groups based on the spam indicator variable. Figure 5.2 shows the confidence interval for the difference of quantiles. The confidence intervals of the difference of quantiles are above x-axis. Thus, the quantiles of the spam observation in 24th attribute are clearly different from the non-spam observation when $p > 0.6$.

5.5 Discussion

Motivated by the challenge and importance of the difference of quantiles, we develop smoothed jackknife empirical likelihood inference methods. JEL includes jackknife pseudo-sample procedure and reduces the number of variables in optimization. It makes the com-

Table 5.2 Average length of 95% confidence interval for the difference of quantiles with two samples at p.

m	n	p	(A)	(B)	(C)
50	50	0.2	0.5121	0.4303	0.3513
100	100	0.2	0.3679	0.3072	0.2484
200	150	0.2	0.2872	0.2429	0.1958
50	50	0.6	0.4483	0.7543	0.8796
100	100	0.6	0.3285	0.5311	0.6389
200	150	0.6	0.2577	0.3957	0.4943

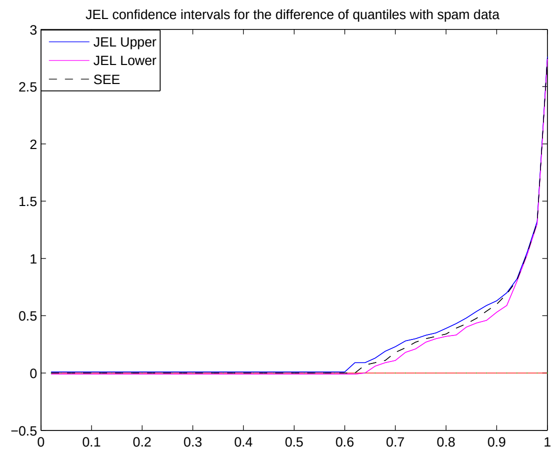

Figure 5.2 95% point-wise jackknife empirical likelihood confidence interval for the difference of quantiles from spam data, where JEL Upper indicates the upper bound of jackknife empirical likelihood confidence interval, JEL Lower indicates the lower bound of jackknife empirical likelihood confidence interval and SEE means smoothed empirical estimator.

putation less intensive tremendously. Our simulation studies show this intuitive argument without sacrificing the accuracy in terms of coverage probability. It is close to the nominal level. We develop a smoothed estimation equation for the difference of quantiles and implement the jackknife empirical likelihood studies. We prove the Wilk's theorem and verify the conclusion with extensive simulation studies.

5.5.1 Missing data

Moreover, we can combine the incomplete data mechanism into our setup, such as missing completely at random (MCAR). Following the proof of Lemma A.1 in Yang and Zhao (2012), we can derive a generalization of Wilk's theorem for the difference of two quantiles with two populations. Denote P_1 and P_2 be response rates of two populations. We similarly prove the theorem as below.

$$l_{MCAR}(p,\theta) \xrightarrow{\mathcal{D}} c(P_1,P_2)\chi_1^2,$$

where $c(P_1,P_2) = \sigma_{MCAR}^2/\sigma^2$ and

$$\sigma_{MCAR}^2 = (1-P_1+P_1^{-1})\frac{1+r}{r}(1-p)p + (1-P_2+P_2^{-1})(1+r)(1-p)pD^{'}(\theta,p)^2.$$

Furthermore, for one sample case, it is straightforward to extend JEL with missing data.

Table 5.3 Coverage probability of 95% confidence interval for inter-quartile range with one sample at $s = 0.25$ and $t = 0.75$.

n	s	t	(D)	(E)	(F)
50	0.25	0.75	0.947	0.946	0.900
80	0.25	0.75	0.956	0.956	0.920
100	0.25	0.75	0.947	0.948	0.925

5.5.2 Dependence

Independent two distributions are a special case of a two dimensional joint distribution. Hence, to extend our conclusions, we can construct jackknife empirical likelihood for the difference of two quantiles based on a joint distribution with a well-defined copula. Wilks' theorem should be valid even if we have not checked it in details. Furthermore, we can consider the covariates into the difference of two quantiles while the dependency of two distributions is associated with the covariates.

5.5.3 Quantile ratio and low income proportion

Moreover, it is straightforward to consider another useful statistics, the ratio of quantiles, i.e., $\theta_r = F_1^{-1}(p)/F_2^{-1}(p)$ for two samples or $\eta_r = F^{-1}(s)/F^{-1}(t)$ for one sample case. Motivated by the JEL for the ROC curve by Gong et al. (2010) and one sample difference of two quantiles in Section 5.3, we can derive the Wilk's theorem, i.e., two negative log likelihood ratio converges to χ_1^2.

Furthermore, the income distribution F_I is strongly concerned in welfare economics and a key index, low income proportion, $\xi(\alpha, \beta) = F_I(\alpha F_I^{-1}(\beta))$, has been investigated by many statisticians and econometricians, such as Zheng (2001) and Yang et al. (2011). The index is a simple transformation from quantile ratio in analytical sense. Without changing the estimation equation, the JEL for quantile ratio can be applied to the inference for the low income proportion.

More specifically, we provide the estimation equation for the ratio of quantiles and low

Table 5.4 Average length of 95% confidence interval for inter-quartile range with one sample at $s = 0.25$ and $t = 0.75$.

n	s	t	(D)	(E)	(F)
50	0.25	0.75	0.7139	1.7805	0.3646
80	0.25	0.75	0.5800	1.3937	0.2965
100	0.25	0.75	0.5293	1.2621	0.2712

income proportion based on the smoothing method. Under the same kernel function $K(\cdot)$, we have the simple estimation equation as follows

$$\Psi(s, t, \eta_r) = \frac{1}{m} \sum_{i=1}^{n} K\left(\frac{t - F_m(X_j - \eta_r)}{h}\right) - s.$$

Following the jackknife pseudo sample, it is not difficult to construct log empirical likelihood ratio and prove the Wilks's theorem as before.

CHAPTER 6

NEW EMPIRICAL LIKELIHOOD INFERENCE FOR LINEAR TRANSFORMATION MODELS

6.1 Introduction

The well-known proportional hazards model was introduced by Cox (1972). Later, Andersen and Gill (1982) explored the Cox model using martingale theory. The Cox model is the most popular method utilized broadly in survival analysis. An alternative method in survival analysis is the proportional odds model (Pettitt, 1982; Bennett 1983). The transformation model is a natural generalization of those two models and provides many other potential choices. Cheng, Wei and Ying (1995) derived a limiting theory of the transformation model using martingale theory. Based on new estimation equations, Chen, Jin and Ying (2002) develop the inference procedure for the linear transformation model. Let T be the failure time; Z, a corresponding p-dimensional covariate; $S_z(\cdot)$ is the survival function of T conditioned on covariate Z. Then, the semiparametric transformation model is (see Cheng et al., 1995)

$$g\{S_z(t)\} = h(t) + Z^T\beta, \tag{6.1}$$

where $h(\cdot)$ is a strictly increasing unspecified function and $g(\cdot)$ is a given decreasing function. An alternative expression of (6.1) is (see Cheng et al., 1995)

$$h(T) = -Z^T\beta + \varepsilon, \tag{6.2}$$

where ε is a random variable indepedent of covariate Z with the distribution function $F(x) = 1 - g^{-1}(x)$. Fine et al. (1998) considered the truncated t_0 to place a finite limit on survival time and guaranteed the uniform convergence of Gaussian processes on interval $[0, t_0]$. Other related studies about the transformation model include Cai et al. (2000) and Cai et al. (2005). Recently, Kong et al. (2006) investigated the case-cohort problems using semi-parameteric linear transformation models.

6.1.1 Empirical likelihood

Recently, empirical likelihood method has been extended to some diverging number of dimensionality, such as Hjort et al. (2009), Chen et al. (2009). Moreover, Zhao (2010) demonstrated that the empirical likelihood method for transformation models can outperform traditional methods in small samples. However, the methodology of Zhao (2010) sacrificed the tremendous computational resource on estimating the covariance matrix. Motivated by Yu et al. (2011), we can construct new empirical likelihood for the transformation model which avoids estimating the complicated matrix.

The rest of the chapter is organized as follows. In Section 6.2, we develop the new empirical likelihood method for the linear transformation model. Then, we report results of simulation studies in terms of coverage probability in Section 6.3. A discussion about JEL method is given in Section 6.4. The proofs are provided in the Appendix.

6.2 Inference Procedure

6.2.1 Preliminaries

Throughout the chapter, we use same notations as Fine et al. (1998). Let T_i be the failure time which might not be observed fully. The censoring variables C_i with distribution function $G(t)$ are independent of failure time T_i. Define bivariate vector (X_i, δ_i), $i = 1, ..., n$, where $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$. Let $\{Z_i\}_{i=1}^n$ be the corresponding covariate vectors, where $Z_i \in \mathrm{R}^p$. We denote $Z_{ij} = Z_i - Z_j$, $i = 1, ..., n, j = 1, ..., n$. Fine et al. (1998) introduced a known constant t_0, where $\Pr\{\min(T, C) > t_0\} > 0$. Denote the h_0 and α_0 as true values of $h(\cdot)$ and $\alpha = h(t_0)$. Define $\theta = (\alpha, \beta^T)^T$ and true $\theta_0 = (\alpha_0, \beta_0^T)^T$ and

$$\eta_{ij}(\theta_0) = \eta(Z_{ij}^T \beta_0) - \Pr(T_i \geq T_j \geq t_0 | Z_i, Z_j) \tag{6.3}$$

where $\eta(Z_{ij}^T\beta_0) = \Pr(\varepsilon_i - \varepsilon_j \geq Z_{ij}^T\beta_0)$. Let $\hat{G}(\cdot)$ be the Kaplan-Meier estimator of G. Combining the positive weight function $w_{ij}(\cdot)$, Fine et al. (1998) proposed the following estimating equation $U_w(\theta)$,

$$U_w(\theta) = \sum_{i=1}^{n} \sum_{j=1, i\neq j}^{n} w_{ij}(\theta)\dot{\eta}_{ij}(\theta) \left\{ \frac{\delta_j I\{\min(X_i, t_0) \geq X_j\}}{\hat{G}^2(X_j)} - \eta_{ij}(\theta) \right\}, \tag{6.4}$$

where

$$\begin{aligned} \dot{\eta}_{ij}(\theta) = (1, Z_j^T)^T \textstyle\int_{-\infty}^{\alpha} \{1 - F(t + Z_i^T\beta)\}\, df(t + Z_j^T\beta) \\ -(1, Z_i^T)^T \textstyle\int_{-\infty}^{\alpha} \{1 - f(t + Z_i^T\beta)\}\, dF(t + Z_j^T\beta), \end{aligned} \tag{6.5}$$

and $f(t) = dF(t)/dt$. Cheng et al. (1995) and Fine et al. (1998) proposed the following notations.

$$e_{ij}(\theta) = w_{ij}(\theta)\dot{\eta}_{ij}(\theta)[\tfrac{\delta_j I(\min(X_i, t_0) \geq X_j)}{G^2(X_j)} - \eta_{ij}(\theta)],$$

$$\pi(t) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I(X_i \geq t),$$

$$q(\theta, t) = \lim_{n\to\infty} \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} w_{ij}(\theta)\dot{\eta}_{ij}(\theta) \frac{\delta_j I(\min(X_i, t_0) \geq X_j)}{G^2(X_j)} I(X_j \geq t).$$

6.2.2 Methodology

We will develop new empirical likelihood method in this section. Denote notations like Fine et el. (1998)

$$d_i(\theta) = 2\int_0^{t_0} \frac{q(\theta, t)}{\pi(t)}\, dM_i(t), \tag{6.6}$$

where

$$dM_i(t) = I(X_i \leq t, \delta_i = 0) - \textstyle\int_0^t I(X_i \geq u) d\Lambda_G(u), \tag{6.7}$$

with $\Lambda_G(u)$ is the cumulative hazard function of censoring time C_i. Denote as Fine et el. (1998)

$$\hat{q}(\theta,t) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} w_{ij}(\theta)\dot{\eta}_{ij}(\theta) \frac{\delta_j I\{\min(X_i,t_0) \geq X_j\}}{\hat{G}^2(X_j)} I(X_j \geq t), \tag{6.8}$$

$$\hat{\pi}(t) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \geq t),$$

$$\hat{d}_i(\theta) = 2\int_0^{t_0} \frac{\hat{q}(\theta,t)}{\hat{\pi}(t)}\, d\hat{M}_i(t),$$

where

$$\hat{M}_i(t) = I(X_i \leq t, \delta_i = 0) - \int_0^t I(X_i \geq u) d\hat{\Lambda}_G(u),$$

and $\hat{\Lambda}_G(t)$ is the Nelson-Aalon estimator of $\Lambda_G(u)$. Denote $U_i = (Z_i^T, X_i, C_i)$. We define the symmetric kernel of U-statistics like Zhao (2010)

$$b(U_i, U_j; \theta) = \{e_{ij}(\theta) + d_i(\theta) + e_{ji}(\theta) + d_j(\theta)\}$$

for given $G(\cdot)$ and

$$\hat{b}(U_i, U_j; \theta) = \{\hat{e}_{ij}(\theta) + \hat{d}_i(\theta) + \hat{e}_{ji}(\theta) + \hat{d}_j(\theta)\}$$

where

$$\hat{e}_{ij}(\theta) = w_{ij}(\theta)\dot{\eta}_{ij}(\theta)\{\frac{\delta_j I(\min(X_i,t_0) \geq X_j)}{\hat{G}^2(X_j)} - \eta_{ij}(\theta)\}.$$

We denote

$$W_i(\theta) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} \{b(U_i, U_j; \theta)\},$$

and

$$\hat{W}_i(\theta) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} \{\hat{b}(U_i, U_j; \theta)\}.$$

Then, we have $p+1$ dimensional multivariate U-statistics

$$V(\theta) = \frac{1}{n}\sum_{i=1}^{n} W_i(\theta),$$

and

$$\hat{V}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_i(\theta).$$

One can define empirical likelihood $L(\theta)$ as follows,

$$L(\theta) = \sup\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \hat{W}_i(\theta) = 0, p_i \geq 0\right\}. \tag{6.9}$$

By using the Lagrange multiplier method (Owen, 1988, 1990), we have

$$l(\theta) = -2\log\{n^n L(\theta)\} = 2\sum_{i=1}^{n} \log\{1 + \lambda(\theta)^T \hat{W}_i(\theta)\}, \tag{6.10}$$

where $\lambda(\theta)$ is a $p+1$ dimensional Lagrange multiplier θ which satisfies the following equation

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{W}_i(\theta)}{1 + \lambda(\theta)^T \hat{W}_i(\theta)} = 0. \tag{6.11}$$

Here and throughout this chapter, we assume the following regularity conditions (see Fine et al., 1998 and Kong et al., 2005):

Assumption 1. $Z_i \in \mathcal{R}^p,\ i = 1, ..., n$ are bounded.

Assumption 2. For any $i = 1, ..., n$, $\partial F(t - Z_i^T\beta)/\partial t$ and $\partial f(t - Z_i^T\beta)/\partial t$ and $\partial^2 f(t - Z_i^T\beta)/\partial t^2$ exist on $t \in [0, \alpha_0]$ and they are uniformly continuous on a compact set Θ of θ.

Assumption 3. For any $i = 1, ..., n$ and $j = 1, ..., n$, $w_{ij}(\theta) > 0$ and $\partial w_{ij}(\theta)/\partial\theta$ exist on Θ and they are uniformly continuous on Θ.

Assumption 4. The functions $w(\cdot)$ and $\eta(\cdot)$ are first derivative continuous.

Assumption 5. The $D(\theta_0)$ and $\Gamma(\theta_0)$ are positive definite matrices, where $D(\theta_0)$ and

$\Gamma(\theta_0)$ are given in Fine et al. (1998).

Theorem 6.1. *Under Assumptions 1-5 consider the null hypothesis $\theta = \theta_0$, as $n \longrightarrow \infty$*

$$\frac{1}{4}l(\theta_0) \xrightarrow{\mathfrak{D}} \chi^2_{p+1}, \tag{6.12}$$

where χ^2_{p+1} is a standard chi-squared random variable with $p+1$ degrees of freedom.

Thus, an asymptotic $100(1-\alpha)\%$ empirical likelihood confidence region for θ can be established as

$$R_\alpha = \left\{\theta : \frac{1}{4}l(\theta) \leq \chi^2_{p+1}(\alpha)\right\},$$

where $\chi^2_{p+1}(\alpha)$ is the upper α-quantile of distribution of χ^2_{p+1}.

Next, we construct the empirical likelihood confidence region for θ_1, a sub-vector of θ. Define $\theta_0 = (\theta_{10}^T, \theta_{20}^T)^T$, The hypothesis is $H_0 : \theta_1 = \theta_{10}$, where $\theta_1 \in R^q$ and $\theta_2 \in R^{p+1-q}$. Based on the above proposed method, the profile empirical likelihood ratio is defined as

$$l^*(\theta_1) = \inf_{\theta_2} l(\theta_1, \theta_2).$$

Following Qin and Lawless (1994), we have Theorem 2.2 for the profile log-empirical likelihood ratio $l^*(\theta_1)$,

Theorem 6.2. *Under Assumptions 1-5, consider the null hypothesis $H_0 : \theta_1 = \theta_{10}$, as $n \longrightarrow \infty$,*

$$\frac{1}{4}l^*(\theta_{10}) \xrightarrow{\mathfrak{D}} \chi^2_q. \tag{6.13}$$

Thus, we can construct the empirical likelihood confidence region for θ_{10} with $100(1-\alpha)\%$ level. Define the EL confidence region

$$R^*_\alpha = \left\{\theta_1 : \frac{1}{4}l^*(\theta_1) \leq \chi^2_q(\alpha)\right\}.$$

We establish the theorem with four chi-squared distribution which avoids estimating the complicated matrix in Zhao (2010).

6.3 Numerical Studies

We conduct several simulation studies to verify our theorems. We compare the coverage probability of the new empirical likelihood confidence region for relatively small samples with normal approximation (NA) confidence region proposed by Fine et al. (1998). The link function h is the natural logarithm function. The ϵ is generated from a standard extreme value function, which specifies the transformation model as the proportional hazards model. The survival time is obtained from above settings. Let $w_{ij}(\cdot) = 1$. t_0 is corresponding to 20% upper quantile of censoring data. The censoring time follows uniform distribution from 0 to c, where c is the parameter used to adjust the censoring rate, such as 0.1, 0.2, 0.3 and 0.4. In the first data setting (**A**), $\beta = (-0.5, 0.5)$. Z_1 follows uniform distribution between 0 and 1, and Z_2 follows Bernoulli distribution with a parameter of 0.2. For the second data setting (**B**), we let $\beta_1 = 1$ and $\beta_2 = 0$. Z_1 follows uniform distribution $[0, 1]$, and Z_2 follows Bernoulli distribution with 0.2. We choose the sample sizes 60 and 100. With 1000 repetitions, coverage probabilities of 95% empirical likelihood and normal approximation confidence regions for θ are reported in Table 6.1. For the sample size 60, the new empirical likelihood has better performance than normal approximation does. When the sample size increases to 100, the estimated coverage probabilities for both methods are close to 95% nominal level.

6.4 Discussion

In this chapter, we proposed the empirical likelihood procedure for the semiparametric transformation model. After adjusting each term of the estimating equations, we derived the limiting distribution of log-empirical likelihood ratio. In the proof, we combined the properties of U-statistics and martingale techniques. Moreover, we conducted a simulation study in terms of coverage probability and observed that empirical likelihood method has an empirical advantage in small sample settings.

In recent years, high dimensional data analysis has dominated in statistical community. The diverging number of dimensions p should be considered into the transformation model in the future. However, problems of the uniform convergence in high dimensional data analysis will be an extremely challenging topic for statistical researchers. On the other hand, the simulation algorithm needs to be optimized appropriately because the computational burden would be extremely heavy in the high dimensional situation.

We propose a jackknife EL for the transformations model in order to improve the efficiency. We use the jackknife procedure for our current estimation equation $\hat{V}(\theta)$ and obtain the pseudo-sample to establish the empirical likelihood. The following estimator from partial samples without lth observation is as follows.

$$\hat{V}_{-l}(\theta) = \frac{1}{n(n-1)} \sum_{i=1, i\neq l}^{n} \sum_{j=1, j\neq i, j\neq l}^{n} \{\hat{b}(U_i, U_j; \theta)\}.$$

Thus, we define jackknife pseudo samples.

$$\hat{Q}_l(\theta) = n\hat{V}(\theta) - (n-1)\hat{V}_{-l}(\theta), l = 1, ..., n.$$

As Jing et al. (2009) showed, we propose the jackknife empirical likelihood procedure. The

Table 6.1 Coverage probability of 95% empirical likelihood confidence region for β.

n	censoring rate	EL (A)	NA (A)	EL (B)	NA (B)
60	10%	0.951	0.909	0.924	0.906
60	20%	0.940	0.927	0.955	0.926
60	30%	0.958	0.926	0.969	0.943
60	40%	0.942	0.914	0.956	0.929
100	10%	0.939	0.939	0.948	0.956
100	20%	0.951	0.925	0.952	0.955
100	30%	0.949	0.944	0.958	0.951
100	40%	0.955	0.938	0.960	0.934

empirical likelihood ratio is

$$l_J(\theta) = \frac{\sup\left\{\prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{Q}_i(\theta) = 0, p_i \geq 0\right\}}{\sup\left\{\prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, p_i \geq 0\right\}}. \tag{6.14}$$

Using the Lagrange multiplier method (Owen, 1988, 1990), we have

$$l_J(\theta) = -2\log\{n^n L_J(\theta)\} = 2\sum_{i=1}^n \log\{1 + \lambda(\theta)^T \hat{Q}_i(\theta)\}, \tag{6.15}$$

where θ satisfies the following equation

$$\frac{1}{n}\sum_{i=1}^n \frac{\hat{Q}_i(\theta)}{1 + \lambda(\theta)^T \hat{Q}_i(\theta)} = 0. \tag{6.16}$$

Thus, the asymptotical theory about jackknife empirical likelihood for the transformation model is shown as

Theorem 6.3. *Under regulation conditions, consider the null hypothesis $\theta = \theta_0$, as $n \longrightarrow \infty$*

$$l_J(\theta_0) \xrightarrow{\mathfrak{D}} \chi^2_{p+1}, \tag{6.17}$$

where χ^2_{p+1} is a standard chi-squared random variable with $p+1$ degrees of freedom.

For the proof of Theorem 6.3, using the consequence about jackknife empirical likelihood method for U-statistics proposed by Jing et al. (2009) and the scheme of proofs of Zhang and Zhao (2012), we can prove the Wilks' theorem and obtain the jackknife empirical likelihood confidence interval for β.

CHAPTER 7

CONCLUSIONS

In the first part of this dissertation, we apply smoothed jackknife empirical likelihood (JEL) method to construct confidence intervals for the receiver operating characteristic (ROC) curve with missing data. After using hot deck imputation, we generate pseudo-jackknife sample to develop jackknife empirical likelihood. Comparing to traditional empirical likelihood method, the smoothed JEL has a great advantage in saving computational cost. The smoothed jackknife empirical likelihood ratio converges to a scaled chi-squared distribution. Furthermore, simulation studies support our conclusion.

