Georgia State University
ScholarWorks @ Georgia State University

Mathematics Dissertations | Department of Mathematics and Statistics

6-12-2012

Statistical Evaluation of Continuous-Scale Diagnostic Tests with Missing Data

Binhuan Wang
Georgia State University

Follow this and additional works at: http://scholarworks.gsu.edu/math_diss

Recommended Citation
Wang, Binhuan, "Statistical Evaluation of Continuous-Scale Diagnostic Tests with Missing Data." Dissertation, Georgia State University, 2012.
http://scholarworks.gsu.edu/math_diss/8

This Dissertation is brought to you for free and open access by the Department of Mathematics and Statistics at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Mathematics Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

STATISTICAL EVALUATION OF CONTINUOUS-SCALE DIAGNOSTIC TESTS WITH MISSING DATA

by

BINHUAN WANG

Under the Direction of Dr. Gengsheng Qin

ABSTRACT

The receiver operating characteristic (ROC) curve methodology is the statistical methodology for assessment of the accuracy of diagnostics tests or bio-markers. Currently most widely used statistical methods for the inferences of ROC curves are complete-data based parametric, semi-parametric or nonparametric methods. However, these methods cannot be used in diagnostic applications with missing data. In practical situations, missing diagnostic data occur more commonly due to various reasons such as medical tests being too

expensive, too time consuming or too invasive. This dissertation aims to develop new nonparametric statistical methods for evaluating the accuracy of diagnostic tests or biomarkers in the presence of missing data. Specifically, novel nonparametric statistical methods will be developed with different types of missing data for (i) the inference of the area under the ROC curve (AUC, which is a summary index for the diagnostic accuracy of the test) and (ii) the joint inference of the sensitivity and the specificity of a continuous-scale diagnostic test. In this dissertation, we will provide a general framework that combines the empirical likelihood and general estimation equations with nuisance parameters for the joint inferences of sensitivity and specificity with missing diagnostic data. The proposed methods will have sound theoretical properties. The theoretical development is challenging because the proposed profile log-empirical likelihood ratio statistics are not the standard sum of independent random variables. The new methods have the power of likelihood based approaches and jackknife method in ROC studies. Therefore, they are expected to be more robust, more accurate and less computationally intensive than existing methods in the evaluation of competing diagnostic tests.

INDEX WORDS: AUC, Bootstrap, Diagnostic tests, Empirical likelihood, Estimating equations, Imputation, Jackknife, Missing data, ROC curve, Sensitivity, Specificity, Verification bias

STATISTICAL EVALUATION OF CONTINUOUS-SCALE DIAGNOSTIC TESTS WITH MISSING DATA

by

BINHUAN WANG

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

Copyright by
Binhuan Wang
2012

STATISTICAL EVALUATION OF CONTINUOUS-SCALE DIAGNOSTIC TESTS WITH MISSING DATA

by

BINHUAN WANG

Committee Chair: Dr. Gengsheng Qin

Committee: Dr. Ruiyan Luo
Dr. Xin Qi
Dr. Ying Zhu

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
August 2012

DEDICATION

To my parents, my advisor,

and all my dear friends.

ACKNOWLEDGEMENTS

These printed pages of this dissertation contain not only cumulative years of study and research, but also help and encouragement from many generous and inspiring people I have met since the beginning of my doctorate career at the Department of Mathematics and Statistics, Georgia State University. I'd like to acknowledge and dedicate my gratitude to the following people who have made the completion of this dissertation possible.

I'm deeply indebted to my advisor Dr. Gengsheng Qin. Dr. Gengsheng Qin helped me to discover my potentials and agreed to be my mentor in the PhD program with concentration in biostatistics. I cannot imagine how much time I would have wasted or how much longer I would have struggled without his help. Dr. Qin spent a lot of time and effort to advise me on how to be a better student, a more successful instructor and a more efficient researcher. Dr. Qin is always available whenever I needed his help, and he is willing to have casual conversations with me and inspired me with his own experiences. With his patient and enlightening directions, my research ran more smoothly so that I could finish this dissertation.

Also, I'd like to express my special gratitude to Dr. Yixin Fang, an energetic young researcher. As a professor in our department, he always encouraged me to be self-motivated and reminded me not to be complacent with small advances. We had many discussions on statistics, which guided me on what are the next steps to take. When I was planning to carry on further research career, he provided me with a great opportunity to work as a researcher at the renowned New York University.

My special thanks also go to Dr. Huazhen Lin, my advisor in Sichuan University. After graduating from college, I was at a crossroad where I have to choose the next path to take. I was fortunate enough to meet Dr. Lin, who guided me into the field of statistics and introduced many inspiring people in statistics. With her help and encouragements, I discovered my interests in about statistics, and decided to pursue my career in the field of statistics.

I also would like to express my gratitude to my dissertation committee members, Dr. Ruiyan Luo, Dr. Xin Qi and Dr. Ying Zhu. Reviewing the manuscript is very time-consuming, and they spent much precious time on correcting mistakes, and provided many insightful suggestions to improve the dissertation.

Additionally, I wish to show my great appreciation to all the professors who have taught and helped me, Dr. Yu-Sheng Hsu, Dr. Guantao Chen, Dr. Changyong Zhong, Dr. Jiawei Liu, Dr. Jun Han, Dr. Renato Monteiro, Dr. Richard Luger, Dr. Richard Rothenberg, Dr. Xu Zhang, Dr. Yichuan Zhao, Dr. Yuanhui Xiao, and Dr. Zhongshan Li.

I'm forever indebted to my beloved parents, Caili Wang and Kaihua Wu, for their ceaseless love, understanding, patience, and support. They have always accompanied me and given me strength and encouragement whenever I needed. At this moment, I believe they would be very proud of me for the completion of this dissertation. Also, I would love to express my heartfelt gratitude to my cousin, Ruya Zhao, who read my dissertation and corrected many grammar mistakes.

My life would not be so fantastic in Atlanta without my dearest friends, Meng Zhao, Shuang Liu, Zi Zhu, Haochuan Zhou, Hongwei Wang, Xin Huang, Yichao Yin, Zhengbo Ma, Zhibo Wang, Garrick Brim, Amy Fomo, Dong Yang, Fei Gong, Huayu Liu, Shan Luo, Shuman Guo, Ye Cui, Xiaoxi Wei, Xiaoxue Gao, Xue Qi, Yueheng An, and many others. I would never forget the happy moments we shared and the many wise suggestions my friends have made.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . . **v**

LIST OF TABLES . . . . . **xi**

LIST OF FIGURES . . . . . **xv**

LIST OF ABBREVIATIONS . . . . . **xviii**

CHAPTER 1 INTRODUCTION **1**

1.1 Receiver Operating Characteristic Curve 1

1.2 Missing Data 3

1.3 Methodology 5

1.3.1 **Empirical Likelihood Method and Its Application to the ROC curve and Estimating Equations** 5

1.3.2 **Random Hot Deck Imputation** 7

1.3.3 **Bias-corrected ROC Curve** 9

1.4 Aims of the Dissertation 11

1.5 Significance 13

1.6 Organization of the Dissertation 14

CHAPTER 2 EMPIRICAL LIKELIHOOD-BASED CONFIDENCE INTERVALS FOR THE SENSITIVITY OF A CONTINUOUS-SCALE DIAGNOSTIC TEST WITH MISSING COMPLETELY AT RANDOM DATA **15**

2.1 Introduction 15

2.2 Imputation-based Empirical Likelihood for the Sensitivity with MCAR Data 15

2.2.1 Missing Data 16
2.2.2 **Profile Empirical Likelihood for the Sensitivity** 18
2.2.3 **Imputation-based Empirical Likelihood Intervals for the Sensitivity with MCAR Data** 21
2.3 Simulation Studies 24
2.4 An Illustrate Example 26

CHAPTER 3 IMPUTATION-BASED EMPIRICAL LIKELIHOOD INFERENCE FOR THE AREA UNDER THE ROC CURVE WITH MISSING COMPLETELY AT RANDOM DATA 34
3.1 Introduction 34
3.2 Imputation-based Empirical Likelihood for the AUC 34
3.2.1 **Empirical Likelihood for the AUC** 35
3.2.2 **Imputation-based Empirical Likelihood Interval for the AUC** 36
3.2.3 **Imputation-based EL Intervals for the AUC with Stratified Samples** 39
3.3 Simulation Studies 41
3.4 A Real Example 42