Next, the difference of two correlated receiver operating characteristic (ROC) curves is used to identify diagnostic tests with stronger discriminant ability. We employ JEL method to construct confidence intervals for the difference of two correlated continuous-scale ROC curves. Under mild conditions, we prove that the smoothed jackknife empirical log likelihood ratio is asymptotically chi-squared distribution. We carry out an extensive simulation study to demonstrate the goo d performance. A real data set is used to illustrate our method.

Partial AUC is a practical and useful measurement for assessing the diagnostic test. We proposed the JEL method for the inference of the partial AUC and the difference of two pAUC's. We prove that the Wilks' theorem for JEL method still holds. Using the jackknife pseudo-sample, we can avoid estimating several nuisance variables which have to be estimated in existing methods. Furthermore, we conduct the simulation studies to demenstrate the good performance with a moderate computational cost.

Quantile is a well-known robust statistics measure. Some derivatives, such as the difference of two quantiles, are natural measure to compare two populations and check the pattern of its distribution, such as inter-quartile range and tail-behavior. We propose a smoothed nonparametric estimation equation for the difference of two quantiles with one sample or two samples. Using the jackknife pseudo-sample technique for the estimation equation, we construct the empirical likelihood (EL) ratio and study its asymptotical properties. Due to avoiding estimating link variables, the simulation studies demonstrate that jackknife empirical likelihood method has computational efficiency compared with traditional EL methods. Coverage probability and average length of confidence intervals support our methods. Furthermore, we can apply the JEL to make inference for the low income proportion, ratio of quantiles, etc.

The transformation model plays an important role in survival analysis. We study the linear transformation model based on new empirical likelihood. Motivated by Fine et al. (1998) and Yu et al. (2011), we introduce the truncated survival time t_0 and adjust each term of estimating equations to improve the accuracy of coverage probability. We prove that the log-likelihood ratio has the asymptotic distribution $4\chi^2_{p+1}$. The new empirical likelihood method avoids estimating the complicated covariance matrix in contrast to normal approximation method and empirical likelihood method developed by Zhao (2010). In the simulation study, compared to the normal approximation method, our method demonstrates better performance in the small samples. The JEL can be used to make inference for the transformation model in order to improve the efficiency of the existing EL methods.

REFERENCES

Aalen, O., Borgan, O. and Gjessing, H., Survival and Event History Analysis: A Process Point of View, *Springer*, 2008.

Adimari, G. and Chiogna, M., Jackknife empirical likelihood based confidence intervals for partial areas under ROC curves, *Statistica Sinica*, Vol. 22, 2011.

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

Andridge, R. and Little, R., A review of hot deck imputation for survey non-response, *International Statistical Review*, Vol. 78, 40-64, 2010.

An, Y., Empirical likelihood confidence interval for ROC curves with missing data, *Master thesis, Georiga State University*, 2010.

Balakrishnan, N. and Lai, C.D., Continuous bivariate distributions, *Springer*, 2009.

Baysal, E. and Staum, J., Empirical likelihood for value-at-risk and expected shortfall, *Journal of Risk*, Vol. 11, 3-32, 2008.

Bennett S., Analysis of survival data by the proportional odds model, *Stat. Med.*, Vol. 2, 273-277, 1983.

Bickel, P. J. and Freedman, D. A., Some asymptotic theory for the bootstrap, *The Annals of Statistics*, Vol. 9, 1196-1217, 1981.

Cai T., Wei L.J., Wilcox M., Semiparametric regression analysis for clustered failure time data, *Biometrika*, Vol.87, 867-878, 2000.

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

Cai et al., Variable selection for multivariate failure time data, *Biometrika*, Vol. 92, 303-316, 2005.

Chen, J., Peng, L. and Zhao, Y., Empirical likelihood based confidence intervals for copulas, *Journal of Multivariate Analysis*, Vol. 100, 137-151, 2009.

Chen K., Jin Z., Ying Z., Semiparametric analysis of transformation models with censored data, *Biometrika*, Vol. 89, 659-668, 2002.

Chen, S. and Hall, P., Smoothed empirical likelihood confidence intervals for quantiles, *Annals of Statistics*, Vol. 21, 1166-1181, 1993.

Chen S., Peng L. and Qin Y., Effects of data dimensionon empirical likelihood, *Biometrika*, Vol. 96, 711-722, 2009.

Cheng S.C., Wei L.J., Ying Z., Analysis of transformation models with censored data, *Biometrika*, Vol. 82, 835-845, 1995.

Claeskens, G., Jing, B-Y., Peng, L. and Zhou, W., Empirical likelihood confidence regions for comparison distributions and ROC curves, *Canadian J. Statist.*, Vol. 31, 173–190, 2003

Cox, D., Regression models and life tables, *J. R. Stat. Soc.*, Vol.34, 187-220, 1972.

Csörgo, M., Quantile Processes With Statistical Applications, *Society for Industrial Mathematics*, 1987.

Dehardt, J., Generalizations of the Glivenko-Cantelli Theorem, *The Annals of Mathematical Statistics*, Vol. 42, 2050-2055, 1971.

DeLong, E., DeLong, D., Daniel, L. and Clarke-Pearson, D. L., Comparing the areas under two or more correlated Receiver Operating Characteristic curves: a nonparametric approach, *Biometrics*, Vol. 44, 837-845, 1988.

DiCiccio, T., Hall, P. and Romano, J., Empirical likelihood is bartlett-correctable, *Ann. Statist.*, Vol.19, 1053-1061, 1991.

Diehl, S. and Stute, W., Kernel density and hazard function estimation in the presence of censoring, *J. Multivariate Anal.*, Vol. 25, 299-310, 1988.

Einmahl, J.H. and McKeague, I.W., Confidence tubes for multiple quantile plots via empirical likelihood, *Ann. Statist.*, Vol. 27, 1348-1367, 1999.

Fleming, T.R. and Harrington, D.P., *Counting Processes and Survival Analysis*, 1991.

Fine, J.P., Ying, Z. and Wei, L.J., On the linear transformation model with censored data,, *Biometrika*, Vol. 85, 980-986, 1998.

Fine, J.P., Analyzing competing risks data with transformation models, *J. R. Stat. Soc.* Vol.61, 817-830, 1999.

Gill, R.D., Censoring and Stochastic Integrals, *Mathematisch Centrum*, 1980.

Giné, E. and Zinn, J., Bootstrapping general empirical measures, *Annals of Probability*, Vol. 18, 851-869, 1990.

Gong, Y., Peng, L. and Qi, Y., Smoothed jackknife empirical likelihood method for ROC curve, *Journal of Multivariate Analysis*, Vol. 70, 1520-1531, 2010.

Gumbel, E.J., Bivariate exponential distributions, *Journal of the American Statistical Association*, Vol. 55, 698-707, 1960.

Hall et al., Geometric representation of high dimension, low sample size data, *J. R. Stat. Soc.*, Vol.67, 427-444, 2005.

Hanley, J. A. and McNeil, B. J., A method of comparing the areas under receiver operating characteristic curves derived from the same cases, *Radiology*, Vol. 148, 839-843, 1983.

Hess, K.R., Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions, *Statist. Med.*, Vol. 13, 1045-1062, 1994.

Hjort, N.L., McKeague, I. W. and Van Keilegom, I., Extending the scope of empirical likelihood, *Annals of Statistics*, Vol. 37, 1079-1111, 2009.

Hollander, M., McKeague, I. W., and Yang, J., Likelihood ratio-based confidence bands for survival functions, *Journal of the American Statistical Association*, Vol. 92, 215-226, 1997.

Hsieh, F. and Turnbull, B.W., Non-parametric and semi-parametric estimation of the receiver operating characteristic curve, *The Annals of Statistics*, Vol. 24, 25-40, 1996.

Kaplan, E. and Meier, P., Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc.*, Vol. 53, 457-481, 1958.

Kosorok, M., Two-sample quantile tests under general conditions, *Biometrika*, Vol. 86, 909-921, 1999.

Kosorok, M., Bootstrapping the Grenander estimator, *Institute of Mathematical Statistics*, Vol. 1, 282-292, 2008.

Kong, L., Cai, J. and Sen, P., Asymptotic results for fitting semiparametric transformation models to failure time data from case-cohort studies, *Statistica Sinica*, Vol. 16, 135-151, 2006.

Jing, B., Yuan, J., and Zhou, W., Jackknife empirical likelihood, *Journal of the American Statistical Association*, Vol. 104, 1224-1232, 2009.

Lee, A.J., U-Statistics, Theory and Practice, *CRC Press*, 1990.

Li, G. and Van Keilegom, I., Likelihood ratio confidence bands in nonparametric regression with censored data, *Scand. J. Statist.*, Vol. 29, 547-562, 2002.

Lin D.Y., Wei L.J., Ying Z., Semiparametric transformation models for point processes, *J. Amer. Statist. Assoc.* , Vol. 96, 620-628, 2001.

Linnet. Comparison of quantitative diagnostic tests: Type I error, power, and sample size, *Statistics in Medicine*, Vol. 6, 147-158, 1987.

Little, R. and Rubin, D., *Statistical Analysis with Missing Data* (John Wiley, New York).

Lloyd, C.J. The use of smoothed ROC curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association*, Vol. 93, 1356-1364, 1998.

Lloyd, C.J. and Yong, Z. Kernel estimators for the ROC curve are better than empirical. *Statistics and Probability Letters*, Vol. 44, 221-228, 1999.

Kosorok, M., Bootstrapping the Grenander estimator, *Institute of Mathematical Statistics*, Vol. 1, 282-292, 2008.

Metz, C.E., Herman J., B.A. and Shen, H., Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data, *Statistics in Medicine*, Vol. 17, 1033-1053, 1998.

Mojirsheibani, M., The Glivenko-Cantelli theorem based on data with randomly imputed missing values, *Statistics and Probability Letters*, Vol. 55, 385-396, 2001.

McKeague, I.W. and Zhao, Y., Simultaneous confidence bands for ratios of survival functions via empirical likelihood, *Statist. Probab. Lett.*, Vol. 60, 405-415, 2002.

McKeague, I. W. and Zhao, Y., Comparing distribution functions via empirical likelihood, *International Journal of Biostatistics*, Vol. 1, 2005.

McKeague, I.W. and Zhao, Y., Width-scaled confidence bands for survival functions, *Statistics and Probability Letters*, Vol. 76, 327-339, 2006.

Metz, C.E., Herman J., B.A. and H. Shen., Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data, *Statistics in Medicine*, Vol. 17, 1033-1053, 1998.

Owen, A., Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, Vol. 75, 237-249, 1988.

Owen, A., Empirical likelihood and confidence regions, *Ann. Statist.*, Vol. 18, 90-120, 1990.

Owen, A., *Empirical Likelihood*, (Chapman & Hall, New York), 2001.

Pepe, M.S., A regression modeling framework for ROC curves in medical diagnostic testing, *Biometrika*, Vol. 84, 595-608, 1997.

Pepe, M.S., The Statistical Evaluation of Medical Tests for Classification and Prediction, *Oxford University Press*, 2003.

Peng, L. and Qi, Y., Smoothed jackknife empirical likelihood method for tail copulas, *TEST*, Vol. 3, 514-536, 2010.

Pepe, M.S., A regression modeling framework for ROC curves in medical diagnostic testing, *Biometrika*, Vol. 84, 595-608, 1997.

Pettitt A.N., Inference for the linear model using a likelihood based on ranks, *J. R. Stat. Soc.*, Vol. 44, 234-243, 1982.

Qin, G. and Zhou, X., Empirical Likelihood Inference for the Area under the ROC Curve, *Biometrics*, Vol. 62, 613-622, 2006.

Qin, J. and Lawless, J., Empirical likelihood and general estimating equations, *Annals of Statistics*, Vol. 22, 300-325, 1994.

Qin, Y. and Qian, Y., Empirical likelihood confidence intervals for the differences of quantiles with missing data, *Acta Mathematicae Applicatae Sinica*, Vol. 25, 105-116, 2009.

Qin, Y. and Zhang, S., Empirical likelihood confidence intervals for differences between two datasets with missing data, *Pattern Recognition Letters*, Vol. 29, 803-812, 2008.

Qin, Y.S. and Zhao, L.C., Empirical likelihood ratio intervals for various differences of two populations, *Chinese Systems Sci. Math. Sci.*, Vol. 13, 23-30, 2000.

Sheather, S. and Marron, J., Kernel quantile estimators, *Journal of the American Statistical Association*, Vol. 85, 410-416, 1990.

Shen, J. and He, S., Empirical likelihood for the difference of two survival functions under right censorship, *Statistical and Probability Letters*, Vol. 76, 169-181, 2006.

Shen, J. and He, S., Empirical likelihood for the difference of quantiles under censorship, *Statistical Papers*, Vol. 48, 437-457, 2007.

Shen, J. and He, S., Empirical likelihood intervals for hazard and density functions under right censorship, *Annals of the Institute of Statistical Mathematics*, Vol. 60, 575-589, 2008.

Shorack, G. R. and Wellner, J. A., *Empirical Processes with Application to Statistics*, Wiley, 1986.

Silverman, B.W., Weak and strong uniform consistency of the kernel estimate of a density and its derivatives, *Ann. Statist.*, 177-184, 1978.

Swets, I.A. and Pickett, R.M., Evaluation of Diagnostic Systems: Methods from Signal Detection Theory, *New York: Academic Press*, 1982.

Thomas, D.R. and Grunkemeier, G.L., Confidence interval estimation of survival probabilities for censored data, *J. Amer. Statist. Assoc.*, Vol. 70, 865-871, 1975.

Tosteson, A.N. and Begg, C.B., A general regression methodology for ROC curve estimation, *Medical Decision Making*, Vol. 8, 205-215, 1988.

Venkatraman, E.S. and Begg, C.B., A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment, *Biometrika*, Vol. 83, 835-848, 1996.

van der Vaart, A. W., Asymptotic Statistics, *Cambridge University Press*, 1998.

van der Vaart, A. W. and Wellner, J., Weak Convergence and Empirical Processes: With Applications to Statistics, *Springer Series in Statistics*, 1996.

Veraverbeke, N., Estimation of the quantiles of the duration of old age, *Journal of Statistical Planning and Inference*, Vol. 98, 101-106, 2001.

Wang, Q.H. and Rao, J.N.K., Empirical likelihood for linear regression models under imputation for missing responses, *Canad. J. Statist.*, Vol. 29, 597-608, 1988.

Wang, Q.H. and Rao, J.N.K., Empirical likelihood-based inference in linear models with missing data, *Scand. J. Statist.* , Vol. 29, 563-576, 2002.

Wang, Z. and Chang, Y., Marker selection via maximizing the partial area under the ROC curve of linear risk scores, *Biostatistics*, Vol. 12, 369-385, 2011.

Wellner, J., Bootstrap limit theorems: a partial survey, *Elsevier Science Pulishers*, 313-329, 1992.

Wieand, S, Gail, M. and James, B. A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data, *Biometrika*, Vol. 76, 585-92, 1989.

Yang, B., Qin, G. and Qin, J., Empirical likelihood-based inferences for a low income proportion, *The Canadian Journal of Statistics*, Vol. 39, 1-16, 2011.

Yang, H. and Zhao, Y., Smoothed empirical likelihood for ROC curves with censored data, *Journal of Multivariate Analysis*, Vol. 109, 254-263, 2012.

Yau, C., Empirical likelihood confidence intervals for the difference of two quantiles with right censoring, *Georgia State University Thesis*, 2009.

Yu, W., Sun, Y. and Zheng, M., Empirical likelihood method for linear transformation models, *Annals of the Institute of Statistical Mathematics*, Vol. 63, 331-346, 2011.

Zhang, Z. and Zhao, Y., Empirical likelihood for linear transformation models with interval-censored failure time data, *Journal of Multivariate Analysis*, 2012.

Zhao Y., Semiparametric inference for transformation models via empirical likelihood, *Journal of Multivariate Analysis*, Vol. 101, 1846-1858, 2010.

Zheng, B., Statistical inference for poverty measures with relative poverty lines, *Journal of Econometrics*, Vol. 101, 337-356, 2001.

Zhou, W. and Jing, B-Y., Smoothed empirical likelihood confidence intervals for the difference of quantiles, *Statistica Sinica*, Vol. 19, 71-81, 2003.

Zhou, W. and Jing, B-Y., Adjusted empirical likelihood method for quantiles, *Annals of the Institute of Statistical Mathematics*, Vol. 55, 689-703, 2003.

Zou, K.H., Hall, W.J. and Shapiro, D.E., Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests, *Stat Med.*, Vol. 16, 2143-2156, 1997.

Zou, H., The adaptive lasso and its oracle properties, *J. Amer. Statist. Assoc*, Vol. 101, 1418-1429, 2006.

APPENDICES

Appendix A

Proof of theorems for Chapter 2

Lemma A.1. *Under conditions in Theorem 2.1, as* $n \longrightarrow \infty$*, we have*

$$\sqrt{m+n}\left\{\hat{R}_{m,n}(p) - R(p)\right\} \xrightarrow{\mathfrak{D}} N(0, \sigma_1^2(p)),$$

where $\sigma_1^2(p)$ *is defined in Theorems 2.1 and* $R(p)$ *is the true ROC curve at point* $p \in (a,b)$.

Proof. Since $R^{''}$ is continuous at $p \in (a,b)$, $R^{'}$ and $R^{''}$ are bounded in (a,b). Denote $\sigma - algebra$ $\mathcal{B}_{r_1} = \{\sigma(x_i, \delta_{xi}, i \in S_{r_1})\}$ and $\mathcal{B}_m = \{\sigma(x_i, \delta_{xi}, i = 1, ..., m)\}$. Because x_i^* are only dependent on $\mathcal{B}_{r_1}$, from Qin and Zhang (2008), we have

$$E(I(x_i^* \leq x)|\mathcal{B}_{r_1}) = E(I(x_i^* \leq x)|\mathcal{B}_m) = \frac{1}{r_1}\sum_{i \in S_{r_1}} I(x_i \leq x).$$

and

$$\begin{aligned}\sqrt{m}\{F_m(x) - F(x)\} = &\frac{\sqrt{m}}{\sqrt{r_1}}\frac{1}{\sqrt{r_1}}\sum_{i \in S_{r_1}}\{I(x_i \leq x) - F(x)\} \\ &+ \frac{\sqrt{m_1}}{\sqrt{m}}\frac{1}{\sqrt{m_1}}\sum_{i \in S_{m_1}}\{I(x_i^* \leq x) - E(I(x_i^* \leq x)|\mathcal{B}_{r_1})\}.\end{aligned}$$

Denote the first term as $V_m(x)$ and the second term as $U_m(x)$. The response rate $P_1 > 0$ assures the $r_1 \to \infty$ and $m_1 \to \infty$ when $m \to \infty$. We define the empirical distribution $F_{r1}(x) = 1/r_1 \sum_{i \in S_{r_1}} I(x_i \leq x)$ of $x_1, \ ... \ , x_m$ and x_i^*, $i \in S_{m_1}$ with the distribution function $F_{r1}(x)$. Denote $F_{r_1,m_1}(x) = 1/m_1 \sum_{i \in S_{m_1}} \{I(x_i^* \leq x)\}$. $F_{r_1,m_1}(x)$ is the empirical distribution of x_1^*, ..., $x_{m_1}^*$ and the bootstrapped version of $F_{r1}(x)$ with weighting mechanism $\mathcal{M}_{m_1}$ which is independent of $\mathcal{B}_{r_1}$ since we can rewrite that $F_{r_1,m_1}(x) = 1/m_1 \sum_{i\in S_{r_1}} \left\{M_{m_1,i}^* I(x_i \le x)\right\}$ from the equation (4.4) in Wellner (1992), where the weight, $\mathcal{M}_{m_1} = \{M_{m_1,1}^*, ..., M_{m_1,m_1}^*\}$, follows multinomial distribution. By Theorem 4.1 of Bickel and Freedman (1981), we have $\sqrt{m_1}\{F_{r_1,m_1}(x) - F_{r1}(x)\} \Longrightarrow B(F(x))$, where $B(\cdot)$ is the Brownian bridge on $[0,1]$. Hence, $E(U_m(s)U_m(t)) \xrightarrow{P} (1-P_1)\{F(\min(s,t)) - F(s)F(t)\}$ and $E(U_m(x)|\mathcal{B}_{r_1}) = 0$. By Donsker's theorem and multivariate central limit theorem from Theorem 19.3 of van der Vaart (1998), $\sqrt{r_1}\left\{F_{r1}(x) - F(x)\right\} \Longrightarrow B(F(x))$ and $B(F(x))$ is tight. $E(V_m(s)V_m(t)) \xrightarrow{P} P_1^{-1}\{F(\min(s,t)) - F(s)F(t)\}$, where $B(\cdot)$ is the Brownian bridge on $[0,1]$. Then, we consider

$$(V_m(x), U_m(x)) = \left(\frac{\sqrt{m}}{\sqrt{r_1}}\sqrt{r_1}\left\{F_{r1}(x) - F(x)\right\}, \frac{\sqrt{m_1}}{\sqrt{m}}\sqrt{m_1}\{F_{r_1,m_1}(x) - F_{r1}(x)\}\right).$$

We know that Brownian bridge $B(F(x))$ is tight and $V_m(x)$ and $U_m(x)$ marginally converge to Brownian bridge, i.e., $V_m(x) \Longrightarrow \sqrt{1-P_1}B(F(x))$ and $U_m(x) \Longrightarrow \sqrt{P_1^{-1}}B(F(x))$, respectively. From the equation (3.2) in Giné and Zinn (1990), we know that $U_m(x) \underset{W}{\overset{P}{\rightsquigarrow}} \sqrt{P_1^{-1}}B(F(x))$, where weak convergence $\underset{W}{\overset{P}{\rightsquigarrow}}$ is defined as follows by Kosorok (2008),

$$\sup_{h\in BL_1(\mathcal{F})} \|E_{\cdot|\mathcal{B}_{r_1}} h\{U_m(x)\} - Eh\{\sqrt{P_1^{-1}}B(F(x))\}\| \to 0.$$

Note $\mathcal{M}_{m_1}$ is measurable conditional on $\mathcal{B}_{r_1}$. $V_m(s)$ and $U_m(s)$ are uncorrelated since

$$E(V_m(s)U_m(s)) = E(V_m(s)E(U_m(s)|\mathcal{B}_{r_1}))) = 0.$$

By p.180 in van der Vaart and Wellner (1996) and Theorem 2.2 in Kosorok (2008),

$$(V_m(x), U_m(x)) \Longrightarrow \left(\sqrt{P_1^{-1}}\tilde{B}_1(F(x)), \sqrt{1-P_1}\tilde{B}_2(F(x))\right),$$

where $\tilde{B}_1(F(x))$ and $\tilde{B}_2(F(x))$ are independent copies of $B(F(x))$. The sequence converges jointly in distribution to two independent Brownian bridges, which implies that

$$W_1(x) = \sqrt{m}\{F_m(x) - F(x)\} = V_m(x) + U_m(x) \Longrightarrow \sqrt{1 - P_1 + P_1^{-1}}B(F(x)). \qquad \text{(A.1)}$$

Similarly, we have $W_2(y) = \sqrt{n}\{G_n(x) - G(x)\} \Longrightarrow \sqrt{1 - P_2 + P_2^{-1}}B(G(x))$.

Then, we consider the uniform convergence of empirical distribution function after hot deck imputation. Mojirsheibani (2001) derived the Glivenko-Cantelli Theorem with completely randomly missing data.

$$\sup_{x\in\mathcal{R}}|F_m(x) - F(x)| \longrightarrow 0 \quad a.s. \quad and \quad \sup_{y\in\mathcal{R}}|G_n(y) - G(y)| \longrightarrow 0 \quad a.s. \qquad \text{(A.2)}$$

We write

$$\begin{aligned}
&1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G(x_{I,j})}{h}\right) - R(p)\\
&= F(G^{-1}(1-p)) - \int_{-\infty}^{\infty} K\left(\frac{1-p-G(x)}{h}\right)\mathrm{d}F_m(x)\\
&= F(G^{-1}(1-p)) - K\left(\frac{1-p-G(x)}{h}\right)F_m(x)|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} F_m(x)\mathrm{d}K\left(\frac{1-p-G(x)}{h}\right)\\
&= F\{G^{-1}(1-p)\} - K\left(\frac{-p}{h}\right) - \int_{-\infty}^{\infty} F_m(x)w\left(\frac{1-p-G(x)}{h}\right)h^{-1}\mathrm{d}G(x)\\
&= F\{G^{-1}(1-p)\} - K\left(\frac{-p}{h}\right) - \int_{-1}^{1} F_m\{G^{-1}(1-p-xh)\}w(x)\mathrm{d}x\\
&= F\{G^{-1}(1-p)\} - F_m(G^{-1}(1-p)) - \int_{-1}^{1}\left[F_m\{G^{-1}(1-p-xh)\} - F_m\{G^{-1}(1-p)\}\right]w(x)\mathrm{d}x\\
&= F\{G^{-1}(1-p)\} - F_m\{G^{-1}(1-p)\} - \int_{-1}^{1}\left[F\{G^{-1}(1-p-xh)\} - F\{G^{-1}(1-p)\}\right]w(x)\mathrm{d}x\\
&- \int_{-1}^{1}([F_m\{G^{-1}(1-p-xh)\} - F\{G^{-1}(1-p-xh)\}]\\
&- \left[F_m\{G^{-1}(1-p)\} - F\{G^{-1}(1-p)\}\right])w(x)\mathrm{d}x.
\end{aligned}$$

Because $-p/h$ is beyond the support of kernel function K as $h \to 0$, $K(-p/h) = 0$ when

$p \in (a,b)$.

$$
\begin{aligned}
&\int_{-\infty}^{\infty}[F\{G^{-1}(1-p-xh)\}-F\{G^{-1}(1-p)\}]w(x)\mathrm{d}x \\
&= -\int_{-p/h}^{(1-p)/h} R^{'}(p)xhw(x)\mathrm{d}x - \frac{1}{2}\int_{-p/h}^{(1-p)/h} R^{''}(p^*)(xh)^2 w(x)\mathrm{d}x \\
&= -\frac{1}{2}\int_{-1}^{1} R^{''}(p^*)(xh)^2 w(x)\mathrm{d}x \\
&= O(h^2),
\end{aligned}
\tag{A.3}
$$

where p^* is between p and $p+xh$. Because $p \in (a,b)$, we have $F_m\{G^{-1}(1-p-xh)\} - F\{G^{-1}(1-p-xh)\} - m^{-1/2}\sqrt{1-P_2+P_2^{-1}}B[F\{G^{-1}(1-p-xh)\}] = o_p(m^{-1/2})$, for any $x \in [-1,1]$. Using the conditions on h and the continuity of $B_F(x)$,

$$
\begin{aligned}
&\int_{-p/h}^{(1-p)/h}(F_m\{G^{-1}(1-p-xh)\} - F\{G^{-1}(1-p-xh)\} - [F_m\{G^{-1}(1-p)\} \\
&- F\{G^{-1}(1-p)\}])w(x)\mathrm{d}x \\
&= \int_{-1}^{1} F_m\{G^{-1}(1-p-xh)\} - F\{G^{-1}(1-p-xh)\} \\
&\quad - m^{-1/2}\sqrt{1-P_2+P_2^{-1}}B[F\{G^{-1}(1-p-xh)\}]w(x)\mathrm{d}x \\
&\quad - \int_{-1}^{1}\left\{F_m\{G^{-1}(1-p)\} - F\{G^{-1}(1-p)\} - m^{-1/2}\sqrt{1-P_2+P_2^{-1}}B[F\{G^{-1}(1-p)\}]\right\}w(x)\mathrm{d}x \\
&\quad + \sqrt{1-P_2+P_2^{-1}}\int_{-1}^{1}\left\{m^{-1/2}B[F\{G^{-1}(1-p-xh)\}] - m^{-1/2}B[F\{G^{-1}(1-p)\}]\right\}w(x)\mathrm{d}x \\
&= o_p(m^{-1/2}).
\end{aligned}
\tag{A.4}
$$

Hence, by (A.1), (A.2), (A.3) and (A.4), we have

$$
\begin{aligned}
&\sqrt{m}\left[1-\frac{1}{m}\sum_{j=1}^{m}K\left\{\frac{1-p-G(x_{I,j})}{h}\right\}-R(p)\right]\\
&=\sqrt{m}[F\{G^{-1}(1-p)\}-F_m\{G^{-1}(1-p)\}]+o_p(m^{-1/2}m^{1/2})+O(m^{1/2}h^2)\\
&\xrightarrow{\mathfrak{D}} N(0,(1-P_1+P_1^{-1})F\{G^{-1}(1-p)\}\{1-F[G^{-1}(1-p)]\})\\
&=N(0,(1-P_1+P_1^{-1})R(p)\{1-R(p)\}). \qquad \text{(A.5)}
\end{aligned}
$$

Write

$$
\begin{aligned}
&\frac{\sqrt{n}}{m}\sum_{j=1}^{m}K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)-\frac{\sqrt{n}}{m}\sum_{j=1}^{m}K\left(\frac{1-p-G(x_{I,j})}{h}\right)\\
&=\int_{-\infty}^{\infty}K\left(\frac{1-p-G_n(x)}{h}\right)\mathrm{d}\sqrt{n}F_m(x)-\int_{-\infty}^{\infty}K\left(\frac{1-p-G(x)}{h}\right)\mathrm{d}\sqrt{n}F_m(x).
\end{aligned}
\qquad \text{(A.6)}
$$

Notice that

$$
\begin{aligned}
&\int_{-\infty}^{\infty}K\left(\frac{1-p-G_n(x)}{h}\right)\mathrm{d}\sqrt{m}\{F_m(x)-F(x)\}-\int_{-\infty}^{\infty}K\left(\frac{1-p-G(x)}{h}\right)\mathrm{d}\sqrt{m}\{F_m(x)-F(x)\}\\
&=K\left(\frac{1-p-G_n(x)}{h}\right)\sqrt{m}\{F_m(x)-F(x)\}\mid_{-\infty}^{\infty}-K\left(\frac{1-p-G(x)}{h}\right)\sqrt{m}\{F_m(x)-F(x)\}\mid_{-\infty}^{\infty}\\
&-\int_{-\infty}^{\infty}W_1(x)\mathrm{d}K\left(\frac{1-p-G_n(x)}{h}\right)+\int_{-\infty}^{\infty}W_1(x)\mathrm{d}K\left(\frac{1-p-G(x)}{h}\right)\\
&=\frac{1}{h}\int_{-\infty}^{\infty}W_1(x)w\left(\frac{1-p-G(x)}{h}\right)\mathrm{d}G(x)-\frac{1}{h}\int_{-\infty}^{\infty}W_1(x)w\left(\frac{1-p-G_n(x)}{h}\right)\mathrm{d}G_n(x)\\
&=\int_{-1}^{1}W_1\{G^{-1}(1-p-hu)\}w(x)\mathrm{d}u-\int_{-1}^{1}W_1\{G_n^{-1}(1-p-hu)\}w(x)\mathrm{d}u\\
&=\sqrt{1-P_2+P_2^{-1}}\int_{-1}^{1}B[F\{G^{-1}(1-p-hu)\}]-B[F\{G_n^{-1}(1-p-hu)\}]w(x)\mathrm{d}u+o_p(1)\\
&=o_p(1),
\end{aligned}
$$

because of the continuity of $B(F(x))$ and the proof in P. 1525 of Gong et al. (2010). Thus, we can adjust the (A.6) as follows

$$
\begin{aligned}
&\int_{-\infty}^{\infty} K\left(\frac{1-p-G_n(x)}{h}\right) d\sqrt{n}F_m(x) - \int_{-\infty}^{\infty} K\left(\frac{1-p-G(x)}{h}\right) \mathrm{d}\sqrt{n}F_m(x) \\
&= \sqrt{n}\int_{-\infty}^{\infty} K\left(\frac{1-p-G_n(x)}{h}\right) - K\left(\frac{1-p-G(x)}{h}\right) \mathrm{d}F(x) \\
&= \sqrt{n}\int_{-\infty}^{\infty} \left(\frac{G(x)-G_n(x)}{h}\right) w\left(\frac{1-p-G(x)}{h}\right) \mathrm{d}F(x) \\
&+ \sqrt{n}\int_{-\infty}^{\infty} \frac{1}{2}\left(\frac{G(x)-G_n(x)}{h}\right)^2 w^{'}\left(\frac{1-p-G(x)+\xi_x}{h}\right) \mathrm{d}F(x).
\end{aligned} \tag{A.7}
$$

Denote $R^{'}(p)$ as the first derivative of $R(p)$. The Brownian bridge $B_1(G(x))$ and $B_2(G(x))$ are uniformly bounded for $x \in (a, b)$. Also, we have the continuities of Brownian bridge $B(\cdot)$ and $R^{'}(p)$. Thus, $B(x)$ is uniformly bounded. We have

$$
\begin{aligned}
&\sqrt{n}\int_{-\infty}^{\infty} \frac{1}{2}\left(\frac{G(x)-G_n(x)}{h}\right)^2 w^{'}\left(\frac{1-p-G(x)+\xi_x}{h}\right) \mathrm{d}F(x) \\
=&\frac{1}{2\sqrt{n}h^2}\int_{-\infty}^{\infty} \left\{\sqrt{n}(G(x)-G_n(x))\right\}^2 w^{'}\left(\frac{1-p-G(x)+\xi_x}{h}\right) \mathrm{d}F(x) \\
=&\frac{1}{2\sqrt{n}h^2}\int_{-1}^{1} \left[W_2\{G^{-1}(1-p-uh+\xi_x)\}\right]^2 w^{'}(u)\mathrm{d}F\{G^{-1}(1-p-uh+\xi_x)\} \\
=&\frac{1}{2\sqrt{n}h}\int_{-1}^{1} \left[W_2\{G^{-1}(1-p-uh+\xi_x)\}\right]^2 w^{'}(u)R^{'}(p+uh+\xi_x)\mathrm{d}u \\
=&\frac{\sqrt{1-P_2+P_2^{-1}}}{2\sqrt{n}h}\int_{-1}^{1} \left\{B[G\{G^{-1}(1-p-uh+\xi_x)\}]\right\}^2 w^{'}(u)R^{'}(p+uh+\xi_x)\mathrm{d}u + o_p(1) \\
=&\frac{\sqrt{1-P_2+P_2^{-1}}}{2\sqrt{n}h}\int_{-1}^{1} \{B(1-p-uh+\xi_x)\}^2 w^{'}(u)R^{'}(p+uh+\xi_x)\mathrm{d}u + o_p(1) \\
=&\frac{\sqrt{1-P_2+P_2^{-1}}}{2\sqrt{n}h}\int_{-1}^{1} \{B(1-p)\}^2 w^{'}(u)R^{'}(p)\mathrm{d}u + o_p(1) \\
=&o_p(1).
\end{aligned} \tag{A.8}
$$

Recall that

$$
\begin{aligned}
&\sqrt{n}\int_{-\infty}^{\infty}\left(\frac{G(x)-G_n(x)}{h}\right)w\left(\frac{1-p-G(x)}{h}\right)\mathrm{d}F(x)\\
&=\int_{-1}^{1}W_2\{G^{-1}(1-p-hu)\}w(u)h^{-1}dF(G^{-1}(1-p-hu))\\
&=\int_{-1}^{1}\sqrt{1-P_2+P_2^{-1}}B[G\{G^{-1}(1-p-hu)\}]w(u)(R^{'}(p+uh))du+o_p(1)\\
&=\sqrt{1-P_2+P_2^{-1}}R^{'}(p)B(1-p)\int_{-1}^{1}w(u)du+o_p(1)\\
&\xrightarrow{\mathfrak{D}}N(0,(1-P_2+P_2^{-1})p(1-p)R^{'2}(p)).
\end{aligned}
\tag{A.9}
$$

Then, we have

$$
\sqrt{m+n}\{\hat{R}_{m,n}(p)-R(p)\}
$$

$$
\begin{aligned}
&=\frac{\sqrt{m+n}}{\sqrt{n}}\sqrt{n}\left\{\frac{1}{m}\sum_{j=1}^{m}K\left(\frac{1-p-G(x_{I,j})}{h}\right)-\frac{1}{m}\sum_{j=1}^{m}K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\}\\
&+\frac{\sqrt{m+n}}{\sqrt{m}}\sqrt{m}\left\{1-\frac{1}{m}\sum_{j=1}^{m}K\left(\frac{1-p-G(x_{I,j})}{h}\right)-R(p)\right\}.
\end{aligned}
$$

Combining (A.6), (A.7), (A.8) and (A.9) and the independence of first term and second term, we can obtain the conclusion as follows,

$$
\sqrt{m+n}\{\hat{R}_{m,n}(p)-R(p)\}\xrightarrow{\mathfrak{D}}N(0,\sigma_1^2(p)).
$$

□

Lemma A.2. *Under conditions in Theorem 2.1, for $p\in(a,b)$, as $n\longrightarrow\infty$, we have*

$$
\sqrt{m+n}\left\{\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i(p)-R(p)\right\}\xrightarrow{\mathfrak{D}}N(0,\sigma_1^2(p)),
$$

where $\sigma_1^2(p)$ is defined in Lemma A.1.

Proof. From the definition of $\hat{V}_i(p)$, we have

$$
\begin{aligned}
&\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i(p) \\
&= \frac{1}{m+n}\sum_{i=1}^{m+n}\{(m+n)\hat{R}_{m,n}(p) - (m+n-1)\hat{R}_{m,n,i}(p)\} \\
&= \frac{1}{m+n}\sum_{i=1}^{m}[(m+n)\left\{1 - \frac{1}{m}\sum_{j=1}^{m}K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\} \\
&\quad - (m+n-1)\left\{1 - \frac{1}{m-1}\sum_{j=1,j\neq i}^{m}K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\}] \\
&\quad + \frac{1}{m+n}\sum_{i=m+1}^{m+n}[(m+n)\left\{1 - \frac{1}{m}\sum_{j=1}^{m}K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\} \\
&\quad - (m+n-1)\left\{1 - \frac{1}{m}\sum_{j=1}^{m}K\left(\frac{1-p-G_{n,m-i}(x_{I,j})}{h}\right)\right\}]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{m+n}\sum_{i=1}^{m} 1 - \frac{m+n}{m}\left\{\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) - \sum_{j=1,j\neq i}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\} \\
&\quad + \left(\frac{m+n-1}{m-1} - \frac{m+n}{m}\right)\sum_{j=1,j\neq i}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) \\
&\quad + \frac{1}{m+n}\sum_{i=1}^{n}[1 + \frac{m+n-1}{m}\{\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,-i}(x_{I,j})}{h}\right) \\
&\quad - \sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\} + \left(\frac{m+n-1}{m} - \frac{m+n}{m}\right)\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)] \\
&= \frac{1}{m+n}(m+n - \frac{m+n}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) \\
&\quad + \frac{m+n-1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,-i}(x_{I,j})}{h}\right) - K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\}) \\
&\quad + \frac{1}{m+n}\{\left(\frac{m+n-1}{m-1} - \frac{m+n}{m}\right)(m-1) \\
&\quad + \left(\frac{m+n-1}{m} - \frac{m+n}{m}\right)n\}\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) \\
&= \frac{1}{m+n}(m+n - \frac{m+n}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) \\
&\quad + \frac{m+n-1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,-i}(x_{I,j})}{h}\right) - K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\}). \qquad \text{(A.10)}
\end{aligned}
$$

Write

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,-i}(x_{I,j})}{h}\right) - K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\} \\
&= \sum_{j=1}^{m}\left\{\sum_{i=1}^{n}\frac{G_{n,-i}(x_{I,j}) - G_n(x_{I,j})}{h}\right\} w\left(\frac{1-p-G_n(x_{I,j})}{h}\right) \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{2}\left\{\frac{G_{n,-i}(x_{I,j}) - G_n(x_{I,j})}{h}\right\}^2 w^{'}\left(\frac{1-p-\xi_{n,i,j}}{h}\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{2}\left\{\frac{G_{n,-i}(x_{I,j}) - G_n(x_{I,j})}{h}\right\}^2 w^{'}\left(\frac{1-p-\xi_{n,i,j}}{h}\right), \qquad \text{(A.11)}
\end{aligned}
$$

where $\xi_{n,i,j}$ is between the $G_n(x_{I,j})$ and $G_{n,-i}(x_{I,j})$,

$$G_n(x_{I,j}) - G_{n,-i}(x_{I,j}) = \frac{1}{n-1}\{G_n(x_{I,j}) - I(Y_{I,i} \leq x_{I,j})\} = O_p\left(\frac{1}{n-1}\right), \tag{A.12}$$

and

$$\sum_{i=1}^{n}\{G_{n,-i}(x_{I,j}) - G_n(x_{I,j})\} = 0,$$

because

$$\begin{aligned}
&G_n(x_{I,j}) - G_{n,-i}(x_{I,j}) \\
&= \frac{1}{n}\sum_{k=1}^{n} I(y_{I,i} \leq x_{I,j}) - \frac{1}{n-1}\sum_{i=k,k\neq i}^{n} I(y_{I,i} \leq x_{I,j}) \\
&= \left(\frac{1}{n} - \frac{1}{n-1}\right)\sum_{k=1}^{n} I(y_{I,i} \leq x_{I,j}) - \frac{1}{n-1}\{\sum_{k=1}^{n} I(y_{I,i} \leq x_{I,j}) - \sum_{i=k,k\neq i}^{n} I(y_{I,i} \leq x_{I,j})\} \\
&= \frac{1}{n-1}\{G_n(x_{I,j}) - I(y_{I,i} \leq x_{I,j})\}.
\end{aligned}$$

By similar steps in (A.11) and (A.12), we have

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,-i}(x_{I,j})}{h}\right) - K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\} = O_p\left(\frac{mn}{(n-1)^2 h}\right). \tag{A.13}$$

Combining (A.10), (A.13) and Lemma A.1, we have

$$\begin{aligned}
&\sqrt{m+n}\left\{\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i(p) - R(p)\right\} \\
&= \sqrt{m+n}\left\{1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) + O_p\left(\frac{m+n-1}{(m+n)m}\frac{mn}{h(n-1)^2}\right) - R(p)\right\} \\
&= \sqrt{m+n}\left\{\hat{R}_{m,n}(p) - R(p) + O_p\left(\frac{(m+n-1)n}{(m+n)(n-1)^2 h}\right)\right\} \\
&\xrightarrow{\mathfrak{D}} N(0, \sigma_1^2(p)).
\end{aligned}$$

$\square$

Lemma A.3. *Under conditions in Theorem 2.1, for $p \in (a,b)$, as $n \longrightarrow \infty$, we have*

$$\frac{1}{m+n}\sum_{i=1}^{m+n}\left\{\hat{V}_i(p) - R(p)\right\}^2 \xrightarrow{\mathfrak{D}} \sigma_2^2(p),$$

where $\sigma_2^2(p)$ is defined in Theorem 2.2.