CHAPTER 4 JOINT EMPIRICAL LIKELIHOOD CONFIDENCE REGIONS FOR THE EVALUATION OF CONTINUOUS-SCALE DIAGNOSTIC TESTS WITH MISSING COMPLETELY AT RANDOM DATA **48**
4.1 Introduction 48
4.2 Imputation-based Bivariate Empirical Likelihood Confidence Regions with MCAR Data 49
4.2.1 **Bivariate Nonparametric Confidence Regions with Complete Data** 49

4.2.2 **Imputation-based Bivariate Empirical Likelihood Confidence Regions** 51

4.3 Simulation Studies 52

CHAPTER 5 EMPIRICAL LIKELIHOOD CONFIDENCE REGIONS FOR THE EVALUATION OF CONTINUOUS SCALE DIAGNOSTIC TEST IN THE PRESENCE OF VERIFICATION BIAS **58**

5.1 Introduction 58

5.2 Bias-corrected Empirical Likelihood Confidence Regions 59

5.2.1 **Empirical Likelihood and General Estimation Equations with Nuisance Parameters** 59

5.2.2 **Bias-corrected Empirical Likelihood Confidence Regions** . 63

5.3 Simulation Studies 72

5.3.1 **Correct Models** 73

5.3.2 **Misspecified Models** 74

5.4 Study of Depression in Elderly Patients Recruited from Primary-care Practices 75

CHAPTER 6 JACKKNIFE EMPIRICAL LIKELIHOOD CONFIDENCE REGIONS FOR THE EVALUATION OF CONTINUOUS SCALE DIAGNOSTIC TEST WITH VERIFICATION BIAS **88**

6.1 Introduction 88

6.2 Jackknife Empirical Likelihood Confidence Regions in the Presence of Verification Bias 90

6.3 Simulation Studies 96

6.3.1 **Correct Models** 96

6.3.2 **Misspecified Models** 98

6.4 Study of Neonatal Hearing Screening Data 100

CHAPTER 7 DISCUSSION AND FUTURE WORK **109**

REFERENCES **111**

APPENDICES **117**

Appendix A PROOFS OF CHAPTER 2 . . . . . **117**

Appendix B PROOFS OF CHAPTER 3 . . . . . **122**

Appendix C PROOFS OF CHAPTER 4 . . . . . **128**

Appendix D PROOFS OF CHAPTER 5 . . . . . **129**

LIST OF TABLES

Table 2.1 Model setting (1): Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 90% 28

Table 2.2 Model setting (1): Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 95% 29

Table 2.3 Model setting (2): Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 90% 30

Table 2.4 Model setting (2): Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 95% 31

Table 2.5 Model setting (1) with observed data only: Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 90% and 95% 32

Table 2.6 Model setting (2) with observed data only: Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 90% and 95% 33

Table 2.7 A real example: 95% IPEL and IHBEL confidence intervals for $R(p)$ of CA19-9 at different levels of specificities with various observation rates. 33

Table 3.1 Model setting (1): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 90% and various observation rates (π_1, π_2) 44

Table 3.2 Model setting (1): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 95% and various observation rates (π_1, π_2) 45

Table 3.3 Model setting (2): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 90% and various observation rates (π_1, π_2) 46

Table 3.4 Model setting (2): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 95% and various observation rates (π_1, π_2) 47

Table 3.5 A real example: 95% IPEL confidence intervals for the AUC of CA19-9 with various observation rates. 47

Table 4.1 Model setting (1): Coverage probabilities of the confidence regions obtained by the empirical likelihood statistic $\tilde{l}(\theta, \eta, \tau)$ with nominal confidence level 90% at various observation rates (π_1, π_2) in the presence of the MCAR data. 54

Table 4.2 Model setting (1): Coverage probabilities of the confidence regions obtained by the empirical likelihood statistic $\tilde{l}(\theta, \eta, \tau)$ with nominal confidence level 95% at various observation rates (π_1, π_2) in the presence of the MCAR data. 55

Table 4.3 Model setting (2): Coverage probabilities of the confidence regions obtained by the empirical likelihood statistic $\tilde{l}(\theta, \eta, \tau)$ with nominal confidence level 90% at various observation rates (π_1, π_2) in the presence of the MCAR data. 56

Table 4.4 Model setting (2): Coverage probabilities of the confidence regions obtained by the empirical likelihood statistic $\tilde{l}(\theta, \eta, \tau)$ with nominal confidence level 95% at various observation rates (π_1, π_2) in the presence of the MCAR data. 57

Table 5.1 Correct models: Coverage probabilities of various joint empirical likelihood confidence regions with nominal confidence level 90% in the presence of verification bias. Pre. means disease prevalence; EL(F) stands for the proposed method based on weighted chi-squared method; EL(R) stands for the reduced method from Corollary 2; NA means the normality approximation method. 80

Table 5.2 Correct models: Coverage probabilities of various joint empirical likelihood confidence regions with nominal confidence level 95% in the presence of verification bias. Pre. means disease prevalence; EL(F) stands for the proposed method based on weighted chi-squared method; EL(R) stands for the reduced method from Corollary 2; NA means the normality approximation method. 81

Table 5.3 Misspecified disease models: Coverage probabilities of SPE-based joint empirical likelihood confidence regions with nominal confidence levels 90% and 95% in the presence of verification bias. Pre. means disease prevalence; EL(F) stands for the proposed method based on weighted chi-squared method; EL(R) stands for the reduced method from Corollary 2; NA means the normality approximation method. 82

Table 5.4 Misspecified verification models: Coverage probabilities of SPE-based joint empirical likelihood confidence regions with nominal confidence levels 90% and 95% in the presence of verification bias. Pre. means disease prevalence; EL(F) stands for the proposed method based on weighted chi-squared method; EL(R) stands for the reduced method from Corollary 1; NA means the normality approximation method. 82

Table 6.1 Correct models: Coverage probabilities of various jackknife empirical likelihood confidence regions with nominal confidence levels 90% and 95% in the presence of verification bias. Pre. means disease prevalence. 103

Table 6.2 Misspecified disease models: Coverage probabilities of SPE-based joint empirical likelihood confidence regions with nominal confidence levels 90% and 95% in the presence of verification bias. Pre. means disease prevalence. 104

Table 6.3 Misspecified verification models: Coverage probabilities of SPE-based joint empirical likelihood confidence regions with nominal confidence levels 90% and 95% in the presence of verification bias. Pre. means disease prevalence. 104

LIST OF FIGURES

Figure 5.1 Left panel: contour curves of $\widehat{l}_{SPE}(\theta, 0.75, \tau)$, giving confidence regions for the pair (cut-off level, sensitivity) at the fixed specificity level of 0.75. Right panel: contour curves of $\widehat{l}_{SPE}(0.75, \eta, \tau)$, giving confidence regions for the pair (cut-off level, specificity) at the fixed sensitivity level of 0.75. Contours in both panels correspond to nominal confidence levels 90%, 95% and 99%. 83

Figure 5.2 Left panel: contour curves of $\widehat{l}_{MSI}(\theta, 0.75, \tau)$, giving confidence regions for the pair (cut-off level, sensitivity) at the fixed specificity level of 0.75. Right panel: contour curves of $\widehat{l}_{MSI}(0.75, \eta, \tau)$, giving confidence regions for the pair (cut-off level, specificity) at the fixed sensitivity level of 0.75. Contours in both panels correspond to nominal confidence levels 90%, 95% and 99%. 83

Figure 5.3 Left panel: contour curves of $\widehat{l}(\theta, 0.75, \tau)$ from full data, giving confidence regions for the pair (cut-off level, sensitivity) at the fixed specificity level of 0.75. Right panel: contour curves of $\widehat{l}(0.75, \eta, \tau)$ from full data, giving confidence regions for the pair (cut-off level, specificity) at the fixed sensitivity level of 0.75. Contours in both panels correspond to nominal confidence levels 90%, 95% and 99%. 84

Figure 5.4 In red, the 95% SPE-based confidence region for the pair (cut-off level, sensitivity) at the fixed 0.75 level of specificity; In blue, the 95% SPE-based confidence region for the pair (cut-off level, specificity,) at the fixed 0.75 level of sensitivity. 84

Figure 5.5 In red, the 95% MSI-based confidence region for the pair (cut-off level, sensitivity) at the fixed 0.75 level of specificity; In blue, the 95% MSI-based confidence region for the pair (cut-off level, specificity,) at the fixed 0.75 level of sensitivity. 85

Figure 5.6 In red, the 95% confidence region for the pair (cut-off level, sensitivity) at the fixed 0.75 level of specificity with full data; In blue, the 95% confidence region for the pair (cut-off level, specificity,) at the fixed 0.75 level of sensitivity with full data. 85

Figure 5.7 Contour curves of $\widehat{l}_{SPE}(\theta, \eta, 8)$, offering the confidence regions for the pair (specificity, sensitivity), at different nominal coverage levels 90%, 95% and 99%. 86

Figure 5.8 Contour curves of $\widehat{l}_{MSI}(\theta, \eta, 8)$, offering the confidence regions for the pair (specificity, sensitivity), at different nominal coverage levels 90%, 95% and 99%. 86

Figure 5.9 Contour curves of $\widehat{l}(\theta, \eta, 8)$ from full data, offering the confidence regions for the pair (specificity, sensitivity), at different nominal coverage levels 90%, 95% and 99%. 87

Figure 6.1 Left panel: contour curves of $l^J_{\mathrm{IPW}}(\theta, 0.6, \tau)$, offering confidence regions for the pair (cut-off level, sensitivity) at the fixed specificity level of 0.6. Right panel: contour curves of $l^J_{\mathrm{IPW}}(0.6, \eta, \tau)$, offering confidence regions for the pair (cut-off level, specificity) at the fixed sensitivity level of 0.6. Contours in both panels correspond to nominal confidence levels 90%, 95% and 99%. 105

Figure 6.2 Left panel: contour curves of $l^J_{\text{MSI}}(\theta, 0.6, \tau)$, offering confidence regions for the pair (cut-off level, sensitivity) at the fixed specificity level of 0.6. Right panel: contour curves of $l^J_{\text{MSI}}(0.6, \eta, \tau)$, offering confidence regions for the pair (cut-off level, specificity) at the fixed sensitivity level of 0.6. Contours in both panels correspond to nominal confidence levels 90%, 95% and 99%. 106

Figure 6.3 IPW-based joint confidence regions: in red, the 95% confidence region for the pair (cut-off level, sensitivity) at the fixed 0.6 level of specificity; In blue, the 95% confidence region for the pair (cut-off level, specificity,) at the fixed 0.6 level of sensitivity. 107

Figure 6.4 MSI-based joint confidence regions: in red, the 95% confidence region for the pair (cut-off level, sensitivity) at the fixed 0.6 level of specificity; In blue, the 95% confidence region for the pair (cut-off level, specificity,) at the fixed 0.6 level of sensitivity. 107

Figure 6.5 Contour curves of $l^J_{\text{IPW}}(\theta, \eta, -4)$, offering the confidence regions for the pair (specificity, sensitivity) when the cut-off level τ is fixed at -4, at nominal coverage levels 90%, 95% and 99%. 108

Figure 6.6 Contour curves of $l^J_{\text{MSI}}(\theta, \eta, -4)$, offering the confidence regions for the pair (specificity, sensitivity) when the cut-off level τ is fixed at -4, at nominal coverage levels 90%, 95% and 99%. 108

LIST OF ABBREVIATIONS

- EL - Empirical likelihood
- JEL - Jackknife empirical likelihood
- SEL - Smoothed empirical likelihood
- HBEL - Hybrid empirical likelihood
- IPEL - Imputation-based profile empirical likelihood
- IHBEL - Imputation-based hybrid empirical likelihood
- EL(F) regions - Profile empirical likelihood confidence regions
- EL(R) regions - Reduced profile empirical likelihood confidence regions
- ROC curve - Receiver operating characteristic curve
- AUC - Area under the curve
- WMW - Wilcoxon-Mann-Whitney
- MCAR - Missing completely at random
- MAR - Missing at random
- MNAR - Missing not at random
- FI - Full imputation
- MSI - Mean score imputation
- IPW - Inverse probability weighting
- SPE - Semiparametric efficient estimator

- GEE - Generalized estimating equations
- HAM-D - Hamilton depression rating scale
- SCID - Structured clinical interview for DSM-IV
- CIRS - Cumulative illness rating scale
- INHI - Identification of neonatal hearing impairment
- DPOAE - Distorting product otoacoustic emissions
- TEOAE - Transient evoked otoacoustic emissions
- VRA - Visual reinforcement audiometry

CHAPTER 1

INTRODUCTION

1.1 Receiver Operating Characteristic Curve

In medical studies, diagnostic tests are widely used to detect the occurrence of a disease, and to monitor the disease progression. The sensitivity and specificity are common measures used to evaluate the performance of a diagnostic test. For a continuous-scale test, the diagnosis is dependent upon whether the test result is above or below a specified cut-off point. Let X and Y be results of a continuous-scale test for a non-diseased and a diseased subject, and we assume that F and G are cumulative distribution functions of X and Y, respectively. For a given cut-off level τ, the sensitivity (true positive rate), denoted by θ, and the specificity (true negative rate), denoted by η, of the test are defined by

$$\theta(\tau) = P(Y > \tau) = 1 - G(\tau), \text{ and } \eta(\tau) = P(X \leq \tau) = F(\tau). \tag{1.1}$$

Alternatively, if we use a common notation T as the test result for both diseased and non-diseased groups, and let D be the disease indicator with 1 as a diseased subject and 0 as a non-diseased subject, then the sensitivity and the specificity can be written as follows:

$$\theta(\tau) = P(T > \tau | D = 1), \text{ and } \eta(\tau) = P(T \leq \tau | D = 0). \tag{1.2}$$

When the cut-off level τ varies throughout the entire real line, the resulting plot of sensitivity against 1-specificity is called the *Receiver Operating Characteristic* (ROC) curve. In practice, we are always interested in a test which has a higher sensitivity at a fixed level of specificity. When a specificity of the test is p $(0 < p < 1)$, the corresponding sensitivity of the test is

$$R(p) = 1 - G(F^{-1}(p)), \tag{1.3}$$

where F^{-1} is the inverse function of F. The plot $\{(1-p, R(p)) : 0 < p < 1\}$ is also the ROC curve.

The area under the curve (AUC), defined as $\delta = \int_0^1 R(p)\mathrm{d}p$, is a commonly used summary measure of the ROC curve. AUC has been frequently used to assess the ability of a diagnostic test to discriminate between individuals with and without a disease. Bamber [1] showed that the AUC, $\delta = P(Y \geq X)$, which can be interpreted as the probability that in a randomly selected pair of diseased and non-diseased subjects, the test value of the diseased subject is higher than or equal to that of the non-diseased subject. In a more general context, Wolfe and Hogg [2] recommended the use of this index as a general measure for the difference between two distributions. One important problem for the inference on the AUC is how to construct a confidence interval for δ. Let $X_1, \ldots, X_m$ be test results of a random sample of non-diseased subjects and $Y_1, \ldots, Y_n$ be test results of a random sample of diseased subjects. Traditionally, the classical Wilcoxon-Mann-Whitney (WMW) [3] two-sample rank statistic, defined by

$$\delta_{m,n} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I(Y_j \geq X_i), \tag{1.4}$$

is employed as a nonparametric estimator of the AUC. Based on the asymptotic normality of the WMW statistic, we can construct a confidence interval (hereafter WMW interval) for the AUC. Although the WMW estimator of the AUC is known to be unbiased, the normal approximation-based WMW interval suffers from low coverage accuracy for high values of the AUC (e.g., 0.90 to 0.95, which are of most interest in diagnostic tests) when sample sizes of diseased and non-diseased subjects are small and unequal.

Recently, Adimari and Chiogna [4] considered joint inferences on both the (specificity, cut-off level) and the (sensitivity, cut-off level). Joint confidence regions depict the association of sensitivity, specificity and cut-off level for a continuous-scale test. By visually inspecting confidence regions, one can select a reasonable cut-off level τ in order to obtain a desirable sensitivity $\theta(\tau)$ and an acceptable specificity $\eta(\tau)$ simultaneously, because it is

well known that there is a trade-off between the sensitivity and the specificity. Moreover, by constructing joint confidence regions, one can investigate the within-pair relationship of (θ, τ) or (η, τ), respectively. In a diagnostic study, the AUC is a widely used summary index of the diagnostic accuracy. However it can not be used to select a cut-off level because the AUC masks the effect of cut-off level.

Extensive studies have been done in literature on estimating the ROC curve in cases of complete data. For more details of the ROC curve, we refer readers to Metz [5], Swets and Pickett [6], Pepe [7], and Metz, Herman and Shen [8]. Linnet [9] proposed both parametric and non-parametric methods for constructing confidence intervals for the sensitivity of a test at a fixed value of specificity. Hsieh and Turnbull [10] estimated the ROC curve by replacing F and G by their corresponding empirical distribution. Zou *et al.* [11] and Lloyd [12] suggested smoothing kernel estimators for $R(p)$. Pepe [13] and Zhou *et al.* [14] reviewed many statistical methods for the evaluation of diagnostic tests. Currently most widely used statistical methods for inferences of ROC curves are parametric, semi-parametric, or nonparametric methods with complete data. However, these methods cannot be used in diagnostic applications with missing data directly.