Proof. For $1 \le i \le m$,

$$\begin{aligned}
\hat{V}_i(p) &= 1 - \frac{m+n}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) + \frac{m+n-1}{m-1}\sum_{j=1, j\neq i}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) \\
&= 1 + \left(\frac{m+n-1}{m-1} - \frac{m+n}{m}\right)\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) \\
&\quad - \frac{m+n-1}{m-1}\left\{\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) - \sum_{j=1, j\neq i}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\} \\
&= 1 + \frac{n}{(m-1)m}\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) - \frac{m+n-1}{m-1} K\left(\frac{1-p-G_n(x_{I,i})}{h}\right),
\end{aligned}$$

and

$$\begin{aligned}
\hat{V}_i^2(p) &= \left\{1 - \frac{m+n-1}{m-1} K\left(\frac{1-p-G_n(x_{I,i})}{h}\right)\right\}^2 + \left\{\frac{n}{(m-1)m}\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\}^2 \\
&\quad + 2\left[1 - \frac{m+n-1}{m-1} K\left(\frac{1-p-G_n(x_{I,i})}{h}\right)\left\{\frac{n}{(m-1)m}\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\}\right],
\end{aligned}$$

which implies that

$$\sum_{i=1}^{m} \hat{V}_i^2(p) = m + \frac{(m+n-1)^2}{(m-1)^2} \sum_{i=1}^{m} K^2 \left(\frac{1-p-G_n(x_{I,i})}{h} \right)$$
$$- \frac{2(m+n-1)}{m-1} \sum_{j=1}^{m} K \left(\frac{1-p-G_n(x_{I,i})}{h} \right) + m \left\{ \frac{n}{(m-1)m} \Sigma_{i=1}^{m} K \left(\frac{1-p-G_n(x_{I,j})}{h} \right) \right\}^2$$
$$+ \frac{2n}{(m-1)m} \left\{ m - \frac{m+n-1}{m-1} \sum_{j=1}^{m} K \left(\frac{1-p-G_n(x_{I,i})}{h} \right) \right\} \left\{ \sum_{j=1}^{m} K \left(\frac{1-p-G_n(x_{I,j})}{h} \right) \right\}. \tag{A.14}$$

Since K^2 is a distribution function, from Gong et al. (2010) and (A.5), we have that

$$\frac{1}{m} \sum_{i=1}^{m} K^2 \left(\frac{1-p-G_n(x_{I,i})}{h} \right) \xrightarrow{\mathcal{P}} F\{G^{-1}(1-p)\}.$$

Hence, by (A.14) and Lemma A.1,

$$\frac{1}{m+n} \sum_{i=1}^{m} \hat{V}_i^2(p) = \frac{m}{m+n} + \frac{(m+n-1)^2}{(m+n)(m-1)^2} \sum_{i=1}^{m} K^2 \left(\frac{1-p-G_n(x_{I,i})}{h} \right)$$
$$- \frac{2(m+n-1)}{(m-1)(m+n)} \sum_{i=1}^{m} K \left(\frac{1-p-G_n(x_{I,i})}{h} \right)$$
$$+ \frac{n^2}{(m-1)^2 m(m+n)} \left\{ \sum_{j=1}^{m} K \left(\frac{1-p-G_n(x_{I,j})}{h} \right) \right\}^2$$
$$+ \frac{2n}{(m+n)(m-1)m} \left\{ m - \frac{m+n-1}{m-1} \sum_{i=1}^{m} K \left(\frac{1-p-G_n(x_{I,i})}{h} \right) \right\}$$
$$\left\{ \sum_{j=1}^{m} K \left(\frac{1-p-G_n(x_{I,j})}{h} \right) \right\}$$
$$\xrightarrow{\mathfrak{D}} \frac{r}{r+1} - 2F\{G^{-1}(1-p)\} + \frac{r+1}{r} F\{G^{-1}(1-p)\} + \frac{1}{r(r+1)} [F\{G^{-1}(1-p)\}]^2$$
$$+ \frac{2}{r+1} F\{G^{-1}(1-p)\}[1 - \frac{r+1}{r} F\{G^{-1}(1-p)\}]$$
$$= \frac{r+1}{r} R(p) - \frac{2r+1}{r(r+1)} R^2(p). \tag{A.15}$$

Next, for $m+1 \leq i \leq m+n$, we can write that

$$\hat{V}_i(p) = 1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right) + \frac{m+n-1}{m}\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,m-i}(x_{I,j})}{h}\right) - K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\},$$

and

$$\hat{V}_i^2(p) = \left\{1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\}^2$$

$$\begin{aligned}
&+ \left[\frac{m+n-1}{m}\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,m-i}(x_{I,j})}{h}\right) - K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\}\right]^2 \\
&+ 2\left\{1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\} \\
&\left[\frac{m+n-1}{m}\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,m-i}(x_{I,j})}{h}\right) - K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\}\right] \\
&= I_i(p) + II_i(p) + III_i(p), \qquad \text{(A.16)}
\end{aligned}$$

By (A.13), we have

$$\frac{1}{m+n}\sum_{i=m+1}^{m+n} III_i(p) = O_p((nh)^{-1}). \qquad \text{(A.17)}$$

Define $A_i = \left[\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,-i}(x_{I,j})}{h}\right) - K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\}\right]^2$. By (A.13), we have

$$\begin{aligned}
A_i &= \left[\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,-i}(x_{I,j})}{h}\right) - K\left(\frac{1-p-G_n(x_{I,j})}{h}\right)\right\}\right]^2 \\
&= \left\{\int_{-\infty}^{\infty} mK\left(\frac{1-p-G_{n,-i}(x)}{h}\right) dF_m(x) - \int_{-\infty}^{\infty} mK\left(\frac{1-p-G_n(x)}{h}\right) \mathrm{d}F_m(x)\right\}^2 \\
&= \{m\int_{-\infty}^{\infty}\left(\frac{G_{n,-i}(x)-G_n(x)}{h}\right) w\left(\frac{1-p-G_n(x)}{h}\right)\mathrm{d}F_m(x) \\
&+ m\int_{-\infty}^{\infty}\frac{1}{2}\left(\frac{G_{n,-i}(x)-G_n(x)}{h}\right)^2 w^{'}\left(\frac{1-p-G_n(x)+\xi}{h}\right)\mathrm{d}F_m(x)\}^2 + o_p(1) \\
&= \left\{\int_{-\infty}^{\infty} m\left(\frac{G_{n,-i}(x)-G_n(x)}{h}\right) w\left(\frac{1-p-G_n(x)}{h}\right)\mathrm{d}F_m(x)\right\}^2 + o_p(1).
\end{aligned}$$

By (A.1), (A.2), the continuity of $R^{'}$, and Assumptions A.4 and A.5

$$\begin{aligned}
\frac{1}{m+n}\sum_{i=1}^{n} A_i = \frac{m^2}{m+n}\sum_{i=1}^{n}\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{G_{n,-i}(x_1)-G_n(x_1)}{h}\right)\left(\frac{G_{n,-i}(x_2)-G_n(x_2)}{h}\right) \\
w\left(\frac{1-p-G_n(x_1)}{h}\right) w\left(\frac{1-p-G_n(x_2)}{h}\right)\mathrm{d}F_m(x_1)\mathrm{d}F_m(x_2)\} + o_p(1)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{m^2}{(m+n)(n-1)^2h^2} \sum_{i=1}^{n} [\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{G_n(x_1) - I(Y_{I,i} \le x_1)\}\{G_n(x_2) - I(Y_{I,i} \le x_2)\} \\
&\quad w\left(\frac{1-p-G_n(x_1)}{h}\right) w\left(\frac{1-p-G_n(x_2)}{h}\right) \mathrm{d}F_m(x_1)\mathrm{d}F_m(x_2)] + o_p(1) \\
&= \frac{m^2}{(m+n)(n-1)^2h^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{\sum_{i=1}^{n} \{G_n(x_1)G_n(x_2) + I(Y_{I,i} \le x_1)I(Y_{I,i} \le x_2) \\
&\quad - I(Y_{I,i} \le x_1)G_n(x_2) - G_n(x_1)I(Y_{I,i} \le x_2)\} \\
&\quad w\left(\frac{1-p-G_n(x_1)}{h}\right) w\left(\frac{1-p-G_n(x_2)}{h}\right) \mathrm{d}F_m(x_1)\mathrm{d}F_m(x_2)\} + o_p(1) \\
&= \frac{nm^2}{(m+n)(n-1)^2h^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\{G_n(x_1 \wedge x_2) - G_n(x_1)G_n(x_2)\} \\
&\quad w\left(\frac{1-p-G_n(x_1)}{h}\right) w\left(\frac{1-p-G_n(x_2)}{h}\right) \mathrm{d}F_m(x_1)\mathrm{d}F_m(x_2)] + o_p(1) \\
&= \frac{nm^2}{(m+n)(n-1)^2h^2} \int_{-1}^{1}\int_{-1}^{1} [G_n\{F_m^{-1}(v_1) \wedge F_m^{-1}(v_2)\} - G_n\{F_m^{-1}(v_1)\}G_n\{F_m^{-1}(v_2)\}] \\
&\quad w\left(\frac{1-p-G_n(F_m^{-1}(v_1))}{h}\right) w\left(\frac{1-p-G_n(F_m^{-1}(v_2))}{h}\right) \mathrm{d}v_1\mathrm{d}v_2 + o_p(1).
\end{aligned}
$$

From the proof of Lemma A.1 of Gong et al. (2010), the above equation is

$$
\begin{aligned}
&= \frac{nm^2}{(m+n)(n-1)^2h^2} \int_{-1}^{1}\int_{-1}^{1} [G\{F^{-1}(v_1) \wedge F^{-1}(v_2)\} - G\{F^{-1}(v_1)\}G\{F^{-1}(v_2)\}] \\
&\quad w\left(\frac{1-p-G\{F^{-1}(v_1)\}}{h}\right) w\left(\frac{1-p-G\{F^{-1}(v_2)\}}{h}\right) \mathrm{d}v_1\mathrm{d}v_2 + o_p(1) \\
&= \frac{nm^2}{(m+n)(n-1)^2} \int_{-1}^{1}\int_{-1}^{1} [G\{G^{-1}(1-p-hu_1) \wedge G^{-1}(1-p-hu_2)\} - G(G^{-1}(1-p-hu_2)) \\
&\quad G_n\{G^{-1}(1-p-hu_2)\}]w(u_1)w(u_2)\mathrm{d}F\{G^{-1}(1-p-hu_2)\}\mathrm{d}F\{G^{-1}(1-p-hu_1)\} + o_p(1) \\
&= \frac{nm^2}{(m+n)(n-1)^2} \int_{-1}^{1}\int_{-1}^{1} [G\{G^{-1}(1-p-hu_1) \wedge G^{-1}(1-p-hu_2)\} \\
&\quad - G\{G^{-1}(1-p-u_2)\}G_n\{G^{-1}(1-p-u_2)\}] \quad w(u_1)w(u_2)R^{'}(p+hu_2)R^{'}(p+hu_1)\mathrm{d}(u_2)\mathrm{d}(u_1) + o_p(1) \\
&= \frac{nm^2}{(m+n)(n-1)^2} \int_{-1}^{1}\int_{-1}^{1} [\{(1-p-hu_1) \wedge (1-p-hu_2) - (1-p-hu_1)(1-p-hu_2)\} \\
&w(u_1)w(u_2)R'(p)R^{'}(p)]\mathrm{d}(u_2)\mathrm{d}(u_1) + o_p(1) \\
&= \frac{nm^2}{(m+n)(n-1)^2}\{(1-p) \wedge (1-p) - (1-p)^2\}R^{'2}(p) \int_{-1}^{1} w(u_1)\mathrm{d}(u_1) \int_{-1}^{1} w(u_2)\mathrm{d}(u_2) + o_p(1)
\end{aligned}
$$

$$= \frac{nm^2}{(m+n)(n-1)^2}\{(1-p)-(1-p)^2\}R'^2(p)+o_p(1)$$

$$\xrightarrow{P} \frac{r^2}{(1+r)}p(1-p)R'^2(p). \tag{A.18}$$

By (A.16), (A.17), (A.18) and Lemma A.1, we have

$$\frac{1}{m+n}\sum_{i=m+1}^{m+n}\hat{V}_i^2(p) \xrightarrow{\mathcal{P}} \frac{1}{1+r}R^2(p)+(r+1)p(1-p)R'^2(p). \tag{A.19}$$

Hence, it follows from (A.15), (A.19) and Lemma A.2 that

$$\begin{aligned}
&\frac{1}{m+n}\sum_{i=1}^{m+n}\{\hat{V}_i(p)-R(p)\}^2 \\
&= \frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i^2(p)+R^2(p)-\frac{2}{m+n}R(p)\sum_{i=1}^{m+n}\hat{V}_i(p) \\
&\xrightarrow{P} \frac{1+r}{r}R(p)-\frac{2r+1}{r(r+1)}R^2(p)+\frac{1}{1+r}R^2(p)+(r+1)p(1-p)R'^2(p)-2R^2(p)+R^2(p) \\
&= \left(1+\frac{1}{r}\right)R(p)\{1-R(p)\}+(r+1)p(1-p)R'^2(p) \\
&= \sigma_2^2(p).
\end{aligned}$$

□

Proof of Theorem 2.1 It follows directly from Lemmas A.2 and A.3. □

Proof of Theorem 2.2 Throughout let $\theta = R(p)$. Recall $1/(m+n)\sum_{i=1}^{m+n}(\hat{V}_i(p)-\theta)/\{1+\lambda(\hat{V}_i(p)-\theta)\}=0$. Define $\gamma_i=\lambda\{\hat{V}_i(p)-\theta\}$. Following similar steps as Gong et al. (2010), we have

$$|\lambda| = O_p((m+n)^{-1/2}), \tag{A.20}$$

and

$$\max_{1\le i\le m+n}|\gamma_i| = o_p(1). \tag{A.21}$$

Using (A.20) and (A.21), we have

$$0 = \frac{1}{m+n} \sum_{i=1}^{m+n} \frac{\hat{V}_i(p) - \theta}{1+\gamma_i} = \frac{1}{m+n} \sum_{i=1}^{m+n} \{\hat{V}_i(p) - \theta\} - S_{m+n}\lambda + O_p((m+n)^{-1}),$$

which implies that

$$\lambda = S_{m+n}^{-1} \frac{1}{m+n} \sum_{i=1}^{m+n} \{\hat{V}_i(p) - \theta\} + O_p((m+n)^{-1}), \tag{A.22}$$

where $S_{m+n} = 1/(m+n) \sum_{i=1}^{m+n} \{\hat{V}_i(p) - \theta\}^2$. Using Taylor expansion, (A.20), (A.22), Lemma A.1 and Lemma A.2, we have

$$\begin{aligned} l(R(p), p) &= 2 \sum_{i=1}^{m+n} \log(1+\gamma_i) \\ &= \frac{(\sqrt{m+n}[\frac{1}{m+n} \sum_{i=1}^{m+n} \{\hat{V}_i(p) - \theta\}])^2}{\frac{1}{m+n} \sum_{i=1}^{m+n} \{\hat{V}_i(p) - \theta\}^2} + o_p(1) \\ &\xrightarrow{\mathfrak{D}} \frac{\sigma_1^2(p)\chi_1^2}{\sigma_2^2(p)}. \end{aligned}$$

$\square$

Appendix B

Proof of Theorems for Chapter 3

Lemma B.1. *Under the same assumptions as Theorem 3.1, for $p \in (a,b)$, we have*

$$\sqrt{m+n}\{\hat{\Delta}_{m,n}(p) - \Delta(p)\} \xrightarrow{\mathfrak{D}} N(0, \sigma^2(p)), \tag{B.1}$$

where $\sigma^2(p)$ is defined in Theorem 3.1.

Proof. Denote the empirical processes, $W_{x1}(x) = \sqrt{m}\{F_{m,1}(x) - F_1(x)\}$, $W_{x2}(x) = \sqrt{m}\{F_{m,2}(x) - F_2(x)\}$, $W_{y1}(y) = \sqrt{n}\{G_{n,1}(y) - G_1(y)\}$ and $W_{y2}(y) = \sqrt{n}\{G_{n,2}(y) - G_2(y)\}$.

By Donsker's theorem and the multivariate central limit theorem from p. 266 of Van der Vaart (2000), we have

$$\begin{aligned}
Cov\,\{W_{x1}(s), W_{x1}(t)\} &\xrightarrow{P} F_1\{\min(s,t)\} - F_1(s)F_1(t),\\
Cov\,\{W_{x2}(s), W_{x2}(t)\} &\xrightarrow{P} F_2\{\min(s,t)\} - F_2(s)F_2(t),\\
Cov\,\{W_{y1}(s), W_{y1}(t)\} &\xrightarrow{P} G_1\{\min(s,t)\} - G_1(s)G_1(t),\\
Cov\,\{W_{y2}(s), W_{y2}(t)\} &\xrightarrow{P} G_2\{\min(s,t)\} - G_2(s)G_2(t).
\end{aligned}$$

From the Glivenko-Cantelli theorem, we have

$$\begin{aligned}
&\sup_{x \in \mathcal{R}} |F_{m,1}(x) - F_1(x)| \longrightarrow 0 \quad a.s.,\\
&\sup_{x \in \mathcal{R}} |F_{m,2}(x) - F_2(x)| \longrightarrow 0 \quad a.s.,\\
&\sup_{y \in \mathcal{R}} |G_{n,1}(y) - G_1(y)| \longrightarrow 0 \quad a.s.,\\
&\sup_{y \in \mathcal{R}} |G_{n,2}(y) - G_2(y)| \longrightarrow 0 \quad a.s.
\end{aligned}$$

and

$$\sup_{x_1,x_2\in\mathcal{R}} |F_m(x_1,x_2) - F(x_1,x_2)| \longrightarrow 0 \quad a.s.,$$

$$\sup_{y_1,y_2\in\mathcal{R}} |G_n(y_1,y_2) - G(y_1,y_2)| \longrightarrow 0 \quad a.s.,$$

which are the generalization of the Glivenko-Cantelli theorem from the Corollary of Dehardt (1971) in p. 2055. Also, we have the Glivenko-Cantelli theorem for quantile process by the Corollary 1.4.1 of Csörgo (1987),

$$\begin{aligned}
\sup_{x\in[0,1]} |F_{m,1}^{-1}(x) - F_1^{-1}(x)| &\longrightarrow 0 \quad a.s.,\\
\sup_{x\in[0,1]} |F_{m,2}^{-1}(x) - F_2^{-1}(x)| &\longrightarrow 0 \quad a.s.,\\
\sup_{y\in[0,1]} |G_{n,1}^{-1}(y) - G_1^{-1}(y)| &\longrightarrow 0 \quad a.s.,\\
\sup_{y\in[0,1]} |G_{n,2}^{-1}(y) - G_2^{-1}(y)| &\longrightarrow 0 \quad a.s.
\end{aligned}$$

Thus, $\sup_{p\in[0,1]} |1 - F_{m,1}\{G_{n,1}^{-1}(1-p)\} - R(p)| = \sup_{p\in[0,1]} |F_{m,1}\{G_{n,1}^{-1}(1-p)\} - F_1\{G_1^{-1}(1-p)\}| \longrightarrow 0 \;\; a.s.$ Then, we split the difference of two ROC curves into the following components,

$$\hat{\Delta}_{m,n}(p) - \Delta(p) = \{\hat{R}_{m,n,1}(p) - R_1(p)\} - \{\hat{R}_{m,n,2}(p) - R_2(p)\}. \tag{B.2}$$

For the ROC curve,

$$\begin{aligned}
\hat{R}_{m,n,1}(p) - R_1(p) = {} & \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_1(X_{1,j})}{h}\right) - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)\\
& + 1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_1(X_{1,j})}{h}\right) - R_1(p).
\end{aligned} \tag{B.3}$$

By Lemma 1 of Gong et al. (2010), we can obtain

$$
\begin{aligned}
& 1-\frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_1(X_{1,j})}{h}\right)-R_1(p) \\
&= F_1\{G_1^{-1}(1-p)\}-F_{m,1}\{G_1^{-1}(1-p)\}+o_p(m^{-1/2})-K\left(-\frac{p}{h}\right)+O(h^2) \\
&= F_1\{G_1^{-1}(1-p)\}-F_{m,1}\{G_1^{-1}(1-p)\}+o_p(m^{-1/2}). \qquad \text{(B.4)}
\end{aligned}
$$

By Gong et al. (2010), we have

$$
\begin{aligned}
& \sqrt{n}\left\{\frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)-\frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_1(X_{1,j})}{h}\right)\right\} \\
&= \int_{-\infty}^{\infty} K\left(\frac{1-p-G_{n,1}(x)}{h}\right)\mathrm{d}\sqrt{n}F_{m,1}(x)-\int_{-\infty}^{\infty} K\left(\frac{1-p-G_1(x)}{h}\right)\mathrm{d}\sqrt{n}F_{m,1}(x). \qquad \text{(B.5)}
\end{aligned}
$$

Notice that

$$
\begin{aligned}
& \frac{\sqrt{n}}{\sqrt{m}}\int_{-\infty}^{\infty}\left\{K\left(\frac{1-p-G_{n,1}(x)}{h}\right)-K\left(\frac{1-p-G_1(x)}{h}\right)\right\}\mathrm{d}\sqrt{m}\{F_{m,1}(x)-F_1(x)\} \\
&= \frac{\sqrt{n}}{\sqrt{m}}\left\{K\left(\frac{1-p-G_{n,1}(x)}{h}\right)-K\left(\frac{1-p-G_1(x)}{h}\right)\right\}\sqrt{m}\{F_{m,1}(x)-F_1(x)\}\,|_{-\infty}^{\infty} \\
&- \frac{\sqrt{n}}{\sqrt{m}}\int_{-\infty}^{\infty} W_{x1}(x)\mathrm{d}K\left(\frac{1-p-G_{n,1}(x)}{h}\right)+\frac{\sqrt{n}}{\sqrt{m}}\int_{-\infty}^{\infty} W_{x1}(x)\mathrm{d}K\left(\frac{1-p-G_1(x)}{h}\right) \\
&= \frac{\sqrt{n}}{h\sqrt{m}}\int_{-\infty}^{\infty} W_{x1}(x)w\left(\frac{1-p-G_1(x)}{h}\right)\mathrm{d}G_1(x)-\frac{\sqrt{n}}{h\sqrt{m}}\int_{-\infty}^{\infty} W_{x1}(x)w\left(\frac{1-p-G_{n,1}(x)}{h}\right)\mathrm{d}G_{n,1}(x) \\
&= \frac{\sqrt{n}}{\sqrt{m}}\int_{-1}^{1} W_{x1}\{G_1^{-1}(1-p-hu)\}w(u)\mathrm{d}u-\frac{\sqrt{n}}{\sqrt{m}}\int_{-1}^{1} W_{x1}\{G_{n,1}^{-1}(1-p-hu)\}w(u)\mathrm{d}u \\
&= \frac{\sqrt{n}}{\sqrt{m}}\int_{-1}^{1} B[F_1\{G_1^{-1}(1-p-hu)\}]-B[F_1\{G_{n,1}^{-1}(1-p-hu)\}]w(u)\mathrm{d}u+o_p(1) \\
&= o_p(1),
\end{aligned}
$$

because of the continuity of $B(F_1(x))$ and the proof in p. 1525, Gong et al. (2010). Thus, we can adjust the (B.5) as follows

$$
\begin{aligned}
&\int_{-\infty}^{\infty} K\left(\frac{1-p-G_{n,1}(x)}{h}\right) d\sqrt{n}F_{m,1}(x) - \int_{-\infty}^{\infty} K\left(\frac{1-p-G_1(x)}{h}\right) \mathrm{d}\sqrt{n}F_{m,1}(x)\\
&= \sqrt{n}\int_{-\infty}^{\infty} K\left(\frac{1-p-G_{n,1}(x)}{h}\right) - K\left(\frac{1-p-G_1(x)}{h}\right)\mathrm{d}F_1(x)\\
&= \sqrt{n}\int_{-\infty}^{\infty} \left(\frac{G_1(x)-G_{n,1}(x)}{h}\right) w\left(\frac{1-p-G_{n,1}(x)}{h}\right)\mathrm{d}F_1(x)\\
&+ \sqrt{n}\int_{-\infty}^{\infty} \frac{1}{2}\left(\frac{G_1(x)-G_{n,1}(x)}{h}\right)^2 w^{'}\left(\frac{1-p-G_1(x)+\xi_x}{h}\right)\mathrm{d}F_1(x)\\
&= I + II,
\end{aligned} \tag{B.6}
$$

where ξ_x is between $G_1(x)$ and $G_{1,n}(x)$. Recall that

$$
\begin{aligned}
I =&\sqrt{n}\int_{-\infty}^{\infty} \frac{1}{2}\left(\frac{G_1(x)-G_{n,1}(x)}{h}\right)^2 w^{'}\left(\frac{1-p-G_1(x)+\xi_x}{h}\right)\mathrm{d}F_1(x)\\
=&\frac{1}{2\sqrt{n}h^2}\int_{-\infty}^{\infty}\left[\sqrt{n}\{G_1(x)-G_{n,1}(x)\}\right]^2 w^{'}\left(\frac{1-p-G_1(x)+\xi_x}{h}\right)\mathrm{d}F_1(x)\\
=&\frac{1}{2\sqrt{n}h^2}\int_{-1}^{1}\left[W_{y1}\{G_1^{-1}(1-p-uh+\xi_x)\}\right]^2 w^{'}(u)\mathrm{d}F_1\{G_1^{-1}(1-p-uh+\xi_x)\}\\
=&\frac{1}{2\sqrt{n}h}\int_{-1}^{1}\left[W_{y1}\{G_1^{-1}(1-p-uh+\xi_x)\}\right]^2 w^{'}(u)\{-R_1^{'}(p+uh+\xi_x)\}\mathrm{d}u\\
=&\frac{1}{2\sqrt{n}h}\int_{-1}^{1}\left\{B(G_1(G_1^{-1}(1-p-uh+\xi_x)))\right\}^2 w^{'}(u)\{-R_1^{'}(p+uh+\xi_x)\}\mathrm{d}u + o_p(1)\\
=&\frac{1}{2\sqrt{n}h}\int_{-1}^{1}\{B(1-p-uh+\xi_x)\}^2 w^{'}(u)\{-R_1^{'}(p+uh+\xi_x)\}\mathrm{d}u + o_p(1)\\
=&o_p(1),
\end{aligned}
$$

because of the continuity and boundness of $w^{'}(x)$ and $R^{'}(x)$. From (B.6), we have

$$\begin{aligned}
I + II &= \int_{-\infty}^{\infty} W_{y1}(x_1)h^{-1}w\left(\frac{1-p-G_1(x_1)}{h}\right)\mathrm{d}F_1(x_1) + o_p(1) \\
&= \int_{-1}^{1} W_{y1}\{G_1^{-1}(1-p-u_1h)\}h^{-1}w(u_1)\mathrm{d}F_1\{G_1^{-1}(1-p-u_1h)\} + o_p(1) \\
&= \int_{-1}^{1} W_{y1}\{G_1^{-1}(1-p-u_1h)\}w(u_1)R_1^{'}(p+u_1h)du_1 + +o_p(1) \\
&= \int_{-1}^{1} W_{y1}\{G_{n,1}^{-1}(1-p)\}w(u_1)R_1^{'}(p)du_1 + o_p(1) \\
&= \sqrt{n}\{G_1(G_{n,1}^{-1}(1-p)) - (1-p)\}R_1^{'}(p) + o_p(1). \qquad \text{(B.7)}
\end{aligned}$$

Hence, from (B.3)-(B.7), we have

$$\begin{aligned}
\sqrt{n+m}\left\{\hat{R}_{m,n,1}(p) - R_1(p)\right\} &= \frac{\sqrt{n+m}}{\sqrt{m}}\sqrt{m}\left[F_1\{G_1^{-1}(1-p)\} - F_{m,1}\{G_1^{-1}(1-p)\}\right] \\
&+ \frac{\sqrt{n+m}}{\sqrt{n}}\sqrt{n}\left([(1-p) - G_1\{G_{n,1}^{-1}(1-p)\}]R_1^{'}(p)\right) + o_p(1). \qquad \text{(B.8)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sqrt{n+m}\left\{\hat{R}_{m,n,2}(p) - R_2(p)\right\} &= \frac{\sqrt{n+m}}{\sqrt{m}}\sqrt{m}\left[F_2\{G_2^{-1}(1-p)\} - F_{m,2}\{G_2^{-1}(1-p)\}\right] \\
&+ \frac{\sqrt{n+m}}{\sqrt{n}}\sqrt{n}[(1-p) - G_2\{G_{n,2}^{-1}(1-p)\}]R_2^{'}(p) + o_p(1). \qquad \text{(B.9)}
\end{aligned}$$

Finally, (B.2) re-expressed as follows

$$\begin{aligned}
&\sqrt{m+n}\{\hat{\Delta}_{m,n}(p) - \Delta(p)\} \\
&= \sqrt{m+n}[F_{m,2}\{G_2^{-1}(1-p)\} - F_2\{G_2^{-1}(1-p)\}] - \sqrt{m+n}[F_{m,1}\{G_1^{-1}(1-p)\} - F_1\{G_1^{-1}(1-p)\}] \\
&+ \sqrt{m+n}R_2^{'}(p)[G_2\{G_{n,2}^{-1}(1-p)\} - (1-p)] - \sqrt{m+n}[G_1\{G_{n,1}^{-1}(1-p)\} - (1-p)]R_1^{'}(p) \\
&+ o_p(1). \qquad \text{(B.10)}
\end{aligned}$$

Wieand et al. (1989) presented the asymptotic normality of nonparametric estimation for the difference of two ROC curves in the area or at one point. By the proof of Theorem 3.1 in Wieand et al. (1989) and Shorack et al. (1986), the variance of (B.10) can be obtained as $\sigma^2(p)$. Thus, we have

$$\sqrt{m+n}\{\hat{\Delta}_{m,n}(p) - \Delta(p)\} \xrightarrow{\mathfrak{D}} N(0, \sigma^2(p)).$$

$\square$

Lemma B.2. *Under conditions in Theorem 3.1, for any $p \in (a, b)$, we have*

$$\sqrt{m+n}\left\{\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i(p) - \Delta(p)\right\} \xrightarrow{\mathfrak{D}} N(0, \sigma^2(p)),$$

where $\sigma^2(p)$ is defined in Theorem 3.1.

Proof. By the definition

$$\begin{aligned}
&\hat{V}_i(p) = (m+n)\hat{\Delta}_{m,n}(p) - (m+n-1)\hat{\Delta}_{m,n,i}(p), \quad i = 1, \ldots, m+n,\\
&\hat{\Delta}_{m,n}(p) = \hat{R}_{m,n,1}(p) - \hat{R}_{m,n,2}(p),\\
&\hat{\Delta}_{m,n,i}(p) = \hat{R}_{m,n,1,i}(p) - \hat{R}_{m,n,2,i}(p), \quad i = 1, \ldots, m+n,
\end{aligned}$$

where

$$\hat{R}_{m,n,k,i}(p) = \frac{1}{m-1}\sum_{1\le j\le m, j\ne i} K\left(\frac{1-p-G_{n,k}(X_{k,j})}{h}\right), 1 \le i \le m,$$

and

$$\hat{R}_{m,n,k,i}(p) = \frac{1}{m-1}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,m-i,k}(X_{k,j})}{h}\right), m+1 \le i \le m+n, k = 1, 2.$$

Define that

$$\begin{aligned}
&\hat{V}_{1,i}(p) = (m+n)\hat{R}_{m,n,1}(p) - (m+n-1)\hat{R}_{m,n,1,i}(p) \quad i = 1, \ldots, m+n,\\
&\hat{V}_{2,i}(p) = (m+n)\hat{R}_{m,n,2}(p) - (m+n-1)\hat{R}_{m,n,2,i}(p) \quad i = 1, \ldots, m+n.
\end{aligned}$$

Thus, we obtain

$$\hat{V}_i(p) = \hat{V}_{1,i}(p) - \hat{V}_{2,i}(p).$$

By the proof of Lemma 2 in Gong et al. (2010), we have

$$\frac{1}{m+n}\sum_{j=1}^{m+n}\hat{V}_{1,j}(p) - R_1(p) = 1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right) - R_1(p) + O_p\left(\frac{n(m+n-1)}{(m+n)(n-1)^2 h}\right).$$

Along with Lemma B.1, we establish Lemma B.2 as follows

$$\sqrt{m+n}\left\{\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i(p) - \Delta(p)\right\}$$

$$\begin{aligned}
&= \sqrt{m+n}\left\{\frac{1}{m+n}\sum_{j=1}^{m+n}\hat{V}_{1,j}(p) - R_1(p) - \left(\frac{1}{m+n}\sum_{j=1}^{m+n}\hat{V}_{2,j}(p) - R_2(p)\right)\right\} \\
&= \sqrt{m+n}\left[1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,1}(x_{1,j})}{h}\right) - R_1(p)\right] \\
&- \sqrt{m+n}\left[\left\{1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,2}(x_{2,j})}{h}\right) - R_2(p)\right\} + O_p(n^{-1})\right] \\
&= \sqrt{m+n}\left\{\hat{\Delta}_{m,n}(p) - \Delta(p)\right\} + o_p(1) \\
&\xrightarrow{\mathfrak{D}} N(0, \sigma^2(p)).
\end{aligned}$$

□

Lemma B.3. *Under conditions in Theorem 3.1, for any $p \in (a, b)$, we have*

$$\frac{1}{m+n}\sum_{i=1}^{m+n}\left\{\hat{V}_i(p) - \Delta(p)\right\}^2 \xrightarrow{P} \sigma^2(p),$$

where $\sigma^2(p)$ is defined in Theorem 3.1.

Proof. For $1 \leq i \leq m, k = 1, 2$, we make similar arguments as Gong et al. (2010),

$$
\begin{aligned}
&\hat{V}_{k,i}(p) \\
&= 1 - \frac{m+n}{m} \sum_{j=1}^{m} K\left(\frac{1-p-G_{n,k}(X_{k,j})}{h}\right) + \frac{m+n-1}{m-1} \sum_{j=1, j\neq i}^{m} K\left(\frac{1-p-G_{n,k}(X_{k,j})}{h}\right) \\
&= 1 + \left(\frac{m+n-1}{m-1} - \frac{m+n}{m}\right) \sum_{j=1}^{m} K\left(\frac{1-p-G_{n,k}(X_{k,j})}{h}\right) \\
&\quad - \frac{m+n-1}{m-1} \left\{\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,k}(X_{k,j})}{h}\right) - \sum_{j=1, j\neq i}^{m} K\left(\frac{1-p-G_{n,k}(X_{k,j})}{h}\right)\right\} \\
&= 1 + \frac{n}{(m-1)m} \sum_{j=1}^{m} K\left(\frac{1-p-G_{n,k}(X_{k,j})}{h}\right) - \frac{m+n-1}{m-1} K\left(\frac{1-p-G_{n,k}(X_{k,i})}{h}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&\hat{V}_{1,i}(p)\hat{V}_{2,i}(p) \\
&= \left\{1 - \frac{m+n-1}{m-1} K\left(\frac{1-p-G_{n,1}(X_{1,i})}{h}\right)\right\} \left\{1 - \frac{m+n-1}{m-1} K\left(\frac{1-p-G_{n,2}(X_{2,i})}{h}\right)\right\} \\
&\quad + \frac{n}{(m-1)m} \sum_{j=1}^{m} K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right) \frac{n}{(m-1)m} \sum_{j=1}^{m} K\left(\frac{1-p-G_{n,2}(X_{2,j})}{h}\right) \\
&\quad + \frac{n}{(m-1)m} \sum_{j=1}^{m} K\left(\frac{1-p-G_{n,2}(X_{2,i})}{h}\right) \left\{1 - \frac{m+n-1}{m-1} K\left(\frac{1-p-G_{n,1}(X_{1,i})}{h}\right)\right\} \\
&\quad + \frac{n}{(m-1)m} \sum_{j=1}^{m} K\left(\frac{1-p-G_{n,1}(X_{1,i})}{h}\right) \left\{1 - \frac{m+n-1}{m-1} K\left(\frac{1-p-G_{n,2}(X_{2,i})}{h}\right)\right\}.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
&\frac{1}{m+n}\sum_{i=1}^{m}\hat{V}_{1,i}(p)\hat{V}_{2,i}(p)\\
=&\frac{m}{m+n}+\frac{(m+n-1)^2}{(m-1)^2(m+n)}\sum_{i=1}^{m}K\left(\frac{1-p-G_{n,1}(X_{1,i})}{h}\right)K\left(\frac{1-p-G_{n,2}(X_{2,i})}{h}\right)\\
&-\frac{m+n-1}{m-1}\sum_{i=1}^{m}K\left(\frac{1-p-G_{n,1}(X_{1,i})}{h}\right)-\frac{m+n-1}{m-1}\sum_{i=1}^{m}K\left(\frac{1-p-G_{n,2}(X_{2,i})}{h}\right)\\
&+\frac{n^2}{(m-1)^2m(m+n)}\sum_{j=1}^{m}K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)\sum_{j=1}^{m}K\left(\frac{1-p-G_{n,2}(X_{2,j})}{h}\right)\\
&+\frac{n}{(m-1)m(m+n)}\sum_{j=1}^{m}K\left(\frac{1-p-G_{n,2}(X_{2,j})}{h}\right)\\
&\left\{m-\frac{m+n-1}{m-1}\sum_{i=1}^{m}K\left(\frac{1-p-G_{n,1}(X_{1,i})}{h}\right)\right\}\\
&+\frac{n}{(m-1)m(m+n)}\sum_{j=1}^{m}K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)\\
&\left\{m-\frac{m+n-1}{m-1}\sum_{i=1}^{m}K\left(\frac{1-p-G_{n,2}(X_{2,i})}{h}\right)\right\}.
\end{aligned}
\tag{B.11}
$$

By the uniform convergence of $F_m(x_1,x_2)$ and $G^{-1}_{m,1}(y_1)$ and $G^{-1}_{m,2}(y_2)$, we have that

$$
\begin{aligned}
&|F_m\{G^{-1}_{n,1}(1-p),G^{-1}_{n,2}(1-p)\}-F\{G^{-1}_{1}(1-p),G^{-1}_{2}(1-p)\}|\\
\le\ &|F_m\{G^{-1}_{n,1}(1-p),G^{-1}_{n,2}(1-p)\}-F\{G^{-1}_{n,1}(1-p),G^{-1}_{n,2}(1-p)\}|\\
+\ &|F\{G^{-1}_{n,1}(1-p),G^{-1}_{n,2}(1-p)\}-F\{G^{-1}_{1}(1-p),G^{-1}_{2}(1-p)\}|.
\end{aligned}
$$

Based on the generalization of the Glivenko-Cantelli theorem by Dehardt (1971), we know the first term, $\sup_{p\in[0,1]}|F_m\{G^{-1}_{n,1}(1-p),G^{-1}_{n,2}(1-p)\}-F\{G^{-1}_{n,1}(1-p),G^{-1}_{n,2}(1-p)\}|\xrightarrow{P}0$.

By the Glivenko-Cantelli theorem of quantile process for G_1^{-1} and G_2^{-1}, we have

$$\sup_{x\in[0,1]} |G_{m,1}^{-1}(x) - G_1^{-1}(x)| \longrightarrow 0 \quad a.s.$$

$$\sup_{x\in[0,1]} |G_{m,2}^{-1}(x) - G_2^{-1}(x)| \longrightarrow 0 \quad a.s.$$

Since the $F(x_1, x_2)$ is continuous for x_1 and x_2, for any $p \in (0,1)$, we have $|F\{G_{n,1}^{-1}(1-p), G_{n,2}^{-1}(1-p)\} - F\{G_1^{-1}(1-p), G_2^{-1}(1-p)\}| \xrightarrow{P} 0$. Thus, for any $p \in (0,1)$,

$$|F_m\{G_{n,1}^{-1}(1-p), G_{n,2}^{-1}(1-p)\} - F\{G_1^{-1}(1-p), G_2^{-1}(1-p)\}| \xrightarrow{P} 0.$$

Similarly, for any $p \in (0,1)$, we have

$$|G_m\{G_{n,1}^{-1}(1-p), G_{n,2}^{-1}(1-p)\} - G\{G_1^{-1}(1-p), G_2^{-1}(1-p)\}| \xrightarrow{P} 0.$$

Consider

$$\begin{aligned}
&\frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right) K\left(\frac{1-p-G_{n,2}(X_{2,j})}{h}\right) \\
=& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{1-p-G_{n,1}(x_1)}{h}\right) K\left(\frac{1-p-G_{n,2}(x_2)}{h}\right) \mathrm{d}F_m(x_1, x_2)
\end{aligned}$$

$$
\begin{aligned}
&= K\left(\frac{1-p-G_{n,1}(x_1)}{h}\right) K\left(\frac{1-p-G_{n,2}(x_2)}{h}\right) F_m(x_1,x_2)|_{-\infty}^{\infty}|_{-\infty}^{\infty} \\
&- \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_m(x_1,x_2)\mathrm{d}K\left(\frac{1-p-G_{n,1}(x_1)}{h}\right) K\left(\frac{1-p-G_{n,2}(x_2)}{h}\right) \\
&= \int_{-1}^{1}\int_{-1}^{1} F_m\{G_{n,1}^{-1}(1-p-hu_1), G_{n,2}^{-1}(1-p-hu_2)\}\mathrm{d}K(u_1)\mathrm{d}K(u_2) \\
&= \int_{-1}^{1}\int_{-1}^{1} [F_m(G_{n,1}^{-1}(1-p), G_{n,2}^{-1}(1-p)) - h\frac{\partial F_m\{G_{n,1}^{-1}(u_1), G_{n,2}^{-1}(1-p)\}}{\partial u_1} \\
&\quad h\frac{\partial F_m\{G_{n,1}^{-1}(u_2), G_{n,2}^{-1}(1-p)\}}{\partial u_2}]w(u_1)w(u_2)\mathrm{d}u_1\mathrm{d}u_2 + o_p(h^2) \\
&= F_m\{G_{n,1}^{-1}(1-p), G_{n,2}^{-1}(1-p)\} + \int_{-1}^{1}\int_{-1}^{1} w(u_1)w(u_2)\mathrm{d}u_1\mathrm{d}u_2 + o_p(1) \\
&= F\{G_1^{-1}(1-p), G_2^{-1}(1-p)\} + o_p(1).
\end{aligned} \tag{B.12}
$$

Based on (B.11) and (B.12) and Lemma 1 of Gong et al. (2010), we can conclude that

$$
\begin{aligned}
&\frac{1}{m+n}\sum_{i=1}^{m}\hat{V}_{1,i}(p)\hat{V}_{2,i}(p) \\
&\xrightarrow{\mathcal{P}} \frac{r}{1+r} + \frac{1+r}{r}F\{G_1^{-1}(1-p), G_2^{-1}(1-p)\} - \{1-R_1(p)+1-R_2(p)\} \\
&+ \{1-R_1(p)\}\{1-R_2(p)\}\frac{1}{r(1+r)} + \frac{1}{1+r}\left[1-\frac{1+r}{r}\{1-R_1(p)\}\right]\{1-R_2(p)\} \\
&+ \frac{1}{1+r}\left[1-\frac{1+r}{r}\{1-R_2(p)\}\right]\{1-R_1(p)\} \\
&= \frac{1+r}{r}F\{G_1^{-1}(1-p), G_2^{-1}(1-p)\} - \frac{1+r}{r}\{1-R_1(p)\}\{1-R_2(p)\} + \frac{r}{1+r}R_1(p)R_2(p).
\end{aligned} \tag{B.13}
$$

On the other hand, for $m < i \le m+n, k = 1,2$,

$$
\begin{aligned}
\hat{V}_{k,i}(p) = 1 &- \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,k}(X_{k,j})}{h}\right) + \frac{m+n-1}{m} \\
&\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,m-i,k}(X_{k,j})}{h}\right) - K\left(\frac{1-p-G_{n,k}(X_{k,j})}{h}\right)\right\}.
\end{aligned} \tag{B.14}
$$

From (B.14), we have

$$
\begin{aligned}
&\hat{V}_{1,i}(p)\hat{V}_{2,i}(p) \\
&= \left\{1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)\right\}\left\{1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,2}(X_{2,j})}{h}\right)\right\} \\
&+ \frac{(m+n-1)^2}{m^2}\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,m-i,1}(X_{1,j})}{h}\right) - K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)\right\} \\
&\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,m-i,2}(X_{2,j})}{h}\right) - K\left(\frac{1-p-G_{n,2}(X_{2,j})}{h}\right)\right\} \\
&+ \frac{m+n-1}{m}\left\{1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)\right\} \\
&\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,m-i,2}(X_{2,j})}{h}\right) - K\left(\frac{1-p-G_{n,2}(X_{2,j})}{h}\right)\right\} \\
&+ \frac{m+n-1}{m}\left\{1 - \frac{1}{m}\sum_{j=1}^{m} K\left(\frac{1-p-G_{n,2}(X_{2,j})}{h}\right)\right\} \\
&\sum_{j=1}^{m}\left\{K\left(\frac{1-p-G_{n,m-i,1}(X_{1,j})}{h}\right) - K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)\right\} \quad m < i \le m+n. \qquad \text{(B.15)}
\end{aligned}
$$

We know

$$
\begin{aligned}
&G_{n,k}(X_{k,j}) - G_{n,-i,k}(X_{k,j}) \\
&= \frac{1}{n}\sum_{l=1}^{n} I(Y_{k,l} \le X_{k,j}) - \frac{1}{n-1}\sum_{l=1, l\neq i}^{n} I(Y_{k,l} \le X_{k,j}) \\
&= \left(\frac{1}{n} - \frac{1}{n-1}\right)\sum_{l=1}^{n} I(Y_{k,l} \le X_{k,j}) - \frac{1}{n-1}\left\{\sum_{l=1}^{n} I(Y_{k,l} \le X_{k,j}) - \sum_{l=1, l\neq i}^{n} I(Y_{k,l} \le X_{k,j})\right\} \\
&= \frac{1}{n-1}\{G_n(X_{k,j}) - I(Y_{k,i} \le X_{k,j})\}, \quad k = 1, 2, \quad 1 \le i \le n, \qquad \text{(B.16)}
\end{aligned}
$$

where

$$
G_{n,k}(X_{k,j}) - G_{n,-i,k}(X_{k,j}) = \frac{1}{n-1}\{G_n(X_{k,j}) - I(Y_{k,j} \le X_{k,j})\} = O_p\left(\frac{1}{n-1}\right), \quad k = 1, 2, \quad 1 \le i \le n. \qquad \text{(B.17)}
$$

and

$$\sum_{j=1}^{n} G_{n,k}(X_{k,j}) - G_{n,-i,k}(X_{k,j}) = 0 \quad k = 1, 2, \quad 1 \le i \le n. \tag{B.18}$$

Define that

$$A_i = \sum_{j=1}^{m} \left\{ K\left(\frac{1-p-G_{n,-i,1}(X_{1,j})}{h}\right) - K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right) \right\}$$
$$\sum_{j=1}^{m} \left\{ K\left(\frac{1-p-G_{n,-i,2}(X_{2,j})}{h}\right) - K\left(\frac{1-p-G_{n,2}(X_{2,j})}{h}\right) \right\}, \quad 1 \le i \le n.$$

Then, by Taylor's expansion, we have

$$\begin{aligned}
A_i = &\{\sum_{j=1}^{m} \frac{(G_{n,1}(X_{1,j}) - G_{n,-i,1}(X_{1,j}))}{h} w\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right) \\
&+ \frac{1}{2m} \sum_{j=1}^{m} \left(\frac{G_{n,1}(X_{1,j}) - G_{n,-i,1}(X_{1,j})}{h}\right)^2 w^{'}\left(\frac{1-p-G_{n,1}(X_{1,j}) + \xi_{n,j}}{h}\right)\} \\
&\{\sum_{j=1}^{m} \frac{(G_{n,2}(X_{2,j}) - G_{n,-i,2}(X_{2,j}))}{h} w\left(\frac{1-p-G_{n,2}(X_{2,j})}{h}\right) \\
&+ \frac{1}{2m} \sum_{j=1}^{m} \left(\frac{G_{n,2}(X_{2,j}) - G_{n,-i,2}(X_{2,j})}{h}\right)^2 w^{'}\left(\frac{1-p-G_{n,2}(X_{2,j}) + \xi_{n,j}}{h}\right)\} + o_p(1) \\
= &\frac{1}{h^2} \sum_{j=1}^{m} \{G_{n,1}(X_{1,j}) - G_{n,-i,1}(X_{1,j})\} w\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right) \\
&\sum_{j=1}^{m} \{G_{n,2}(X_{2,j}) - G_{n,-i,2}(X_{2,j})\} w\left(\frac{1-p-G_{n,2}(X_{2,j})}{h}\right) + O(h^{-1}n^{-1}), \quad 1 \le i \le n.
\end{aligned} \tag{B.19}$$