1.2 Missing Data

In making statistical inferences, samples are usually assumed to be complete. However, due to various reasons, missing data instead of complete data occur commonly in practical situations. Rubin [15] and Little and Rubin [16] classified missing data into three categories based on missing mechanisms: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). MCAR and MAR are considered more often in literature.

MCAR means the missing mechanism is independent of both observable variables and unobservable variables. For example, patients involved in a regular blood or urine test in a medical diagnosis would quit the research because they move to other districts, or miss visits to hospitals due to bad weather or schedule conflicts. Regarding these situations, these

kinds of missingness are unrelated to any patients' characteristics. This class of missingness could be assumed to be MCAR.

MAR means missing mechanism only depends on observed data. In medical diagnostics, we hope, based on screening test results, the true disease status for every subject can be verified by applying a gold standard evaluation, which assesses the disease status with certainty. However, in many situations, not all subjects that are given their screening test results ultimately have their true disease statuses verified. There are various reasons that account for this occurrence. For example, gold standard tests may be too expensive, too time consuming, or too invasive. In these situations, subjects with positive test results are more likely to take a gold standard evaluation than the subjects with negative test results. Thus, estimates of accuracy, like the sensitivity and the specificity, can be biased in studies with such designs because the decision of whether or not to verify the subject's true disease status depends on their test results. This bias is called verification bias [17] or work-up bias [18]. In fact, verification biased data could be assumed to be one type of MAR data, i.e. the probability that a subject has the disease status verified only depends on the test result and the subject's observed characteristics. Direct application of complete data inference procedures to MAR problems may produce biased estimation and lose efficiency.

Various methods [16], including imputation-based methods, have been proposed in order to handle problems caused by missing data. Zhou ([19], [20]), Hunink *et al.* [21], and Rodenberg and Zhou [22] proposed bias-corrected methods for binary and ordinal tests. Geert *et al.* [23] evaluated five different methods in dealing with missing values in the empirical data from a study among patients suspected of pulmonary embolism, and they found that imputation is relatively better than others.

To correct verification bias, Alonzo and Pepe [24] broadened Begg and Greenes (BG) method [17] from binary test cases to continuous test cases. In addition, Alonzo and Pepe [24] proposed several imputation and reweighting bias-corrected methods (reviewed by Carroll et al. [25]), including the inverse probability weighting (IPW) estimator, the full imputation (FI) method, the mean score imputation (MSI) method, and the semi-parametric efficient

estimator (SPE) of the sensitivity and the specificity as well as the AUC. But there is no closed-form expression for the variance of the AUC estimator based on their methods. Therefore, resampling methods are needed for inference. Rotnitzky *et al.* [26] proposed a doubly robust estimator of the AUC under both MAR and MNAR assumptions. Later, He *et al.* [27] provided a direct estimate of the AUC in the presence of verification bias, and Fluss *et al.* [28] investigated the properties of the doubly robust method for estimating the ROC curve under verification bias. Liu and Zhou [29] proposed a semi-parametric estimation method for covariate-specific ROC curves with a partial missing gold standard. Long *et al.* [30] developed robust statistical methods for estimating the AUC, and the proposed methods used information from auxiliary variables that are potentially predictive of the missingness of the biomarkers or the missing biomarker values.

1.3 Methodology

Missing diagnostic data bring challenges to inferences of the ROC curve. The first challenge is how to estimate the ROC curve and the AUC with missing data. The second challenge is how to construct confidence intervals/regions for the ROC curve and its related quantities like sensitivity, specificity and AUC with missing data. To solve these problems, the proposed new methodologies will involve some modern statistical techniques such as empirical likelihood method, jackknife and bootstrap methods.

1.3.1 Empirical Likelihood Method and Its Application to the ROC curve and Estimating Equations

Empirical likelihood (EL, see [31], [32], [33]) is a nonparametric method traditionally used for providing confidence intervals or regions for the mean, without assuming distributions of underlying populations. Empirical likelihood-based statistical methods have many good properties, such as good small sample performance, automatically data determined confidence regions for unknown parameters of interest. When diagnostic data are complete, empirical likelihood could be applied to construct confidence intervals/regions for the ROC

curve and its related quantities. Claesken *et al.* [34] developed the smoothed empirical likelihood (SEL) method for $R(p)$. By using jackknife technique, Gong, Peng and Qi [42] proposed the smoothed jackknife empirical likelihood method for the ROC curve. Due to the difficulty of bandwidth selection, Qin, Davis and Jing [36] proposed the hybrid empirical likelihood intervals (HBEL) for $R(p)$, which does not involve in the selection of bandwidth. Chen and Van Keilegom [37] provided a general review on empirical likelihood method for regressions. Based on the consideration of computation, Zhou and Qin [38], and Horváth *et al.* [39] proposed bootstrap intervals for the ROC curve.

For the inference on the AUC with the empirical likelihood method, following research has been done. Based on the mean-like form of WMW estimator, Qin and Zhou [40] proposed an EL approach for the inference on the AUC, which was shown to have good small sample performance. Motivated by the asymptotic independence of pseudo-values from the jackknife technique, Jing, Yuan and Zhou [41] introduced the jackknife empirical likelihood (JEL) method for U-statistics, and used the AUC as an example to illustrate their method because the WMW estimator is a two-sample U-statistics.

Empirical likelihood method has also been applied to estimating equations. Qin and Lawless [43] linked estimating equations and empirical likelihood, and developed methods of combining information about parameters. Hjort *et al.* [44] extended the scope of general empirical likelihood methodology by introducing plug-in estimates of nuisance parameters in estimating equations. But there are no explicit asymptotic results on the empirical likelihood defined by general estimation equations with nuisance parameters, which are estimated by another set of estimating equations. Li *et al.* [45] proposed a jackknife EL method to construct confidence regions for interesting parameters in the presence of nuisance parameters being simply replaced by some estimators under general estimating equation framework. With jackknife pseudo samples generated, the resulting jackknife EL method retains the attractive chi-square limiting distribution. In order to reduce the computation in the jackknife empirical likelihood method when explicit estimators of nuisance parameters are not available, Peng [46] proposed an approximate jackknife empirical likelihood method.

However, current empirical likelihood framework cannot be directly used with missing diagnostic data. Various methods [16], including imputation-based methods, have been proposed in order to handle problems caused by missing data. Wang and Rao [47] developed EL-based confidence intervals for the mean of the response variable using kernel regression imputation, and Wang and Rao [48] also constructed intervals for the mean of the response variable in a linear model with missing data. Liang and Zhou [49] developed smoothed empirical likelihood-based confidence intervals for ROC curves when samples are censored and generated from semi-parametric models. Qin and Qian [50] constructed confidence intervals for the differences of quantiles with missing data. Wang and Chen [51] applied empirical likelihood to estimating equations with missing data based on a nonparametric imputation of missing values from a kernel estimator of the conditional distribution of the missing variable given always observable variables. Qin *et al.* [52] proposed a unified empirical likelihood approach to missing data problems and explored the use of empirical likelihood to effectively combine unbiased estimating equations when the number of estimating equations is greater than the number of unknown parameters.

1.3.2 Random Hot Deck Imputation

In this part, we will review the random hot deck imputation method.

Let $(X_1, \delta_{X_1}), \ldots, (X_m, \delta_{X_m})$ and $(Y_1, \delta_{Y_1}), \ldots, (Y_n, \delta_{Y_n})$ be simple random sample sequences of incomplete data associated with the populations (X, δ_X) and (Y, δ_Y) respectively, where

$$X \sim F, \quad \delta_{X_i} = \begin{cases} 0, & \text{if } X_i \text{ is missing} \\ 1, & \text{if } X_i \text{ is observed} \end{cases}, \ i = 1, \ldots, m,$$

and

$$Y \sim G, \quad \delta_{Y_j} = \begin{cases} 0, & \text{if } Y_j \text{ is missing} \\ 1, & \text{if } Y_j \text{ is observed} \end{cases}, \ j = 1, \ldots, n.$$

We assume both F and G are absolutely continuous for mathematical consideration.

We assume X and Y are MCAR, i.e.,

$$P(\delta_X = 1|X) = \pi_1 \text{ and } P(\delta_Y = 1|Y) = \pi_2,$$

where both π_1 and π_2 are constants belonging to $(0, 1)$.

For convenience, some standard notations are needed. Let $r_X = \sum_{i=1}^{m} \delta_{X_i}$, $r_Y = \sum_{j=1}^{n} \delta_{Y_j}$, $m_X = m - r_X$ and $m_Y = n - r_Y$. Denote the sets of observed data with respect to X and Y as S_{r_X} and S_{r_Y} respectively, and the sets of missing data with respect to X and Y as S_{m_X} and S_{m_Y} respectively. Then the means of the observed data with respect to X and Y are denoted as $\bar{X}_r = \frac{1}{r_X} \sum_{i \in S_{r_X}} X_i$ and $\bar{Y}_r = \frac{1}{r_Y} \sum_{j \in S_{r_Y}} Y_j$, respectively. Furthermore, let X_i^* and Y_j^* be the imputed values for the missing data with respect to X and Y, respectively.

Imputation methods are useful in dealing with missing data. With MCAR type of data, we prefer the random hot deck imputation method to impute missing values rather than the deterministic imputation, because the latter one is not appropriate in making inference of distribution functions [53]. The idea of random hot deck imputation [54] is natural. From S_{r_X}, the random hot deck imputation draws a simple random sample of size m_X with replacement, and then let $X_i^* = X_k$ for some $k \in S_{r_X}$. Therefore, a sample of so-called "complete data" after imputation can be obtained as

$$\tilde{X}_i = \delta_{X_i} X_i + (1 - \delta_{X_i}) X_i^*, i = 1, \ldots, m.$$

Similarly, the imputed "complete data" from S_{r_Y} could be obtained as

$$\tilde{Y}_j = \delta_{Y_j} Y_j + (1 - \delta_{Y_j}) Y_j^*, j = 1, \ldots, n.$$

1.3.3 Bias-corrected ROC Curve

In this part, we review current bias-correction methods in the presence of verification bias. Let T_i denote the continuous test result from a screening test, and let D_i denote the binary disease status without measurement error, $i = 1, \ldots, n$, where $D_i = 1$ indicates the ith patient is diseased and $D_i = 0$ indicates the ith patient is free of disease. Due to various causes, such as cost limits and privacy security, only a subset of patients have their disease statuses verified; let V_i denote the binary verification status of the ith patients, with $V_i = 1$ if the ith patient has the true disease status verified, and $V_i = 0$ if otherwise. In practice, some covariate information, other than the results from the screening test, can be obtained. Let A_i be a vector of observed covariates for the ith patient that may be associated with both D_i and V_i.

When all patients are verified, i.e., $V_i = 1, i = 1, \ldots, n$, a complete data set is obtained. In this case, for any cut-off level τ, the sensitivity $\theta(\tau)$, and the specificity $\eta(\tau)$ can be estimated by

$$\widehat{\theta}_{Full}(\tau) = \frac{\sum_{i=1}^{n} I(T_i > \tau) D_i}{\sum_{i=1}^{n} D_i}, \quad \widehat{\eta}_{Full}(\tau) = \frac{\sum_{i=1}^{n} I(T_i \leq \tau)(1 - D_i)}{\sum_{i=1}^{n} (1 - D_i)}. \tag{1.5}$$

Obviously, $\widehat{\theta}_{Full}(\tau)$ and $\widehat{\eta}_{Full}(\tau)$ are unbiased estimators for θ and η respectively.

Many current studies center on the MAR assumption because it is manageable in practice. Under this assumption, whether one subject has his or her disease status verified is conditionally independent of the true disease status given the test result and the observed covariates, i.e., $V \perp D|T, A$ or $P(V|D, T, A) = P(V|T, A)$ or $P(D|V, T, A) = P(D|T, A)$. In other words, the decision to verify the patient's true disease status only depends on T and A regardless of the true disease status D.

Alonzo and Pepe [24] extended the method proposed by Begg and Greenes [17] on discrete data and saturated model to the bias-corrected ROC problem with continuous data T and A. They used full imputation (FI) over the distribution $P(D|T, A)$ to estimate the

prevalence of disease. The corresponding FI estimators of $\theta(\tau)$ and $\eta(\tau)$ are given as

$$\widehat{\theta}_{FI}(\tau) = \frac{\sum_{i=1}^{n} I(T_i > \tau)\widehat{\rho}_i}{\sum_{i=1}^{n} \widehat{\rho}_i}, \quad \widehat{\eta}_{FI}(\tau) = \frac{\sum_{i=1}^{n} I(T_i \leq \tau)(1-\widehat{\rho}_i)}{\sum_{i=1}^{n}(1-\widehat{\rho}_i)}, \tag{1.6}$$

where $\widehat{\rho}_i$ is an estimator of $\rho_i = P(D_i = 1|T_i, A_i)$ that is obtained by using, for example, probit models only on the data from verified patients.

Based on the recognition that the study with verification-biased sampling could be thought of as the study with a two-phase or double-sampling design ([55], [56]), Alonzon *et al.* [57] proposed the mean score imputation (MSI) approach for estimating the prevalence of disease. Compared with the FI method, the MSI method estimates $P(D_i|T_i, A_i)$ by using D_i from verified subjects, the same as that in the FI method, then only imputes D_i for those who are not in the verification sample. The resulting estimators of $\theta(\tau)$ and $\eta(\tau)$ are defined as follows:

$$\begin{aligned} \widehat{\theta}_{MSI}(\tau) &= \frac{\sum_{i=1}^{n} I(T_i > \tau)(V_i D_i + (1-V_i)\widehat{\rho}_i)}{\sum_{i=1}^{n}(V_i D_i + (1-V_i)\widehat{\rho}_i)}, \\ \widehat{\eta}_{MSI}(\tau) &= \frac{\sum_{i=1}^{n} I(T_i \leq \tau)(V_i(1-D_i) + (1-V_i)(1-\widehat{\rho}_i))}{\sum_{i=1}^{n}(V_i(1-D_i) + (1-V_i)(1-\widehat{\rho}_i))}, \end{aligned} \tag{1.7}$$

where $\widehat{\rho}_i$ is defined as that in the FI method.

An inverse probability weighting (IPW) estimator weighs each observation in the verified sample by the inverse of the verification probability to correct selection bias. This estimator provides another approach that is used to estimate the prevalence of disease in a two-phase design. Let $\pi_i = P(V_i = 1|T_i, A_i)$. By given the estimated weight $\widehat{\pi}_i^{-1}$ to each verified subject, the inverse of the estimated probability that the subject was verified, the estimators of $\theta(\tau)$ and $\eta(\tau)$ are defined as follows:

$$\widehat{\theta}_{IPW}(\tau) = \frac{\sum_{i=1}^{n} I(T_i > \tau)V_i D_i \widehat{\pi}_i^{-1}}{\sum_{i=1}^{n} V_i D_i \widehat{\pi}_i^{-1}}, \quad \widehat{\eta}_{IPW}(\tau) = \frac{\sum_{i=1}^{n} I(T_i \leq \tau)V_i(1-D_i)\widehat{\pi}_i^{-1}}{\sum_{i=1}^{n} V_i(1-D_i)\widehat{\pi}_i^{-1}}, \tag{1.8}$$

where $\widehat{\pi}_i$ is an estimator of $P(V_i = 1|T_i, A_i)$ that is obtained by using, for example, logistic regression from all patients in the sample.

Gao *et al.* [58] and Alonzo *et al.* [57] independently derived the following semiparametric efficient estimator (SPE) of the prevalence of disease in two-phase studies, which deserves the so-called "double robustness" property in the sense that it is consistent if either π_i or ρ_i is estimated consistently:

$$\widehat{P}(D=1) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{V_i D_i}{\widehat{\pi}_i} - \frac{(V_i - \widehat{\pi}_i)\widehat{\rho}_i}{\widehat{\pi}_i}\right),$$

where $\widehat{\pi}_i$ and $\widehat{\rho}_i$ are the same as those defined previously. Following this approach, Alonzo and Pepe [24] proposed the following SPE estimators for $\theta(\tau)$ and $\eta(\tau)$:

$$\begin{aligned}\widehat{\theta}_{SPE}(\tau) &= \frac{\sum_{i=1}^{n} I(T_i > \tau)[V_i D_i - (V_i - \widehat{\pi}_i)\widehat{\rho}_i]\widehat{\pi}_i^{-1}}{\sum_{i=1}^{n}[V_i D_i - (V_i - \widehat{\pi}_i)\widehat{\rho}_i]\widehat{\pi}_i^{-1}},\\ \widehat{\eta}_{SPE}(\tau) &= \frac{\sum_{i=1}^{n} I(T_i \le \tau)[V_i(1-D_i) - (V_i - \widehat{\pi}_i)(1-\widehat{\rho}_i)]\widehat{\pi}_i^{-1}}{\sum_{i=1}^{n}[V_i(1-D_i) - (V_i - \widehat{\pi}_i)(1-\widehat{\rho}_i)]\widehat{\pi}_i^{-1}}.\end{aligned} \tag{1.9}$$

When τ varies throughout the real line, each of above methods provides an empirical bias-corrected ROC curve by using the pair $(1-\widehat{\eta}, \widehat{\theta})$. Alonzo *et al.* [57] proposed a common estimating equation framework to derive consistency and asymptotic normality results for two-phase disease prevalence estimators that accounts for the uncertainty in estimating nuisance parameters corresponding to the estimation of $P(D=1|T,A)$ and/or $P(V=1|T,A)$.