Based on (B.15), (B.16), we consider

$$
\begin{aligned}
\frac{1}{m+n}\sum_{i=1}^{n} A_i =& \frac{1}{m+n}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}\left\{K\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)-K\left(\frac{1-p-G_{n,-i,1}(X_{1,j})}{h}\right)\right\} \\
&\left\{K\left(\frac{1-p-G_{n,2}(X_{2,l})}{h}\right)-K\left(\frac{1-p-G_{n,-i,2}(X_{2,l})}{h}\right)\right\} \\
&= \frac{1}{m+n}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}\{G_{n,1}(X_{1,j})-G_{n,-i,1}(X_{1,j})\}\{G_{n,2}(X_{2,l})-G_{n,-i,2}(X_{2,l})\} \\
&\quad w\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)w\left(\frac{1-p-G_{n,2}(X_{2,l})}{h}\right)+o_p(1) \\
&= \frac{1}{(m+n)h^2(n-1)^2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}\{G_n(X_{1,j})G_n(X_{2,l})-G_n(X_{1,j})I(Y_{2,i}\le X_{2,l}) \\
&\quad -G_n(X_{2,l})I(Y_{1,i}\le X_{1,j})+I(Y_{1,i}\le X_{1,j})I(Y_{2,i}\le X_{2,l})\} \\
&\quad w\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)w\left(\frac{1-p-G_{n,2}(X_{2,l})}{h}\right)+o_p(1) \\
&= \frac{1}{(m+n)h^2(n-1)^2}\sum_{j=1}^{m}\sum_{l=1}^{m}\{nG_{n,1}(X_{1,j})G_{n,2}(X_{2,l})-2nG_{n,1}(X_{1,j})G_{n,2}(X_{2,l})\} \\
&\quad w\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)w\left(\frac{1-p-G_{n,2}(X_{2,l})}{h}\right) \\
&\quad +\frac{m^2}{(m+n)h^2(n-1)^2}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(Y_{1,i}\le x_1)I(Y_{2,i}\le x_2) \\
&\quad w\left(\frac{1-p-G_{n,1}(X_{1,j})}{h}\right)w\left(\frac{1-p-G_{n,2}(X_{2,l})}{h}\right)\mathrm{d}F_{m,1}(x_1)\mathrm{d}F_{m,2}(x_2)+o_p(1) \\
&= \frac{-nm^2}{(m+n)h^2(n-1)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_{n,1}(x_1)G_{n,2}(x_2)w\left(\frac{1-p-G_{n,1}(x_1)}{h}\right) \\
&\quad w\left(\frac{1-p-G_{n,2}(x_2)}{h}\right)\mathrm{d}F_{m,1}(x_1)\mathrm{d}F_{m,2}(x_2)+\frac{m^2}{(m+n)h^2(n-1)^2} \\
&\quad \sum_{i=1}^{n}\int_{-1}^{1}\int_{-1}^{1} I\{Y_{1,i}\le G_{n,1}^{-1}(1-p-hu_1)\}I\{Y_{2,i}\le G_{n,2}^{-1}(1-p-hu_2)\} \\
&\quad w(u_1)w(u_2)\mathrm{d}F_{m,1}\{G_{n,1}^{-1}(1-p-hu_1)\}\mathrm{d}F_{m,2}\{G_{n,2}^{-1}(1-p-hu_2)\}+o_p(n^{-1/2}) \\
&= \frac{-nm^2}{h^2(m+n)(n-1)^2}\int_{-1}^{1}\int_{-1}^{1} G_{n,1}\{G_{n,1}^{-1}(1-p-hu_1)\}G_{n,2}\{G_{n,2}^{-1}(1-p-hu_2)\} \\
&\quad w(u_1)w(u_2)\mathrm{d}F_{m,1}\{G_{n,1}^{-1}(1-p-hu_1)\}\mathrm{d}F_{m,2}\{G_{n,2}^{-1}(1-p-hu_2)\} \\
&\quad +\frac{nm^2}{h^2(m+n)(n-1)^2}\int_{-1}^{1}\int_{-1}^{1} G_n\{G_{n,1}^{-1}(1-p-hu_1),G_{n,2}^{-1}(1-p-hu_2)\} \\
&\quad w(u_1)w(u_2)\mathrm{d}F_{m,1}\{G_{n,1}^{-1}(1-p-hu_1)\}\mathrm{d}F_{m,2}\{G_{n,2}^{-1}(1-p-hu_2)\}+o_p(1).
\end{aligned}
$$

Recall that empirical estimators of ROC curves uniformly converge to true ROC curves, i.e., $\sup_{x\in[-h,-h]}|F_{m,1}(G_{n,1}^{-1}(x)) - F_1(G_1^{-1}(x))| = o_p(1)$. The above equality is,

$$
\begin{aligned}
&= \frac{-nm^2}{(m+n)(n-1)^2}\int_{-h}^{h}\int_{h}^{h} G_{n,1}\{G_{n,1}^{-1}(1-p-v_1)\}G_{n,2}\{G_{n,2}^{-1}(1-p-v_2)\} \\
&\quad R_1^{'}(p)R_2^{'}(p)w\left(\frac{v_1}{h}\right)w\left(\frac{v_2}{h}\right)\mathrm{d}v_1\mathrm{d}v_1 \\
&\quad + \frac{m^2}{(m+n)(n-1)^2}\int_{-h}^{h}\int_{-h}^{h} G_n\{G_{n,1}^{-1}(1-p-v_1), G_{n,2}^{-1}(1-p-v_2)\} \\
&\quad w\left(\frac{v_1}{h}\right)w\left(\frac{v_2}{h}\right)R_1^{'}(p)R_2^{'}(p)\mathrm{d}v_1\mathrm{d}v_1 + o_p(1) \\
&= \frac{-nm^2}{(m+n)(n-1)^2}(1-p)^2R_1^{'}(p)R_2^{'}(p) \\
&+ \frac{m^2}{(m+n)(n-1)^2}R_1^{'}(p)R_2^{'}(p)G_n\{G_{n,1}^{-1}(1-p), G_{n,2}^{-1}(1-p)\} + o_p(1) \\
&= \frac{nm^2}{(m+n)(n-1)^2}R_1^{'}(p)R_2^{'}(p)\left\{G_n(G_{n,1}^{-1}(1-p), G_{n,2}^{-1}(1-p)) - (1-p)^2\right\} + o_p(1) \\
&\xrightarrow{P} \frac{r^2}{1+r}R_1^{'}(p)R_2^{'}(p)[G\{G_1^{-1}(1-p), G_2^{-1}(1-p)\} - (1-p)^2].
\end{aligned} \tag{B.20}
$$

Combining the Taylor's expansion, (B.15), (B.19) and (B.20), we have

$$
\begin{aligned}
&\frac{1}{m+n}\sum_{i=m+1}^{m+n}\hat{V}_{1,i}(p)\hat{V}_{2,i}(p) \\
&\xrightarrow{P} \frac{(1+r)^2}{r^2}\frac{r^2}{1+r}R_1^{'}(p)R_2^{'}(p)[G\{G_1^{-1}(1-p), G_2^{-1}(1-p)\} - (1-p)^2] + \frac{1}{1+r}R_1(p)R_2(p) \\
&= (1+r)R_1^{'}(p)R_2^{'}(p)[G\{G_1^{-1}(1-p), G_2^{-1}(1-p)\} - (1-p)^2] + \frac{1}{1+r}R_1(p)R_2(p).
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_{1,i}(p)\hat{V}_{2,i}(p) \\
&\xrightarrow{P} (1+r)R_1^{'}(p)R_2^{'}(p)[G\{G_1^{-1}(1-p),G_2^{-1}\}(1-p)-(1-p)^2]+\frac{1}{1+r}R_1(p)R_2(p) \\
&\quad+\frac{1+r}{r}F\{G_1^{-1}(1-p),G_2^{-1}(1-p)\}-\frac{1+r}{r}\{1-R_1(p)\}\{1-R_2(p)\}+\frac{r}{1+r}R_1(p)R_2(p) \\
&=(1+r)R_1^{'}(p)R_2^{'}(p)[G\{G_1^{-1}(1-p),G_2^{-1}(1-p)\}-(1-p)^2] \\
&\quad+\frac{1+r}{r}F\{G_1^{-1}(1-p),G_2^{-1}(1-p)\}-\frac{1+r}{r}\{1-R_1(p)\}\{1-R_2(p)\}+R_1(p)R_2(p).
\end{aligned}
$$

Finally, we have

$$
\begin{aligned}
&\frac{1}{m+n}\sum_{i=1}^{m+n}\left\{\hat{V}_i(p)-\Delta(p)\right\}^2 \\
&=\frac{1}{m+n}\sum_{i=1}^{m+n}\left\{\hat{V}_{1,i}(p)-R_1(p)\right\}^2+\frac{1}{m+n}\sum_{i=1}^{m+n}\left\{\hat{V}_{2,i}(p)-R_2(p)\right\}^2 \\
&+\frac{2}{m+n}\sum_{i=1}^{m+n}\left\{\hat{V}_{1,i}(p)\hat{V}_{2,i}(p)-\hat{V}_{1,i}(p)R_2(p)-\hat{V}_{2,i}(p)R_1(p)+R_1(p)R_2(p)\right\} \\
&\xrightarrow{P}\sigma_1^2(p)+\sigma_2^2(p)+2\sigma_{12}^2(p) \\
&=\sigma^2(p).
\end{aligned}
$$

$\square$

Proof of Theorem 3.1 It follows directly from Lemmas B.2 and B.3. $\square$

Proof of Theorem 3.2 From Lemmas B.1 and B.2, we follow the similar arguments as Gong et al. (2010) and prove Theorem 3.2. $\square$

Appendix C

Proof of Theorems for Chapter 5

Proof of Theorem 5.1

Proof. We can decompose $\Pi_{m,n}(p,\theta)$ as

$$\Pi_{m,n}(p,\theta) = \frac{1}{m}\sum_{j=1}^{m} K\left\{\frac{p - F_{n,2}(x_j-\theta)}{h}\right\} - p - \Pi_m(p,\theta) + \Pi_m(p,\theta), \tag{C.1}$$

where $\Pi_m(p,\theta) = \frac{1}{m}\sum_{j=1}^{m} K\left\{\frac{p-F_2(x_j-\theta)}{h}\right\} - p.$

$$\begin{aligned}
\Pi_m(p,\theta) =& \frac{1}{m}\sum_{j=1}^{m} K\left\{\frac{p - F_2(x_j-\theta)}{h}\right\} - p \\
=& \int_{-\infty}^{\infty} K\left\{\frac{p - F_2(x-\theta)}{h}\right\} \mathrm{d}F_{m,1}(x) - p \\
=& K\left\{\frac{p - F_2(x-\theta)}{h}\right\} F_{m,1}(x)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} F_{m,1}(x)\mathrm{d}K\left\{\frac{p - F_2(x-\theta)}{h}\right\} - p \\
=& \frac{1}{h}\int_{-\infty}^{\infty} F_{m,1}(x) w\left\{\frac{p - F_2(x-\theta)}{h}\right\} \mathrm{d}F_2(x-\theta) - p \\
=& \int_{-1}^{1} F_{m,1}\{F_2^{-1}(p+uh)+\theta\} w\,(u)\,\mathrm{d}u - p \\
=& \int_{-1}^{1} [F_{m,1}\{F_2^{-1}(p+uh)+\theta\} - F_1\{F_2^{-1}(p+uh)+\theta\} \\
&+ F_1\{F_2^{-1}(p+uh)+\theta\} - F_1\{F_2^{-1}(p)+\theta\}]w\,(u)\,\mathrm{d}u \\
=& o_p(1).
\end{aligned} \tag{C.2}$$

The above equation is obtained by the Glivenko-Cantelli Theorem of F_1 and the bounded derivative of $D(\theta,p) = F_1(F_2^{-1}(p)+\theta)$. By the equations (10) and (11) in Gong et al. (2010), we can easily extend their result in our case, i.e., $\Pi_m(p,\theta) - \Pi_{m,n}(p,\theta) = o_p(1)$. Hence, we finish the proof

$$\Pi_{m,n}(p,\theta) \xrightarrow{\mathcal{P}} 0. \tag{C.3}$$

□

Lemma C.1. *Under the same assumptions as Theorem 5.1,*

$$\sqrt{m+n}\Pi_{m,n}(p,\theta) \xrightarrow{\mathfrak{D}} N(0,\sigma^2(p)), \tag{C.4}$$

where $\sigma^2(p)$ is defined in Theorem 5.2.

Proof.

$$\sqrt{m+n}\Pi_{m,n}(p,\theta) = \frac{\sqrt{m+n}}{\sqrt{m}}\sqrt{m}\{\Pi_m(p,\theta)\} + \frac{\sqrt{m+n}}{\sqrt{n}}\sqrt{n}\{\Pi_{m,n}(p,\theta) - \Pi_m(p,\theta)\} \tag{C.5}$$

For the first term of (C.5), we have

$$\begin{aligned}
&\sqrt{m}\{\Pi_m(p,\theta)\}\\
=&\int_{-1}^{1}\sqrt{m}[F_{m,1}\{F_2^{-1}(p+uh)+\theta\} - F_1\{F_2^{-1}(p+uh)+\theta\}]w\,(u)\,\mathrm{d}u\\
+&\sqrt{m}\int_{-1}^{1}F_1\{F_2^{-1}(p+uh)+\theta\} - F_1\{F_2^{-1}(p)+\theta\}]w\,(u)\,\mathrm{d}u\\
=&\int_{-1}^{1}W_{x_1}\{F_2^{-1}(p+uh)+\theta\}w\,(u)\,\mathrm{d}u + \sqrt{m}\int_{-1}^{1}D^{'}(\theta,p)uhw\,(u)\,\mathrm{d}u + O_p(\sqrt{m}h^2) && \text{(C.6)}\\
=&I + II + O_p(\sqrt{m}h^2), && \text{(C.7)}
\end{aligned}$$

where $W_{x_1}(t) = \sqrt{m}\{F_{m,1}(t) - F_1(t)\}$. Because of the symmetric property of kernel function, the second term of (C.6) is equal to zero. Due to the Donsker theorem and similar proofs for equation (9) in Gong et al. (2010), $I \xrightarrow{\mathfrak{D}} B_{F_1}\{F_2^{-1}(p)+\theta\}$ and $B_F(\cdot)$ is a Brownian bridge for distribution F. Hence, we have that

$$\sqrt{m}\{\Pi_m(p,\theta)\} \xrightarrow{\mathfrak{D}} B_{F_1}\{F_1^{-1}(p)\}. \tag{C.8}$$

For the second term of (C.5), we propose the procedure similar to Gong et al. (2010),

$$\begin{aligned}
&\sqrt{n}\{\Pi_{m,n}(p,\theta) - \Pi_m(p,\theta)\} \\
&= -\int_{-\infty}^{\infty} W_{F_2}(x) w\left\{\frac{p - F_2(x-\theta)}{h}\right\} \mathrm{d}F_1(x) + O_p(n^{-1/2}h^{-1}) \\
&= \int_{-1}^{1} W_{F_2}\{F_2^{-1}(p)\} w(u) D^{'}(p,\theta)\mathrm{d}u + O_p(n^{-1/2}h^{-1}) \\
&\xrightarrow{\mathfrak{D}} B_{F_2}\{F_2^{-1}(p)\} D^{'}(p,\theta).
\end{aligned} \tag{C.9}$$

Combining (C.5), (C.8), (C.9) and the independence of $B_{F_1}(\cdot)$ and $B_{F_2}(\cdot)$, we have

$$\sqrt{m+n}\Pi_{m,n}(p,\theta) \xrightarrow{\mathfrak{D}} N(0, \sigma^2(p)). \tag{C.10}$$

□

Proof of Theorem 5.2

Proof. First, we introduce some properties of $F_{n,2,-i}$ as follows:

$$F_{n,2}(x_j) - F_{n,2,-i}(x_j) = \frac{1}{n-1}\{F_{n,2}(x_j) - I(Y_i \le x_j)\} = O_p\left(\frac{1}{n-1}\right), i = 1, ..., n \tag{C.11}$$

and

$$\sum_{i=1}^{n}\{F_{n,2,-i}(x_j) - F_{n,2}(x_j)\} = 0, \tag{C.12}$$

because

$$
\begin{aligned}
&F_{n,2}(x_j) - F_{n,2,-i}(x_j) \\
=& \frac{1}{n}\sum_{k=1}^{n} I(y_i \le x_j) - \frac{1}{n-1}\sum_{i=k,k\neq i}^{n} I(y_i \le x_j) \\
=& \left(\frac{1}{n} - \frac{1}{n-1}\right)\sum_{k=1}^{n} I(y_i \le x_j) - \frac{1}{n-1}\left\{\sum_{k=1}^{n} I(y_i \le x_j) - \sum_{i=k,k\neq i}^{n} I(y_i \le x_j)\right\} \\
=& \frac{1}{n-1}\{F_{n,2}(x_j) - I(y_i \le x_j)\}. \qquad \text{(C.13)}
\end{aligned}
$$

For the pseudo sample, based on (C.16) in Gong et al. (2010), we have

$$
\begin{aligned}
&\left\{\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i(p,\theta)\right\} \\
=&\frac{1}{m+n}\left[\sum_{i=1}^{m}\left\{\frac{m+n}{m}\sum_{j=1}^{m}K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} - \frac{m+n-1}{m-1}\sum_{j=1,j\neq i}^{m}K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\}\right\}\right] - p \\
&+\frac{1}{m+n}\left[\sum_{i=m+1}^{m+n}\left\{\frac{m+n}{m}\sum_{j=1}^{m}K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} - \frac{m+n-1}{m}\sum_{j=1}^{m}K\left\{\frac{p-F_{n,2,m-i}(x_j-\theta)}{h}\right\}\right\}\right] \\
=&\frac{1}{m+n}\left[\sum_{j=1}^{m}K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} + \frac{n}{m}\sum_{j=1}^{m}K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\}\right] \\
&+\frac{m+n-1}{(m+n)m}\sum_{i=m+1}^{m+n}\sum_{j=1}^{m}\left[K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} - K\left\{\frac{p-F_{n,2,m-i}(x_j-\theta)}{h}\right\}\right] - p \\
=&\frac{1}{m}\sum_{j=1}^{m}K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} - p + O_p\left\{\frac{mn}{(m+n)(n-1)^2h}\right\}.
\end{aligned}
$$

Using (C.11) and (C.12), we have that

$$
\begin{aligned}
\sqrt{m+n}\left\{\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{V}_i(p,\theta)\right\} =&\sqrt{m+n}\,\Pi_{m,n}(p,\theta) + o_p(1) \\
&\xrightarrow{\mathfrak{D}} N(0,\sigma^2(p)). \qquad \text{(C.14)}
\end{aligned}
$$

□

Proof of Theorem 5.3

Proof. For $1 \le i \le m$,

$$
\begin{aligned}
\hat{V}_i(p,\theta) =& \frac{m+n}{m}\sum_{j=1}^{m} K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} - \frac{m+n-1}{m-1}\sum_{j=1, j\neq i}^{m} K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} - p\\
=& \frac{m+n-1}{m-1}K\left\{\frac{p-F_{n,2}(x_i-\theta)}{h}\right\} - \frac{n}{m(m-1)}\sum_{j=1}^{m} K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} - p
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{V}_i^2(p,\theta) =& \left[\frac{m+n-1}{m-1}K\left\{\frac{p-F_{n,2}(x_i-\theta)}{h}\right\}\right]^2 + \left[\frac{n}{m(m-1)}\sum_{j=1}^{m} K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\}\right]^2\\
&-\frac{2(m+n-1)n}{m(m-1)^2}K\left\{\frac{p-F_{n,2}(x_i-\theta)}{h}\right\}\sum_{j=1}^{m} K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} + p^2\\
&-2p\left[\frac{m+n-1}{m-1}K\left\{\frac{p-F_{n,2}(x_i-\theta)}{h}\right\} - \frac{n}{m(m-1)}\sum_{j=1}^{m} K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\}\right].
\end{aligned}
$$

After tedious computation, we have

$$
\begin{aligned}
\frac{1}{m+n}\sum_{j=1}^{m}\hat{V}_i^2(p,\theta) \xrightarrow{\mathfrak{P}}& \frac{m+n}{m}p + \frac{(n-m-2n)(m+n)}{m(m+n)}p^2\\
=& \frac{m+n}{m}p(1-p).
\end{aligned}
\tag{C.15}
$$

For $m+1 \le i \le m+n$, we have

$$
\begin{aligned}
\hat{V}_i(p,\theta) =& \frac{m+n}{m}\sum_{j=1}^{m} K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} - \frac{m+n-1}{m}\sum_{j=1}^{m} K\left\{\frac{p-F_{n,2,m-i}(x_j-\theta)}{h}\right\} - p\\
=& \frac{m+n-1}{m}\sum_{j=1}^{m}\left[K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} - K\left\{\frac{p-F_{n,2,m-i}(x_j-\theta)}{h}\right\}\right]\\
&+\frac{1}{m}\sum_{j=1}^{m} K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} - p,
\end{aligned}
$$

and

$$\hat{V}_i^2(p,\theta) = \left\{\frac{m+n-1}{m}\right\}^2 \left[\sum_{j=1}^m K\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} - K\left\{\frac{p-F_{n,2,m-i}(x_j-\theta)}{h}\right\}\right]^2 + o_p(1)$$
$$= \left\{\frac{m+n-1}{m}\right\}^2 \left[\sum_{j=1}^m w\left\{\frac{p-F_{n,2}(x_j-\theta)}{h}\right\} \frac{F_{n,2}(x_j-\theta)-F_{n,2,m-i}(x_j-\theta)}{h}\right]^2 + o_p(1).$$

We follow the argument which is similar to Gong et al. (2010),

$$\begin{aligned}
&\frac{1}{m+n}\sum_{j=m+1}^{m+n}\hat{V}_i^2(p,\theta)\\
=&\frac{m+n}{nh^2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}\{F_{n,2}(x_j)F_{n,2}(x_l)-F_{n,2}(x_j)I(y_i\le x_l)\\
&-F_{n,2}(x_l)I(y_i\le x_j)+I(y_i\le x_j)I(y_i\le x_l)\}w\left(\frac{p-F_{n,1}(x_j-\theta)}{h}\right)w\left(\frac{p-F_{n,2}(x_l-\theta)}{h}\right)+o_p(1)\\
=&\frac{m+n}{nh^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{F_{n,2}(x_1\wedge x_2)-F_{n,2}(x_1)F_{n,2}(x_2)\}\\
&w\left(\frac{p-F_{n,1}(x_j-\theta)}{h}\right)w\left(\frac{p-F_{n,2}(x_l-\theta)}{h}\right)\mathrm{d}F_{m,1}(x_1)\mathrm{d}F_{m,2}(x_2)+o_p(1)\\
=&\frac{m+n}{nh^2}\int_{-1}^{1}\int_{-1}^{1}\{F_2\{F_2^{-1}(p-u_1h)\wedge F_2^{-1}(p-u_2h)\}-F_2\{F_2^{-1}(p-u_1h)\}F_2\{F_2^{-1}(p-u_2h)\}\}\\
&w\left(u_1\right)w\left(u_2\right)\mathrm{d}F_1\{F_2^{-1}(p-u_1h)+\theta\}\mathrm{d}F_1\{F_2^{-1}(p-u_2h)+\theta\}+o_p(1)\\
=&\frac{m+n}{n}\int_{-1}^{1}\int_{-1}^{1}p(1-p)\{D'(p,\theta)\}^2w(u_1)w(u_2)\mathrm{d}u_1\mathrm{d}u_2\\
=&\frac{m+n}{n}p(1-p)\{D'(p,\theta)\}^2+o_p(1).
\end{aligned}\tag{C.16}$$