1.4 Aims of the Dissertation

Based on all reviews above, to our knowledge, not much study has been done on the evaluation of continuous-scale diagnostic tests with missing Data. In this dissertation, we consider the inference of ROC curves with two types of missing data, MCAR data and data in the presence of verification under the MAR assumption. We aim to develop new nonparametric statistical methods for evaluating the accuracy of diagnostic tests or biomarkers in the presence of missing data. Specifically, novel nonparametric statistical methods will be developed with different types of missing data for (i) the inference of the area under the ROC curve and (ii) the joint inference of sensitivity and specificity of a continuous-scale

diagnostic test.

With MCAR data, we propose to use the random hot deck imputation [54] method to impute missing values. The idea of random hot deck imputation method is natural: drawing a simple random sample with replacement from the observed data to impute missing data, and then applying empirical likelihood method to the imputed "complete data". Based on the imputed "complete data", inference on ROC curves with the empirical likelihood method could be made.

With MAR data, we will provide a general framework that combines empirical likelihood and general estimation equations with nuisance parameters for joint inferences of the sensitivity and the specificity. We propose to rewrite the empirical estimates of sensitivity, specificity and AUC as the solutions of empirical estimating functions with nuisance parameter. Here verification (or missing) probability and the prevalence of disease are treated as nuisance parameters, but these parameters can be consistently estimated by employing parametric models such as logistic regressions or probit models. Then we can define profile empirical likelihoods for sensitivity and specificity as well as AUC. The resulting profile log-empirical likelihood ratio statistics can be given explicitly. We show that asymptotic distributions of these profile empirical log-likelihood ratio statistics are weighted sums of independent chi-squared distributions. If either disease models or verification models are miss-specified, one of the proposed bias-corrected empirical likelihood methods still performs well with moderate sample size cases ($n \geq 400$). Therefore, the profile empirical log-likelihood ratio statistics can be used as pivotals to construct confidence intervals for the AUC and joint confidence regions for sensitivity and specificity. Furthermore, in order to reduce computation burden of estimating weights of the asymptotic distributions of profile empirical log-likelihood ratio statistics, jackknife technique is applied to construct pseudo samples, and standard chi-squared distributions are retained for jackknife empirical likelihood ratio statistics, which are easier to apply in practice.

The proposed empirical likelihood-based joint confidence regions provide a graphical tool to select a cut-off level which yields the desirable sensitivity and/or specificity by plotting

joint confidence regions for (θ, τ) and (η, τ) in the same graph. Reasonable cut-off levels could be directly identified from the overlapping part of the two regions. Such visual tool is straightforward and easy to implement in practice. It is necessary to point out that the proposed confidence regions preserve many good properties of empirical likelihood method, such as good small sample performance, data determined confidence regions and range-respecting, which could be a problem for normal-approximation based confidence regions.

Simulation studies are conducted to evaluate the finite sample performance of all proposed methods. Additionally, all new methods will also be applied to some real data sets in medical diagnostics to show their practical meanings.

1.5 Significance

In this dissertation, various bias-corrected empirical likelihood confidence intervals for the sensitivity of ROC curves, the AUC and joint confidence regions for the sensitivity and the specificity with missing data are proposed. These confidence regions could provide a good solution to the problem of selecting a reasonable cut-off point for a continuous-scale diagnostic test. The research will make significant contributions to medical diagnostic tests, and will greatly extend the scope of the applications of empirical likelihood methods. The efficacy of the proposed inference procedures will be demonstrated via simulations and empirical applications. The results should be very useful in assessing diagnostic tests because the costs of diagnostic tests can be very high. To select more accurate diagnostic tests for wider use, it is important to develop appropriate statistical methods for evaluating the diagnostic accuracy of competing tests. The use of attractive statistical methods like these proposed in this dissertation for the ROC curve analysis will help diagnostic test users make informed choices of the most reliable diagnostic tests. This dissertation will certainly contribute to the reduction of health care costs in the long run.

1.6 Organization of the Dissertation

The remainder of this dissertation is organized as follows. In Chapter 2 and 3, we present imputation-based empirical likelihood intervals for the sensitivity and the AUC with MCAR data, respectively. Chapter 4 shows imputation-based bivariate empirical likelihood confidence regions with MCAR data. Empirical likelihood confidence regions for the evaluation of continuous-scale diagnostics test in the presence of verification bias are shown in Chapter 5. In Chapter 6, the jackknife technique is applied to the problem in Chapter 5 in order to simplify the calculation. A brief discussion is provided in Chapter 7.

CHAPTER 2

EMPIRICAL LIKELIHOOD-BASED CONFIDENCE INTERVALS FOR THE SENSITIVITY OF A CONTINUOUS-SCALE DIAGNOSTIC TEST WITH MISSING COMPLETELY AT RANDOM DATA

2.1 Introduction

In this chapter, an imputation-based profile empirical likelihood (IPEL) and an imputation-based hybrid empirical likelihood (IHBEL) are proposed to construct confidence intervals for the sensitivity with missing data. The proposed methods preserve the advantage of the method in Qin, Davis and Jing [36], which is free of bandwidth selection, and the advantage of the random hot deck imputation method [54], which preserves the distribution of item values whereas the deterministic imputation methods like the ratio imputation and the regression imputation do not have this appealing property. Both IPEL and IHBEL intervals are easy to apply in practice.

The remainder of this chapter is organized as follows. Section 2.2 presents the imputation-based empirical likelihood method to construct confidence intervals for sensitivity with missing data. In section 2.3, we conduct simulation studies to assess the performance of the proposed methods. In section 2.4, we illustrate the proposed imputation-based empirical likelihood intervals with a real example. All proofs are deferred until the Appendix A.

2.2 Imputation-based Empirical Likelihood for the Sensitivity with MCAR Data

In this section, we aim to construct empirical likelihood-based confidence intervals for $R(p)$ with missing data. We firstly impute the missing data by the random hot deck imputation technique, and then review the profile empirical likelihood method proposed by Qin, Davis and Jing [36]. Finally we develop the imputation-based empirical likelihood method

for $R(p)$.

2.2.1 Missing Data

Let $(X_1, \delta_{X_1}), \ldots, (X_m, \delta_{X_m})$ and $(Y_1, \delta_{Y_1}), \ldots, (Y_n, \delta_{Y_n})$ be simple random sample sequences of incomplete data associated with the populations (X, δ_X) and (Y, δ_Y) respectively, where

$$X \sim F, \quad \delta_{X_i} = \begin{cases} 0, & \text{if } X_i \text{ is missing} \\ 1, & \text{if } X_i \text{ is observed} \end{cases}, \ i = 1, \ldots, m,$$

and

$$Y \sim G, \quad \delta_{Y_j} = \begin{cases} 0, & \text{if } Y_j \text{ is missing} \\ 1, & \text{if } Y_j \text{ is observed} \end{cases}, \ j = 1, \ldots, n.$$

We assume both F and G are absolutely continuous for mathematical consideration.

Throughout this chapter, we assume X and Y are MCAR, i.e.,

$$P(\delta_X = 1|X) = \pi_1 \text{ and } P(\delta_Y = 1|Y) = \pi_2,$$

where both π_1 and π_2 are constants belonging to $(0, 1)$.

For convenience, some standard notations are needed. Let $r_X = \sum_{i=1}^m \delta_{X_i}$, $r_Y = \sum_{j=1}^n \delta_{Y_j}$, $m_X = m - r_X$ and $m_Y = n - r_Y$. Denote the sets of observed data with respect to X and Y as S_{r_X} and S_{r_Y} respectively, and the sets of missing data with respect to X and Y as S_{m_X} and S_{m_Y} respectively. Then the means of the observed data with respect to X and Y are denoted as $\bar{X}_r = \frac{1}{r_X} \sum_{i \in S_{r_X}} X_i$ and $\bar{Y}_r = \frac{1}{r_Y} \sum_{j \in S_{r_Y}} Y_j$, respectively. Furthermore, let X_i^* and Y_j^* be the imputed values for the missing data with respect to X and Y, respectively.

Imputation methods are useful in dealing with missing data. With MCAR type of data, we prefer the random hot deck imputation method to impute missing values rather than

the deterministic imputation, because the latter one is not appropriate in making inference of distribution functions [53]. The idea of random hot deck imputation [54] is natural. From S_{r_X}, the random hot deck imputation draws a simple random sample of size m_X with replacement, and then let $X_i^* = X_k$ for some $k \in S_{r_X}$. Therefore, a sample of so-called "complete data" after imputation can be obtained as

$$\tilde{X}_i = \delta_{X_i} X_i + (1 - \delta_{X_i}) X_i^*, i = 1, \ldots, m.$$

Similarly, the imputed "complete data" from S_{r_Y} could be obtained as

$$\tilde{Y}_j = \delta_{Y_j} Y_j + (1 - \delta_{Y_j}) Y_j^*, j = 1, \ldots, n.$$

We could prove that based on the imputed data $\tilde{X}_i$'s and $\tilde{Y}_j$'s, the empirical distributions

$$\widetilde{F}(x) = \frac{1}{m} \sum_{i=1}^{m} I(\tilde{X}_i \leq x) \tag{2.1}$$

and

$$\widetilde{G}(y) = \frac{1}{n} \sum_{j=1}^{n} I(\tilde{Y}_j \leq y) \tag{2.2}$$

are still consistent and asymptotically normal.

Proposition 1 *$\widetilde{F}(x)$ and $\widetilde{G}(y)$ defined above are uniformly consistent estimates for $F(x)$ and $G(y)$ respectively. Furthermore, they are asymptotically normal, i.e.,*

$$\sqrt{m}(\widetilde{F}(x) - F(x)) \xrightarrow{d} \mathcal{N}(0, \sigma_X^2) \tag{2.3}$$

where $\sigma_X^2 = (1 - \pi_1 + \pi_1^{-1}) F(x)(1 - F(x))$, and

$$\sqrt{n}(\widetilde{G}(y) - G(y)) \xrightarrow{d} \mathcal{N}(0, \sigma_Y^2) \tag{2.4}$$

where $\sigma_Y^2 = (1 - \pi_2 + \pi_2^{-1})G(y)(1 - G(y))$.

If we only use complete observations without applying the random hot deck imputation, we could obtain following results. Define

$$\widetilde{F}^*(x) = \frac{1}{r_X} \sum_{i \in S_{r_X}} I(X_i \leq x) \quad \text{and} \quad \widetilde{G}^*(y) = \frac{1}{r_Y} \sum_{j \in S_{r_Y}} I(Y_j \leq y).$$

Corollary 1 *$\widetilde{F}^*(x)$ and $\widetilde{G}^*(y)$ defined above are uniformly consistent estimates for $F(x)$ and $G(y)$ respectively. Furthermore, they are asymptotically normal, i.e.,*

$$\sqrt{m}(\widetilde{F}^*(x) - F(x)) \xrightarrow{d} \mathcal{N}(0, \sigma_X^{*2}) \tag{2.5}$$

where $\sigma_X^{*2} = \pi_1^{-1}F(x)(1 - F(x))$, *and*

$$\sqrt{n}(\widetilde{G}^*(y) - G(y)) \xrightarrow{d} \mathcal{N}(0, \sigma_Y^{*2}) \tag{2.6}$$

where $\sigma_Y^{*2} = \pi_2^{-1}G(y)(1 - G(y))$.

The results from Corollary 1 and Proposition 1 are slightly different. Without the random hot deck imputation, some terms in Proposition 1 are absent in Corollary 1. Actually, it is equivalent to disregard missing data and apply methods based on complete data to observed data only. For comparison purpose, we will list results based on observed data only in simulation studies.

2.2.2 Profile Empirical Likelihood for the Sensitivity

The sensitivity corresponding to a given specificity p is $R(p) = P(Y \geq F^{-1}(p))$, thus it could be estimated by

$$\widehat{R}(p) = \frac{\sum_{j=1}^{n} I(Y_j \geq F_m^{-1}(p))}{n}, \tag{2.7}$$

where F_m is the empirical distribution function based on X_i's, i.e., $F_m(x) = \frac{1}{m}\sum_{i=1}^{m} I(X_i \leq x)$. Denote the empirical distribution function of G by $G_n(y) = \frac{1}{n}\sum_{i=1}^{n} I(Y_j \leq y)$. Then,

$$\widehat{R}(p) = 1 - G_n(F_m^{-1}(p)). \tag{2.8}$$

Gastwirth [59], and Chakraborti and Mukerjee [60] showed that for a fixed p, $0 < p < 1$,

$$\sqrt{n}(G_n(F_m^{-1}(p)) - G(F^{-1}(p))) \xrightarrow{d} \mathcal{N}(0, \sigma_0^2(p)), \tag{2.9}$$

where $\sigma_0^2(p) = R(p)(1 - R(p)) + \kappa p(1-p)\frac{g^2(F^{-1}(p))}{f^2(F^{-1}(p))}$, and $\kappa = \lim_{m,n\to\infty} \frac{n}{m}$, a fixed quantity, with f and g denoting the density functions of F and G.

By substituting unknown quantities in (2.9) by their corresponding sample estimates, a $(1-\alpha)100\%$ normal approximation-based confidence interval could be constructed. However, this confidence interval could be greatly affected by poor density and quantile estimation, mentioned by Platt, Hanley and Yang [61].

In order to obtain better confidence intervals for $R(p)$, Qin, Davis and Jing [36] proposed a profile empirical likelihood for the sensitivity. Their method is easy to be applied in practice because it does not involve in the selection of bandwidth, which is crucial in the smoothed empirical likelihood based method (SEL), proposed by Claeskens *et al.* [34]. The selection of an optimal bandwidth is always problematic in practical situations and is still an open problem.

For a given test value Y from a diseased subject, let $U = 1 - F(Y)$. The value U can be interpreted as the proportion of the non-diseased population with test values greater than Y. It is easy to obtain the following equality:

$$E(I(U \leq 1-p)) = P(F(Y) \geq p) = P(Y \geq F^{-1}(p)) = R(p).$$

The introduction of U converts the original problem into a mean-based problem. That is, the estimation of the sensitivity $R(p)$ is equivalent to the estimation of the expectation of an

indicator function of U. Based on this relationship between $R(p)$ and the placement value U, an empirical likelihood procedure for the inference of the sensitivity has been derived [36]. Let $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ be a probability vector, i.e., $\sum_{j=1}^{n} p_j = 1$ and $p_j \geq 0$ for all j. The empirical likelihood for $R(p)$ is defined as follows:

$$L(R(p)) = \sup \left\{ \prod_{j=1}^{n} p_j : \sum_{j=1}^{n} p_j = 1, \sum_{j=1}^{n} p_j W_j(p) = 0 \right\},$$

where $W_j(p) = I(U_j \leq 1 - p) - R(p)$ with $U_j = 1 - F(Y_j)$, $j = 1, 2, \ldots, n$. $L(R(p))$ can not be found because it involves in an unknown nuisance parameter F that is the distribution function of the non-diseased population. After plugging-in the empirical estimate F_m for F in $L(R(p))$, a profile empirical likelihood (PEL) for $R(p)$ can be obtained:

$$\widehat{L}(R(p)) = \sup \left\{ \prod_{j=1}^{n} p_j : \sum_{j=1}^{n} p_j = 1, \sum_{j=1}^{n} p_j \widehat{W}_j(p) = 0 \right\}, \tag{2.10}$$

where $\widehat{W}_j(p) = I(\widehat{U}_j \leq 1 - p) - R(p)$ with $\widehat{U}_j = 1 - F_m(Y_j)$, $j = 1, 2, \ldots, n$. By the standard procedure of empirical likelihood method, the empirical likelihood ratio for $R(p)$ could be defined as follows:

$$\widehat{r}(R(p)) = \prod_{j=1}^{n} \left\{ 1 + \widehat{\lambda} \widehat{W}_j(p) \right\}^{-1},$$

where $\widehat{\lambda}$ is the solution of

$$\frac{1}{n} \sum_{j=1}^{n} \frac{\widehat{W}_j(p)}{1 + \widehat{\lambda} \widehat{W}_j(p)} = 0. \tag{2.11}$$

Then the corresponding log-EL ratio is

$$\widehat{l}(R(p)) \equiv -2 \log \widehat{r}(R(p)) = 2 \sum_{j=1}^{n} \log \left\{ 1 + \widehat{\lambda} \widehat{W}_j(p) \right\}. \tag{2.12}$$

Qin, Davis and Jing [36] proved that the limiting distribution of $l(R(p))$ is a scaled chi-

square distribution. They proposed a hybrid bootstrap empirical likelihood (HBEL) interval for $R(p)$ based on the limiting distribution.