Hence,

$$\frac{1}{m+n}\sum_{j=1}^{m+n}\hat{V}_i^2(p,\theta)\xrightarrow{\mathcal{P}}\sigma^2(p).$$

□

Proof of Theorem 5.4 From Theorem 5.2 and Theorem 5.3, we follow the standard arguments in Owen (1990) and prove Theorem 5.4. □

Proof of Theorem 5.5

Proof. We can decompose $\Phi_m(s,t,\eta)$ as follows

$$\Phi_{m,m}(s,t,\eta) = \frac{1}{m}\sum_{j=1}^{m} K\left\{\frac{p - F_m(x_j-\eta)}{h}\right\} - p - \Phi_{m,1}(s,t,\eta) + \Phi_{m,1}(s,t,\eta), \tag{C.17}$$

where $\Phi_{m,1}(s,t,\eta) = \frac{1}{m}\sum_{j=1}^{m} K\left\{\frac{p-F(x_j-\eta)}{h}\right\} - t$.

$$\begin{aligned}
\Phi_{m,1}(s,t,\eta) =& \frac{1}{m}\sum_{j=1}^{m} K\left\{\frac{s-F(x_j-\eta)}{h}\right\} - t \\
=& \int_{-\infty}^{\infty} K\left\{\frac{s-F(x-\eta)}{h}\right\} \mathrm{d}F_m(x) - t \\
=& K\left\{\frac{s-F(x-\eta)}{h}\right\} F_m(x)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} F_m(x)\mathrm{d}K\left\{\frac{s-F(x-\eta)}{h}\right\} - t \\
=& \frac{1}{h}\int_{-\infty}^{\infty} F_m(x) w\left\{\frac{s-F(x-\eta)}{h}\right\} \mathrm{d}F(x-\eta) - t \\
=& -\int_{-1}^{1} F_m\{F^{-1}(s-uh)+\eta\} w\,(u)\,\mathrm{d}u - t \\
=& -\int_{-1}^{1} [F_m\{F^{-1}(s-uh)+\eta\} - F\{F^{-1}(s-uh)+\eta\} \\
&+ F\{F^{-1}(s-uh)+\eta\} - F\{F^{-1}(s)+\eta\}] w\,(u)\,\mathrm{d}u \\
=& o_p(1).
\end{aligned} \tag{C.18}$$

The above equation is obtained by the Glivenko-Cantelli Theorem of F and the bounded derivative of $Q(s,t,\eta) = F(F^{-1}(s)+\eta) - t$. By the equations (10) and (11) in Gong et al. (2010), we can easily extend their result in our case, i.e.,

$$\Phi_{m,1}(s,t,\eta) - \Phi_m(s,t,\eta) = o_p(1). \tag{C.19}$$

Hence, we finish the proof

$$\Phi_m(s,t,\eta) \xrightarrow{\mathcal{P}} 0. \tag{C.20}$$

□

Proof of Theorem 5.6

Proof.

$$\sqrt{m}\Phi_m(s,t,\eta) = \sqrt{m}\{\Phi_{m,1}(s,t,\eta)\} + \sqrt{m}\{\Phi_m(s,t,\eta) - \Phi_{m,1}(s,t,\eta)\}. \tag{C.21}$$

For the first term of (C.5), we have

$$\begin{aligned}
&\sqrt{m}\{\Phi_{m,1}(s,t,\eta)\}\\
=&\int_{-1}^{1}\sqrt{m}[F_m\{F^{-1}(s-uh)+\eta\} - F\{F^{-1}(s-uh)+\eta\}]w\,(u)\,\mathrm{d}u\\
+&\sqrt{m}\int_{-1}^{1}F\{F^{-1}(s-uh)+\eta\} - F\{F^{-1}(p)+\eta\}]w\,(u)\,\mathrm{d}u\\
=&\int_{-1}^{1}W_{x_1}\{F^{-1}(s-uh)+\eta\}w\,(u)\,\mathrm{d}u + \sqrt{m}\int_{-1}^{1}Q^{'}(s,t,\eta)uhw\,(u)\,\mathrm{d}u + O_p(\sqrt{m}h^2)\\
=&I + II + O_p(\sqrt{m}h^2),
\end{aligned} \tag{C.22}$$

where $W_{x_1}(t) = \sqrt{m}\{F_m(t) - F(t)\}$. Because of the symmetric property of kernel function, the second term of (C.6) is equal to zero. For the second term of (C.5), we propose the procedure similar to Gong et al. (2010),

$$\begin{aligned}
&\sqrt{n}\{\Phi_m(s,t,\eta) - \Phi_{m,1}(s,t,\eta)\}\\
=&-\int_{-\infty}^{\infty}W_F(x-\eta)w\left\{\frac{s-F(x-\eta)}{h}\right\}\mathrm{d}F(x) + O_p(n^{-1/2}h^{-1})\\
=&\int_{-1}^{1}W_F\{F^{-1}(s)\}w(u)Q^{'}(s,t,\eta)\mathrm{d}u + O_p(n^{-1/2}h^{-1})
\end{aligned}$$

Due to the Donsker theorem,

$$\begin{aligned}\sqrt{m}\Phi_m(s,t,\eta) =& W_F\{F^{-1}(s)\}Q^{'}(s,t,\eta) + W_F\{F^{-1}(t)\} + O_p(\sqrt{m}h^2) + O_p(n^{-1/2}h^{-1}) \\ &\xrightarrow{\mathfrak{D}} N(0,\sigma_1^2) \qquad \text{(C.23)}\end{aligned}$$

and $B_F(\cdot)$ is Brownian bridge for distribution F.

□

Proof of Theorem 5.7

Proof. First, we introduce some property of $F_{m,-i}$ as follows:

$$F_m(x_j) - F_{m,-i}(x_j) = \frac{1}{n-1}\{F_m(x_j) - I(x_i \le x_j)\} = O_p\left(\frac{1}{n-1}\right), i = 1, ..., n \qquad \text{(C.24)}$$

and

$$\sum_{i=1}^{n}\{F_{m,-i}(x_j) - F_m(x_j)\} = 0, \qquad \text{(C.25)}$$

because

$$\begin{aligned}
&F_m(x_j) - F_{m,-i}(x_j) \\
&= \frac{1}{n}\sum_{k=1}^{n} I(x_i \le x_j) - \frac{1}{n-1}\sum_{i=k,k\ne i}^{n} I(x_i \le x_j) \\
&= \left(\frac{1}{n} - \frac{1}{n-1}\right)\sum_{k=1}^{n} I(x_i \le x_j) - \frac{1}{n-1}\left\{\sum_{k=1}^{n} I(x_i \le x_j) - \sum_{i=k,k\ne i}^{n} I(x_i \le x_j)\right\} \\
&= \frac{1}{n-1}\{F_m(x_j) - I(x_i \le x_j)\}. \qquad \text{(C.26)}
\end{aligned}$$

For the pseudo sample, based on (16) in Gong et al. (2010), we have

$$
\begin{aligned}
&\left\{\frac{1}{m}\sum_{i=1}^{m+n}\hat{U}_i(s,t,\eta)\right\}\\
=&m\Phi_m(s,t,\eta)-\frac{m-1}{n}\sum_{i=1}^{m}\Phi_{m,m,-i}(s,t,\eta)-t\\
=&\sum_{i=1}^{m}K\left\{\frac{s-F_m(x_i-\eta)}{h}\right\}-\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1,j\neq i}^{m}K\left\{\frac{s-F_m(x_j-\eta)}{h}\right\}\\
+&\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1,j\neq i}^{m}\left[K\left\{\frac{s-F_m(x_j-\eta)}{h}\right\}-K\left\{\frac{s-F_{m,-i}(x_j-\eta)}{h}\right\}\right]-t\\
=&\frac{1}{m}\sum_{j=1}^{m}K\left\{\frac{s-F_m(x_j-\eta)}{h}\right\}-t+O_p\left\{\frac{mn}{(m+n)(n-1)^2h}\right\}.
\end{aligned}
\tag{C.27}
$$

Using (C.10) and (C.11), we have

$$
\begin{aligned}
\sqrt{m}\left\{\frac{1}{m+n}\sum_{i=1}^{m+n}\hat{U}_i(s,t,\eta)\right\}=&\sqrt{m}\Phi_m(s,t,\eta)+o_p(1)\\
\xrightarrow{\mathfrak{D}}&N(0,\sigma_1^2).
\end{aligned}
\tag{C.28}
$$

□

Proof of Theorem 5.8

Proof. For $1\le i\le m$,

$$
\hat{U}_i(s,t,\eta)=\sum_{j=1}^{m}\left[K\left\{\frac{s-F_m(x_j-\eta)}{h}\right\}-K\left\{\frac{s-F_{m,-i}(x_j-\eta)}{h}\right\}\right]+K\left\{\frac{s-F_{m,-i}(x_i-\eta)}{h}\right\}-t
$$

and

$$
\begin{aligned}
\hat{U}_i^2(s,t,\eta) = & \left\{ \sum_{j=1}^{m} \left[K\left\{ \frac{s - F_m(x_j - \eta)}{h} \right\} - K\left\{ \frac{s - F_{m,-i}(x_j - \eta)}{h} \right\} \right] \right\}^2 \\
& + \left[K\left\{ \frac{s - F_{m,-i}(x_i - \eta)}{h} \right\} - t \right]^2 \\
& + 2\left\{ \sum_{j=1}^{m} \left[K\left\{ \frac{s - F_m(x_j - \eta)}{h} \right\} - K\left\{ \frac{s - F_{m,-i}(x_j - \eta)}{h} \right\} \right] \left[K\left\{ \frac{s - F_{m,-i}(x_i - \eta)}{h} \right\} - t \right] \right\}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& \frac{1}{m} \sum_{i=1}^{m} \hat{U}_i^2(s,t,\eta) \\
= & \frac{1}{m} \sum_{i=1}^{m} \left\{ \sum_{j=1}^{m} \left[K\left\{ \frac{s - F_m(x_j - \eta)}{h} \right\} - K\left\{ \frac{s - F_{m,-i}(x_j - \eta)}{h} \right\} \right] \right\}^2 \\
+ & \frac{1}{m} \sum_{i=1}^{m} K^2\left\{ \frac{s - F_{m,-i}(x_i - \eta)}{h} \right\} + t^2 - \frac{2t}{m} \sum_{i=1}^{m} K\left\{ \frac{s - F_{m,-i}(x_i - \eta)}{h} \right\} \\
+ & \frac{2}{m} \sum_{i=1}^{m} \left\{ \sum_{j=1}^{m} \left[K\left\{ \frac{s - F_m(x_j - \eta)}{h} \right\} - K\left\{ \frac{s - F_{m,-i}(x_j - \eta)}{h} \right\} \right] \left[K\left\{ \frac{s - F_{m,-i}(x_i - \eta)}{h} \right\} - t \right] \right\}
\end{aligned}
\tag{C.29}
$$

After tedious computation as proof of Theorem 5.4, we have

$$
\frac{1}{m} \sum_{i=1}^{m} \left\{ \sum_{j=1}^{m} \left[K\left\{ \frac{s - F_m(x_j - \eta)}{h} \right\} - K\left\{ \frac{s - F_{m,-i}(x_j - \eta)}{h} \right\} \right] \right\}^2 \xrightarrow{\mathfrak{P}} s(1-s)Q'(s,t,\eta)^2. \tag{C.30}
$$

Following equation (18) of Gong (2010), we know

$$
\frac{1}{m} \sum_{i=1}^{m} K^2\left\{ \frac{s - F_{m,-i}(x_i - \eta)}{h} \right\} + t^2 - \frac{2t}{m} \sum_{i=1}^{m} K\left\{ \frac{s - F_{m,-i}(x_i - \eta)}{h} \right\} \xrightarrow{\mathfrak{P}} t(1-t). \tag{C.31}
$$

Moreover, based on (C.9),

$$\frac{2}{m}\sum_{i=1}^{m}\left\{\sum_{j=1}^{m}\left[K\left\{\frac{s-F_m(x_j-\eta)}{h}\right\}-K\left\{\frac{s-F_{m,-i}(x_j-\eta)}{h}\right\}\right]\left[K\left\{\frac{s-F_{m,-i}(x_i-\eta)}{h}\right\}-t\right]\right\}$$

$$=\frac{2}{m}\sum_{i=1}^{m}\sum_{j=1}^{m}\left[K\left\{\frac{s-F_m(x_j-\eta)}{h}\right\}-K\left\{\frac{s-F_{m,-i}(x_j-\eta)}{h}\right\}\right]^2+o_p(1)$$

$$=\frac{2}{m}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\{\frac{F_m(x_j-\eta)-F_{m,-i}(x_j-\eta)}{h}\right\}^2 w^2\left\{\frac{s-F_{m,-i}(x_j-\eta)}{h}\right\}+o_p(1)$$

$$=\frac{2}{m(m-1)^2h^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\{F_m^2(x_j-\eta)-2F_m(x_j-\eta)I(x_i<x_j-\eta)+I(x_i<x_j-\eta)\right\}$$

$$w^2\left\{\frac{s-F_m(x_j-\eta)}{h}\right\}+o_p(1)$$

$$=\frac{2m}{(m-1)^2h^2}\int_{-\infty}^{\infty}\left\{F_m(x-\eta)-F_m^2(x-\eta)\right\}w^2\left\{\frac{s-F_m(x_j-\eta)}{h}\right\}\mathrm{d}F_m(x)$$

$$\xrightarrow{\mathfrak{P}} 2s(1-s)Q^{'}(s,t,\eta) \tag{C.32}$$

Hence,

$$\frac{1}{m}\sum_{j=1}^{m}\hat{U}_i^2(s,t,\eta)\xrightarrow{\mathcal{P}}\sigma_1^2.$$

□

Proof of Theorem 5.9 From Theorem 5.7 and Theorem 5.8, we follow the standard arguments in Owen (1990) and prove Theorem 5.9. □

Appendix D

Proof of Theorems for Chapter 6

Lemma D.1. *Under Assumptions 1-5, as* $n \to \infty$

$$\sqrt{n}\hat{V}(\theta_0) \xrightarrow{\mathfrak{D}} N(0, 4\Gamma(\theta_0)). \tag{D.1}$$

Proof. By the definition of $\hat{d}_i(\theta_0)$, we can rewrite

$$\hat{d}_i(\theta) = 2\left\{ \frac{\hat{q}(\theta, X_i)(1-\delta_i)}{\hat{\pi}(X_i)} - \sum_{j=1}^{n} \frac{\hat{q}(\theta, X_j)I(X_i \geq X_j)(1-\delta_j)}{n\hat{\pi}(X_j)\hat{\pi}(X_j)} \right\}.$$

Hence, we have $\sum_{i=1}^{n} \hat{d}_i(\theta_0) = 0$. By the asymptotic normality of $U_w(\theta_0)$ (see Fine et al., 1998),

$$\begin{aligned}
\sqrt{n}\hat{V}(\theta_0) &= \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \{\hat{b}(U_i, U_j, \theta_0)\} \\
&= \frac{1}{\sqrt{n}(n-1)} \left[\sum_{i=1}^{n} \sum_{j=1, i\neq j}^{n} \{\hat{e}_{ij}(\theta_0) + \hat{d}_i(\theta_0) + \hat{e}_{ji}(\theta_0) + \hat{d}_j(\theta_0)\} \right] \\
&= \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j=1, i\neq j}^{n} \{\hat{e}_{ij}(\theta_0) + \hat{e}_{ji}(\theta_0)\} \\
&= \frac{2n}{n-1} n^{-3/2} U_w(\theta_0) \\
&\xrightarrow{\mathfrak{D}} N(0, 4\Gamma(\theta_0)).
\end{aligned}$$

□

Lemma D.2. *Under Assumptions 1-5, Let* $\Gamma_n(\theta_0) = 1/n \sum_{i=1}^{n} W_i(\theta_0) W_i^T(\theta_0)$, $\hat{\Gamma}_n(\theta_0) =$

$1/n \sum_{i=1}^{n} \hat{W}_i(\theta_0)\hat{W}_i^T(\theta_0)$. *We have*

$$(i) \quad \Gamma_n(\theta_0) \xrightarrow{P} \Gamma(\theta_0), \tag{D.2}$$

$$(ii) \quad \hat{\Gamma}_n(\theta_0) \xrightarrow{P} \Gamma(\theta_0). \tag{D.3}$$

Proof. For (i), the proof is similar to Lemma D.2 in Zhao (2010). By Cheng et al. (1995) and Fine et al. (1998), we have

$$var(V(\theta_0)) = \frac{4\Gamma(\theta_0)}{n} + o_p(n^{-1}) \ a.s. \tag{D.4}$$

Applying the Strong Law of Large Number for U-statistics, we have

$$V(\theta_0) = O(n^{-1/2}). \tag{D.5}$$

Combining arguments by Lee (1990) and Zhao (2010), we can finish (i).

Note that

$$\begin{aligned} |W_i(\theta_0) - \hat{W}_i(\theta_0)| \leq & \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} \left\{ |\hat{b}(U_i, U_j; \theta_0) - b(U_i, U_j; \theta_0)| \right\} \\ \leq & \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} \{ |e_{ij}(\theta_0) - \hat{e}_{ij}(\theta_0)| \\ & + |e_{ji}(\theta_0) - \hat{e}_{ji}(\theta_0)| + |d_i(\theta_0) - \hat{d}_i(\theta_0)| + |d_j(\theta_0) - \hat{d}_j(\theta_0)| \}. \end{aligned} \tag{D.6}$$

From Gill (1980), we have

$$K_1 = \sup_{0 \leq x \leq X_n} \left| \frac{G(x) - \hat{G}(x)}{G(x)} \right| = o_p(1). \tag{D.7}$$

Define

$$\phi_{ij}(\theta_0) = w_{ij}(\theta_0)\dot{\eta}_{ij}(\theta_0)\delta_j I(\min(X_i, t_0) \geq X_j).$$

Furthermore, under Assumptions 1, 2 and 3, $w_{ij}(\theta)$ and $\dot{\eta}_{ij}(\theta)$ are bounded on compact set Θ. By Zhao (2010), we have that

$$\begin{aligned}|e_{ij}(\theta_0) - \hat{e}_{ij}(\theta_0)| &= w_{ij}(\theta_0)\dot{\eta}_{ij}(\theta_0)\delta_j I\{\min(X_i, t_0) \geq X_j\} \left\{\frac{1}{G^2(X_j)} - \frac{1}{\hat{G}^2(X_j)}\right\} \\ &\leq |w_{ij}(\theta_0)\dot{\eta}_{ij}(\theta_0)\delta_j\{\min(X_i, t_0) \geq X_j\}|(3K_1^2 + 2K_1) \\ &= o_p(1). \qquad \text{(D.8)}\end{aligned}$$

Then,

$$\begin{aligned}|d_i(\theta_0) - \hat{d}_i(\theta_0)| &= 2\left|\int_0^{t_0} \frac{q(\theta_0, t)}{\pi(t)}\, dM_i(t) - \int_0^{t_0} \frac{\hat{q}(\theta_0, t)}{\hat{\pi}(t)}\, d\hat{M}_i(t)\right| \\ &\leq 2\left|\int_0^{t_0} \frac{q(\theta_0, t)}{\pi(t)} - \frac{\hat{q}(\theta_0, t)}{\hat{\pi}(t)}\, dM_i(t)\right| + 2\left|\int_0^{t_0} \frac{\hat{q}(\theta_0, t)}{\hat{\pi}(t)}\, d(M_i(t) - \hat{M}_i(t))\right|. \\ &\qquad \text{(D.9)}\end{aligned}$$

Define

$$q_n(\theta, t) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} w_{ij}(\theta)\dot{\eta}_{ij}(\theta) \frac{\delta_j\{\min(X_i, t_0) \geq X_j\}}{G^2(X_j)} I(X_j \geq t).$$

We have

$$\sup_{0\leq t\leq t_0} \left|\frac{q(\theta_0, t)}{\pi(t)} - \frac{\hat{q}(\theta_0, t)}{\hat{\pi}(t)}\right| \leq \sup_{0\leq t\leq t_0} \left|\frac{q(\theta_0, t)}{\pi(t)} - \frac{q_n(\theta_0, t)}{\hat{\pi}(t)}\right| + \sup_{0\leq t\leq t_0} \left|\frac{q_n(\theta_0, t)}{\hat{\pi}(t)} - \frac{\hat{q}(\theta_0, t)}{\hat{\pi}(t)}\right|. \qquad \text{(D.10)}$$

By the Gilvenko-Cantelli Theorem, $\hat{\pi}(t)$ converges to $\pi(t)$ uniformly on $[0, t_0]$. By the Strong Law of Large Number for U-statistics, $q_n(t)$ converges to $q(t)$ uniformly on $[0, t_0]$. Hence, by the boundness of $\pi(t)$ and $q(t)$,

$$\sup_{0\leq t\leq t_0} \left|\frac{q(\theta_0, t)}{\pi(t)} - \frac{q_n(\theta_0, t)}{\hat{\pi}(t)}\right| \xrightarrow{P} 0. \qquad \text{(D.11)}$$

By Zhao (2010) and Gill (1980) and the boundness of $\dot{\eta}_{ij}(\theta)$ and $w_{ij}(\theta)$,

$$
\begin{aligned}
&\sup_{0\le t\le t_0} |q_n(\theta_0,t) - \hat{q}(\theta_0,t)| \\
&= \sup_{0\le t\le t_0} \left| \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1,j\ne i}^{n} w_{ij}(\theta)\dot{\eta}_{ij}(\theta)\delta_j\{\min(X_i,t_0)\ge X_j\} I(X_j\ge t) \left\{ \frac{1}{G^2(X_j)} - \frac{1}{\hat{G}^2(X_j)} \right\} \right| \\
&\le \sup_{0\le t\le t_0} \left| \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1,j\ne i}^{n} w_{ij}(\theta)\dot{\eta}_{ij}(\theta)\delta_j\{\min(X_i,t_0)\ge X_j\} I(X_j\ge t) \right| (3K_1^2 + 2K_1) \\
&= o_p(1).
\end{aligned}
$$

By the uniform boundness of $\hat{\pi}(t)$

$$
\sup_{0\le t\le t_0} \left| \frac{q_n(\theta_0,t)}{\hat{\pi}(t)} - \frac{\hat{q}(\theta_0,t)}{\hat{\pi}(t)} \right| \xrightarrow{P} 0. \tag{D.12}
$$

From (D.11) and (D.12), we have

$$
\sup_{0\le t\le t_0} \left| \frac{q(\theta_0,t)}{\pi(t)} - \frac{\hat{q}(\theta_0,t)}{\hat{\pi}(t)} \right| \xrightarrow{P} 0. \tag{D.13}
$$

We know $M_i(t) = N_i(t) - \Lambda_i(t)$, where $N_i(t)$ is one jump counting process, i.e., $N_i(t) = I(X_i \le t, \delta_i = 0)$ and $\Lambda_i(t)$ is the corresponding compensator, i.e., $\Lambda_i(t) = \int_0^t I(X_i \ge u)\, d\Lambda_G(u)$, which are uniformly bounded on $[0, t_0]$. Hence,

$$
\begin{aligned}
&\left| \int_0^{t_0} \left\{ \frac{q(\theta_0,t)}{\pi(t)} - \frac{\hat{q}(\theta_0,t)}{\hat{\pi}(t)} \right\} dM_i(t) \right| \\
&\le \left| \int_0^{t_0} \left\{ \frac{q(\theta_0,t)}{\pi(t)} - \frac{\hat{q}(\theta_0,t)}{\hat{\pi}(t)} \right\} dN_i(t) \right| + \left| \int_0^{t_0} \left\{ \frac{q(\theta_0,t)}{\pi(t)} - \frac{\hat{q}(\theta_0,t)}{\hat{\pi}(t)} \right\} I(X_i\ge t)\, d\Lambda_G(t) \right| \\
&\le \sup_{0\le t\le t_0} \left| \frac{q(\theta_0,t)}{\pi(t)} - \frac{\hat{q}(\theta_0,t)}{\hat{\pi}(t)} \right| + \sup_{0\le t\le t_0} \left| \frac{q(\theta_0,t)}{\pi(t)} - \frac{\hat{q}(\theta_0,t)}{\hat{\pi}(t)} \right| \left| \int_0^{t_0} d\Lambda_G(t) \right| \\
&\xrightarrow{P} 0.
\end{aligned} \tag{D.14}
$$

Then, we consider

$$
\begin{aligned}
& d\left\{M_i(t) - \hat{M}_i(t)\right\} \\
&= d\left\{I(X_i \le t, \delta_i = 0) - \int_0^t I(X_i \ge u) d\Lambda_G(u) - I(X_i \le t, \delta_i = 0) + \int_0^t I(X_i \ge u) d\hat{\Lambda}_G(u)\right\} \\
&= I(X_i \ge t) d\{\hat{\Lambda}_G(t) - \Lambda_G(t)\}.
\end{aligned}
$$

From the Martingale Central Limit Theorem, $\sqrt{n}\{\hat{\Lambda}_G(t) - \Lambda_G(t)\}$ converges weakly to a zero-mean Gaussian process on $t \in [0, t_0]$. Hence, $\sup_{0 \le t \le t_0} \left|\hat{\Lambda}_G(t) - \Lambda_G(t)\right| \xrightarrow{P} 0$. By the uniform boundness of $\hat{q}(\theta_0, t)$, $\hat{\pi}(t)$ and the rule of integration by parts,

$$
\begin{aligned}
\left|\int_0^{t_0} \frac{\hat{q}(\theta_0, t)}{\hat{\pi}(t)} d(M_i(t) - \hat{M}_i(t))\right| &= \left|\int_0^{t_0} \frac{\hat{q}(\theta_0, t)}{\hat{\pi}(t)} I(X_i \ge t) d(\hat{\Lambda}_G(u) - \Lambda_G(u))\right| \\
&\xrightarrow{P} 0.
\end{aligned} \tag{D.15}
$$

Thus, combining (D.9), (D.14) and (D.15), we know that

$$
|d_i(\theta_0) - \hat{d}_i(\theta_0)| \xrightarrow{P} 0. \tag{D.16}
$$

From (D.6), (D.8) and (D.16), we obtain that

$$
|b(U_i, U_j, \theta_0) - \hat{b}(U_i, U_j, \theta_0)| \xrightarrow{P} 0 \tag{D.17}
$$

and

$$
|W_i(\theta_0) - \hat{W}_i(\theta_0)| \xrightarrow{P} 0. \tag{D.18}
$$

For any $a \in \mathcal{R}^p$, we have,

$$\begin{aligned}
&a^T \left\{ 1/n \sum_{i=1}^n W_i(\theta_0) W_i^T(\theta_0) - 1/n \sum_{i=1}^n \hat{W}_i(\theta_0) \hat{W}_i^T(\theta_0) \right\} a \\
&= \frac{1}{n} \sum_{i=1}^n \left[a^T \{ W_i(\theta_0) - \hat{W}_i(\theta_0) \} \right]^2 + \frac{2}{n} \sum_{i=1}^n a^T \{ W_i(\theta_0) \} \left[a^T \{ W_i(\theta_0) - \hat{W}_i(\theta_0) \} \right] \\
&= o_p(1). \qquad \text{(D.19)}
\end{aligned}$$

Finally, we prove (ii). □

Lemma D.3. *Under Assumptions 1-5,* $\lambda(\theta_0) = O_p(n^{-1/2})$.

Proof. Write $\lambda(\theta_0) = \rho \upsilon$, where υ is a unit vector. By the Lemma 3 and Corollary 2 of Jing et al. (2008), for the components of U-statistics $|W_i(\theta_0)|$, we have $\max_{i=1,\dots,n} |W_i(\theta_0)| = o(n^{1/2}), \quad a.s.$ From (D.8), we know

$$\begin{aligned}
&|e_{ij}(\theta_0) - \hat{e}_{ij}(\theta_0) + e_{ji}(\theta_0) - \hat{e}_{ji}(\theta_0)| \\
\le& |w_{ij}(\theta_0) \dot{\eta}_{ij}(\theta_0) \delta_j I\{\min(X_i, t_0) \ge X_j\} + w_{ji}(\theta_0) \dot{\eta}_{ji}(\theta_0) \delta_i I\{\min(X_i, t_0) \ge X_j\}| (3K_1^2 + 2K_1) \\
=& |\phi_{ij}(\theta) + \phi_{ji}(\theta)| (3K_1^2 + 2K_1).
\end{aligned}$$

Define $h(U_i, U_j, \theta_0) = |e_{ij}(\theta_0) - \hat{e}_{ij}(\theta_0) + e_{ji}(\theta_0) - \hat{e}_{ji}(\theta_0)|$. By Corollary 2 of Jing et al. (2008), $\max_{i=1,\dots,n} |1/(n-1) \sum_{j=1, j\ne i}^n h(U_i, U_j, \theta_0)| = o(n^{1/2})$. From (D.9),

$$\begin{aligned}
&\max_{i=1,\dots,n} |d_i(\theta_0) - \hat{d}_i(\theta_0)| \\
\le& 2 \max_{i=1,\dots,n} \left| \int_0^{t_0} \frac{q(\theta_0, t)}{\pi(t)} - \frac{\hat{q}(\theta_0, t)}{\hat{\pi}(t)} \, dM_i(t) \right| + 2 \max_{i=1,\dots,n} \left| \int_0^{t_0} \frac{\hat{q}(\theta_0, t)}{\hat{\pi}(t)} \, d(M_i(t) - \hat{M}_i(t)) \right|.
\end{aligned}$$

Denote $y_n(t) = q(\theta_0, t)/\pi(t) - \hat{q}(\theta_0, t)/\hat{\pi}(t)$. We know $M_i(t) = N_i(t) - \Lambda_i(t)$, where $N_i(t) = I(X_i \le t, \delta_i = 0)$ is a counting process and $\Lambda_i(t) = \int_0^t I(X_i \ge u) d\Lambda_G(u)$ is an increasing function of t with uniformly bound on $[0, t_0]$. Then, we will show

$$\max_{i=1,\dots,n} \left| \int_0^{t_0} y_n(t) \, dN_i(t) \right| = o_p(1),$$

i.e., for any $\varepsilon_0 > 0$ $\delta_0 > 0$, exists N_{ε_0},

$$Pr\left\{\max_{i=1,\ldots,n}\left|\int_0^{t_0} y_n(t)\,dN_i(t)\right| \geq \varepsilon_0\right\} \leq \delta_0.$$

By (D.13), for any $\varepsilon_1 > 0$ $\delta_1 > 0$, exists N_{ε_1},

$$Pr\left\{|y_n(t)| \geq \varepsilon_1\right\} \leq \delta_1.$$

Recall that $N_i(t_0)$ has a uniform upper bound 1. Thus, for any $\delta_0 > 0$ and ε_0, $\exists N_{\varepsilon_0} = N_{\varepsilon_1}+1$, we can find $\delta_1 = \delta_0$ and $\varepsilon_1 = 1/2\varepsilon_0$ and obtain that

$$\begin{aligned}
Pr\left\{\max_{i=1,\ldots,n}\left|\int_0^{t_0} y_n(t)\,dN_i(t)\right| \geq \varepsilon_0\right\} \leq & Pr\left\{\max_{i=1,\ldots,n}\int_0^{t_0} |y_n(t)|\ dN_i(t) \geq \varepsilon_0\right\} \\
\leq & Pr\left\{\max_{i=1,\ldots,n}\int_0^{t_0} |y_n(t)|\ dN_i(t) \geq \varepsilon_0, |y_n(t)| \geq \varepsilon_1\right\} \\
+ & Pr\left\{\max_{i=1,\ldots,n}\int_0^{t_0} |y_n(t)|\ dN_i(t) \geq \varepsilon_0, |y_n(t)| < \varepsilon_1\right\} \\
\leq & \delta_1 + Pr\left\{\varepsilon_1 \max_{i=1,\ldots,n}\int_0^{t_0} dN_i(t) \geq \varepsilon_0\right\} \\
\leq & \delta_1 + Pr\left\{\varepsilon_1 \max_{i=1,\ldots,n} N_i(t_0) \geq \varepsilon_0\right\} \\
= & \delta_0.
\end{aligned}$$

Hence, we obtain

$$\left|\max_{i=1,\ldots,n}\int_0^{t_0} y_n(t)\,dN_i(t)\right| = o_p(1).$$

From (D.13) and uniform bounded $\Lambda_G(t)$, we have $\int_0^{t_0} |y_n(t)|\, d\Lambda_G(t) = o_p(1)$.

$$\begin{aligned}
\max_{i=1,\ldots,n} \left| \int_0^{t_0} y_n(t)\, dM_i(t) \right| =& \max_{i=1,\ldots,n} \left| \int_0^{t_0} y_n(t)\, dN_i(t) \right| + \max_{i=1,\ldots,n} \left| \int_0^{t_0} -y_n(t)\, d\Lambda_i(t) \right| \\
\leq& \max_{i=1,\ldots,n} \left\{ \int_0^{t_0} |y_n(t)|\, dN_i(t) \right\} + \max_{i=1,\ldots,n} \int_0^{t_0} |-y_n(t) I(X_i \geq u)|\, d\Lambda_G(t) \\
\leq& \max_{i=1,\ldots,n} \left\{ \int_0^{t_0} |y_n(t)|\, dN_i(t) \right\} + \int_0^{t_0} |y_n(t)|\, d\Lambda_G(t) \\
=& o_p(1).
\end{aligned}$$

We have that $d(M_i(t) - \hat{M}_i(t)) = I(X_i \geq t) d\{\hat{\Lambda}_G(t) - \Lambda_G(t)\}$. By the equation (3.23) of Aalen et al. (2008), $\hat{\Lambda}_G(t) - \Lambda_G(t) = \int_0^{t_0} 1/Y(t)\, dM(t)$, where $Y(t) = \sum_{j=1}^n I(X_j \geq t)$ and $M(t)$ is a martingale, such that $M(t) = N(t) - \Lambda(t)$, where $N(t) = \sum_{t=1}^n I(X_i \leq t, \delta_i = 0)$ is a counting process and $\Lambda(t) = \int_0^t Y(u) d\Lambda_G(u)$ is a compensator. By the Strong Law of Large Number, $Y(t)/n \xrightarrow{P} P(X_i \geq t)$. Because $P(X_i \geq t) > 0$, for any $t \in [0, t_0]$, we have

$$\sup_{t \in [0,t_0]} \left\{ \frac{n}{Y(t)} \right\} = \frac{n}{Y(t_0)} \xrightarrow{P} \frac{1}{P(X_i \geq t_0)}.$$

Then, we have

$$
\begin{aligned}
&\max_{i=1,\dots,n}\left|\int_0^{t_0}\frac{\hat q(\theta_0,t)}{\hat\pi(t)}\,d(M_i(t)-\hat M_i(t))\right|\\
=&\max_{i=1,\dots,n}\left|\int_0^{t_0}\frac{\hat q(\theta_0,t)}{\hat\pi(t)}\,I(X_i\geq t)d(\hat\Lambda_G(t)-\Lambda_G(t))\right|\\
=&\max_{i=1,\dots,n}\left|\int_0^{t_0}\frac{\hat q(\theta_0,t)I(X_i\geq t)}{\hat\pi(t)Y(t)}\,dM(t)\right|\\
=&\max_{i=1,\dots,n}\left|\int_0^{t_0}\frac{\hat q(\theta_0,t)I(X_i\geq t)}{\hat\pi(t)Y(t)}\,dN(t)\right|+\max_{i=1,\dots,n}\left|\int_0^{t_0}-\frac{\hat q(\theta_0,t)I(X_i\geq t)}{\hat\pi(t)Y(t)}\,d\Lambda(t)\right|\\
=&\int_0^{t_0}\max_{i=1,\dots,n}\left|\frac{\hat q(\theta_0,t)I(X_i\geq t)}{\hat\pi(t)Y(t)}\right|\,dN(t)+\int_0^{t_0}\max_{i=1,\dots,n}\left|\frac{\hat q(\theta_0,t)I(X_i\geq t)}{\hat\pi(t)Y(t)}\right|\,d\Lambda(t)\\
=&\int_0^{t_0}\left|\frac{\hat q(\theta_0,t)}{\hat\pi(t)Y(t)}\right|\,dN(t)+\int_0^{t_0}\left|\frac{\hat q(\theta_0,t)}{\hat\pi(t)Y(t)}\right|\,d\Lambda(t)\\
\leq&\frac{n}{Y(t_0)}\int_0^{t_0}\left|\frac{\hat q(\theta_0,t)}{\hat\pi(t)}\right|\,d\left(\frac{N(t)}{n}\right)+\int_0^{t_0}\left|\frac{\hat q(\theta_0,t)}{\hat\pi(t)}\right|\,d\Lambda_G(t)\\
=&\frac{n}{Y(t_0)}\frac{1}{n}\sum_{i=1,X_i<t_0}^{n}\left|\frac{\hat q(\theta_0,X_i)I(\delta_i=0)}{\hat\pi(X_i)}\right|+\int_0^{t_0}\left|\frac{\hat q(\theta_0,t)}{\hat\pi(t)}\right|\,d\Lambda_G(t)\\
\leq&\frac{n}{Y(t_0)}\sup_{i=1,\dots,n}\left|\frac{\hat q(\theta_0,X_i)}{\hat\pi(X_i)}\right|+\int_0^{t_0}\left|\frac{\hat q(\theta_0,t)}{\hat\pi(t)}\right|\,d\Lambda_G(t)\\
\leq&\frac{n}{Y(t_0)}\left\{\sup_{i=1,\dots,n}\left|\frac{q(\theta_0,X_i)}{\pi(X_i)}\right|+C\right\}+\int_0^{t_0}\left|\frac{\hat q(\theta_0,t)}{\hat\pi(t)}\right|\,d\Lambda_G(t),
\end{aligned}
$$

where C is a constant. Because $|q(\theta_0,t)/\pi(t)|$ and $\Lambda_G(t)$ are uniformly bounded and $N(t)/n$ is bounded by 1, we know

$$\max_{i=1,\dots,n}\left|\int_0^{t_0}\frac{\hat q(\theta_0,t)}{\hat\pi(t)}\,d\{M_i(t)-\hat M_i(t)\}\right|=O_p(1)=o_p(\sqrt{n}).$$

Because $\hat\pi(t)$, $\hat q(\theta_0,t)$ and $\Lambda_G(t)$ are uniformly bounded and $N(t)/n$ is bounded by 1. We have

$$\max_{i=1,\dots,n}\left|\int_0^{t_0}\frac{\hat q(\theta_0,t)}{\hat\pi(t)}\,d\{M_i(t)-\hat M_i(t)\}\right|=o_p(\sqrt{n}),$$

and

$$\max_{i=1,\dots,n}|d_i(\theta_0)-\hat d_i(\theta_0)|=o_p(\sqrt{n}).$$

Hence,

$$
\begin{aligned}
&\max_{i=1,\ldots,n} |\hat{W}_i(\theta_0) - W_i(\theta_0)| \\
\leq &\max_{i=1,\ldots,n} \left| \frac{1}{n-1} \sum_{j=1,j\neq i}^{n} h(U_i, U_j, \theta_0) \right| + \max_{i=1,\ldots,n} |d_i(\theta_0) - \hat{d}_i(\theta_0)| + \max_{i=1,\ldots,n} \frac{1}{n-1} \sum_{j=1,j\neq i}^{n} |d_j(\theta_0) - \hat{d}_j(\theta_0)| \\
\leq &\max_{i=1,\ldots,n} \left| \frac{1}{n-1} \sum_{j=1,j\neq i}^{n} h(U_i, U_j, \theta_0) \right| + 2 \max_{i=1,\ldots,n} |d_i(\theta_0) - \hat{d}_i(\theta_0)| \\
= &o_p(n^{1/2}).
\end{aligned}
$$

Combining above equations, we have

$$
\max_{i=1,\ldots,n} |\hat{W}_i(\theta_0)| \leq \max_{i=1,\ldots,n} |\hat{W}_i(\theta_0) - W_i(\theta_0)| + \max_{i=1,\ldots,n} |W_i(\theta_0)| = o_p(n^{1/2}). \tag{D.20}
$$

By Owen (2001, p. 220),

$$
\rho \left[\upsilon^T \hat{\Gamma}_n(\theta_0) \upsilon - \left\{ \max_{i=1,\ldots,n} \hat{W}_i(\theta_0) \right\} \left\{ n^{-1} \sum_{i=1}^{n} \upsilon^T \hat{W}_i(\theta_0) \right\} \right] \leq n^{-1} \sum_{i=1}^{n} \upsilon^T \hat{W}_i(\theta_0). \tag{D.21}
$$

By Lemma D.2, we have

$$
\upsilon^T \hat{\Gamma}_n(\theta_0) \upsilon \leq \upsilon^T \{\hat{\Gamma}_n(\theta_0) - \Gamma\} \upsilon + \upsilon^T \Gamma \upsilon = O_p(1). \tag{D.22}
$$

By Lemma D.1, we have

$$
n^{-1} \sum_{i=1}^{n} \hat{W}_i(\theta_0) = O_p(n^{-1/2}). \tag{D.23}
$$

So, by (D.20-23), we have

$$
0 < \rho\{O_p(1) + o_p(1)\} \leq O_p(n^{-1/2}),
$$

and

$$\lambda(\theta_0) = \rho \upsilon = o_p(1)\upsilon = O_p(n^{-1/2}). \tag{D.24}$$

□

Proof of Theorem 6.1 By Owen (2001),

$$\lambda(\theta_0) = \left\{ \sum_{i=1}^{n} \hat{W}_i(\theta_0)\hat{W}_i^T(\theta_0) \right\}^{-1} \left\{ \sum_{i=1}^{n} \hat{W}_i(\theta_0) \right\} + o_p(n^{-1/2}). \tag{D.25}$$

By the Taylor expansion for (D.10), we have

$$l(\theta_0) = \sum_{i=1}^{n} \lambda(\theta_0)^T \hat{W}_i(\theta_0) + o_p(1). \tag{D.26}$$

Combining (D.25) and (D.26), we have

$$\begin{aligned} l(\theta_0) &= \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{W}_i(\theta_0) \right\}^T \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{W}_i(\theta_0)\hat{W}_i^T(\theta_0) \right\}^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{W}_i(\theta_0) \right\} + o_p(1) \\ &\xrightarrow{D} 4\chi_p^2. \end{aligned}$$

□

Proof of Theorem 6.2 The proof is along the lines of Proposition 3 of Yu et al. (2011). Note that $\theta_0 = (\theta_{10}^T, \theta_{20}^T)^T$ and the corresponding $(Z_1^T, Z_2^T)^T$. $\tilde{\eta}_{ij}(\theta_0)$ is the partial derivative of $\eta_{ij}(\theta_0)$ with respect to θ_2, where

$$\tilde{\eta}_{ij}(\theta_0) = Z_{2j} \int_{-\infty}^{\alpha_0} \{1 - F(t + Z_i^T\beta_0)\}\, df(t + Z_j^T\beta_0) - Z_{2i} \int_{-\infty}^{\alpha_0} \{1 - f(t + Z_i^T\beta_0)\}\, dF(t + Z_j^T\beta_0). \tag{D.27}$$

and $\tilde{D}(\theta_0) = \lim_{n\to\infty} n^{-2} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} w_{ij}(\theta_0)\tilde{\eta}_{ij}(\theta_0)\tilde{\eta}_{ij}^T(\theta_0)$, where $\tilde{\eta}_{ij}(\theta)$ is a $p+1-q$ dimensional vector. Denote $\hat{\theta}_2 = \arg\inf_{\theta_2} l(\theta_{10}, \theta_2)$. Let $\tilde{\Phi}(\theta_0) = \tilde{D}(\theta_0)^{-1}\Gamma(\theta_0)\tilde{D}(\theta_0)^{-1}$. By similar arguments in Qin and Lawless (1994) and Fine et al. (1998), we can obtain

$$\sqrt{n}(\hat{\beta}_2 - \beta_{20}) = -\tilde{\Phi}^{-1}(\theta_0)\tilde{D}(\theta_0)^T\Gamma(\theta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{W}_i(\theta_0) + o_p(1),$$

and the Lagrange multiplier λ_2 satisfies that

$$\sqrt{n}\lambda_2 = \left\{I - \Gamma^{-1}(\theta_0)\tilde{D}(\theta_0)\tilde{\Phi}^{-1}(\theta_0)\tilde{D}(\theta_0)^T\right\}\Gamma(\theta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{W}_i(\theta_0) + o_p(1).$$

Thus,

$$\begin{aligned}
&\frac{1}{4}l^*(\theta_{10}) \\
&= \left\{\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\hat{W}_i(\theta_0)\right\}^T\left\{\Gamma^{-1}(\theta_0) - \Gamma^{-1}(\theta_0)\tilde{D}(\theta_0)\tilde{\Phi}^{-1}(\theta_0)\tilde{D}(\theta_0)^T\Gamma^{-1}(\theta_0)\right\}\left\{\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\hat{W}_i(\theta_0)\right\} + o_p(1) \\
&= \left\{\Gamma^{-1/2}(\theta_0)\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\hat{W}_i(\theta_0)\right\}^T\left\{I - \Gamma^{-1/2}(\theta_0)\tilde{D}(\theta_0)\tilde{\Phi}^{-1}(\theta_0)\tilde{D}(\theta_0)^T\Gamma^{-1/2}(\theta_0)\right\} \\
&\quad \left\{\Gamma^{-1/2}(\theta_0)\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\hat{W}_i(\theta_0)\right\} + o_p(1).
\end{aligned}$$

Define $S = I - \Gamma^{-1/2}(\theta_0)\tilde{D}(\theta_0)\tilde{\Phi}^{-1}(\theta_0)\tilde{D}(\theta_0)^T\Gamma^{-1/2}(\theta_0)$. Note that $trace(S) = q$. We have

$$\frac{1}{4}l^*(\theta_{10}) \xrightarrow{D} \chi^2_q. \qquad \square$$