2.2.3 Imputation-based Empirical Likelihood Intervals for the Sensitivity with MCAR Data

Based on the imputed data $\tilde{X}_i$'s and $\tilde{Y}_j$'s, we can substitute all complete data X_i's and Y_j's in the previous section and obtain the similar log-EL ratio for $R(p)$ as follows:

$$\widetilde{l}(R(p)) = 2\sum_{j=1}^{n} \log\left\{1 + \widetilde{\lambda}\widetilde{W}_j(p)\right\}. \tag{2.13}$$

where $\widetilde{W}_j(p) = I(\widetilde{U}_j \leq 1-p) - R(p)$ with $\widetilde{U}_j = 1 - \widetilde{F}(\tilde{Y}_j)$, $j = 1, 2, \ldots, n$, and $\widetilde{\lambda}$ is the solution of

$$\frac{1}{n}\sum_{j=1}^{n} \frac{\widetilde{W}_j(p)}{1 + \widetilde{\lambda}\widetilde{W}_j(p)} = 0. \tag{2.14}$$

In order to present the asymptotic distribution of the log-EL ratio for sensitivity with missing data, some modifications of previous results are needed. (2.9) is based on complete data, but it could be extended to the missing data situation based on the random hot deck imputation. The result is stated in the following proposition.

Proposition 2 *The random variable* $\sqrt{n}(\widetilde{G}(\widetilde{F}^{-1}(p)) - G(F^{-1}(p)))$ *is asymptotically normally distributed:*

$$\sqrt{n}(\widetilde{G}(\widetilde{F}^{-1}(p)) - G(F^{-1}(p))) \xrightarrow{d} \mathcal{N}(0, \sigma_1^2(p)), \tag{2.15}$$

where $\widetilde{F}$ *and* $\widetilde{G}$ *are defined by (2.1) and (2.2) respectively,* $\sigma_1^2(p) = R(p)(1-R(p))(1-\pi_2+\pi_2^{-1}) + \kappa p(1-p)\frac{g^2(F^{-1}(p))}{f^2(F^{-1}(p))}(1-\pi_1+\pi_1^{-1})$*, and* $\kappa = \lim_{m,n\to\infty}\frac{n}{m} < \infty$*, with* f *and* g *denoting the density functions of* F *and* G*.*

Based on all the previous work, the following theorem establishes the asymptotic distribution of the log-EL ratio for the sensitivity with missing data.

Theorem 1 *Assume that the distribution functions F and G are both continuous with density functions f and g respectively. If $\lim_{m,n\to\infty} \frac{n}{m} = \kappa < \infty$, and $0 < R(p) < 1$ for $0 < p < 1$, then the asymptotic distribution of $\widetilde{l}(R(p))$, defined by (2.13), is a scaled χ^2 distribution with degree of freedom one:*

$$c(p)\widetilde{l}(R(p)) \xrightarrow{d} \chi_1^2, \tag{2.16}$$

where the scale constant $c(p)$ is $c(p) = \frac{\sigma^2(p)}{\sigma_1^2(p)}$ with

$$\begin{aligned}
\sigma^2(p) &= R(p)(1 - R(p)),\\
\sigma_1^2(p) &= \sigma^2(p)(1 - \pi_2 + \pi_2^{-1}) + \kappa p(1-p)\frac{g^2(F^{-1}(p))}{f^2(F^{-1}(p))}(1 - \pi_1 + \pi_1^{-1}).
\end{aligned}$$

The confidence interval for $R(p)$ could be constructed based on Theorem 1 by plugging in consistent estimates of all unknown quantities. Let

$$\begin{aligned}
\widetilde{\sigma}^2(p) &= \widetilde{R}(p)(1 - \widetilde{R}(p)),\\
\widetilde{\sigma}_1^2(p) &= \widetilde{\sigma}^2(p)(1 - \widehat{\pi}_2 + \widehat{\pi}_2^{-1}) + \widehat{\kappa} p(1-p)\frac{\widetilde{g}^2(\widetilde{F}^{-1}(p))}{\widetilde{f}^2(\widetilde{F}^{-1}(p))}(1 - \widehat{\pi}_1 + \widehat{\pi}_1^{-1}),
\end{aligned}$$

where $\widetilde{R}(p) = 1 - \widetilde{G}(\widetilde{F}^{-1}(p))$, $\widehat{\kappa} = \frac{n}{m}$, $\widehat{\pi}_1 = \frac{r_X}{m}$, $\widehat{\pi}_2 = \frac{r_Y}{n}$, $\widetilde{F}^{-1}(p)$ is the p -th sample quantile of $\tilde{X}_i$'s, $\widetilde{f}$ and $\widetilde{g}$ are kernel density estimates of f and g, respectively. Here the kernel density estimation method provided by the R package KS [62] is employed to obtain $\widetilde{f}$ and $\widetilde{g}$. The R package KS implements diagonal and unconstrained data-driven bandwidth matrices for kernel density estimation. Therefore, a $(1-\alpha)100\%$ imputation-based profile empirical likelihood confidence interval for $R(p)$, denoted by IPEL interval, is defined as follows:

$$CI_{1,\alpha}(R(p)) = \{R(p) : \widetilde{c}(p)\widetilde{l}(R(p)) \le \chi_1^2(1-\alpha)\}, \tag{2.17}$$

where $\widetilde{c}(p) = \frac{\widetilde{\sigma}^2(p)}{\widetilde{\sigma}_1^2(p)}$.

It is clear that the performance of the IPEL interval depends on the density estimates $\widetilde{f}$ and $\widetilde{g}$. If the sample size is not large enough or the missing probability is high, the performance of density estimation may not be good, especially for high specificity case. In order to solve this problem, the bootstrap method, which is powerful in small sample case, could be employed to construct an imputation-based hybrid bootstrap and empirical likelihood (IHBEL) ratio confidence interval for $R(p)$. From Proposition 2, $\sigma_1^2(p)$ is the asymptotic variance of $\sqrt{n}(\widetilde{R}(p) - R(p))$. Motivated by this observation, $\sigma_1^2(p)$ could be estimated by the bootstrap method.

The key point of the bootstrap is resampling from the sample. The ordinary bootstrap method draws samples from the observations equally with replacement. However, it may not be appropriate to apply the ordinary bootstrap method to the imputed data, because the original method treats the imputed values as if they were true observations. This would result in an under-estimation of the variance, discussed by Shao and Sitter [63]. Therefore, instead, the bootstrap data set should also be imputed in the same way as the original data set was imputed. In this paper, the bootstrap method for imputed data proposed by Shao and Sitter [63] is used to obtain the variance estimate. We summarize the procedure for computing the bootstrap variance as follows:

1. Draw a resample of size n, denoted by $(\tilde{Y}_j^*, \delta_{\tilde{Y}_j^*})$'s, with replacement from the imputed diseased sample $(\tilde{Y}_j, \delta_{Y_j})$'s, and a separate resample of size m, denoted by $(\tilde{X}_i^*, \delta_{\tilde{X}_i^*})$'s, with replacement from the imputed non-diseased sample $(\tilde{X}_i, \delta_{X_i})$'s.

2. Let $S_{r_Y}^*$ and $S_{m_Y}^*$ denote the sets of observed data with $\delta_{\tilde{Y}_j^*} = 1$, and imputed data with $\delta_{\tilde{Y}_j^*} = 0$, with respect to bootstrap resample $\tilde{Y}_j^*$ of the diseased group, respectively. Similarly, let $S_{r_X}^*$ and $S_{m_X}^*$ denote the sets of observed data with $\delta_{\tilde{X}_i^*} = 1$, and imputed data with $\delta_{\tilde{X}_i^*} = 0$, with respect to bootstrap resample $\tilde{X}_i^*$ of the non-diseased group, respectively.

3. Apply the same random hot deck imputation procedure used in constructing $\tilde{Y}_j$'s and

$\tilde{X}_i$'s to impute the bootstrap analog $\tilde{Y}_j^*$ and $\tilde{X}_i^*$ belonging to $S_{m_Y}^*$ and $S_{m_X}^*$ respectively.

4. Calculate the bootstrap version of $\widetilde{R}(p)$ by

$$\widetilde{R}_b^*(p) = \frac{\sum_{j=1}^n I(\tilde{Y}_j^* \geq \widetilde{F}^{-1*}(p))}{n},$$

where $\widetilde{F}^{-1*}(p)$ is the p -th sample quantile based on the bootstrap resample $\tilde{X}_i^*$'s.

5. Repeat the first four steps B times (In this paper, we use $B = 600$) to obtain the set of bootstrap replicates $\{\widetilde{R}_b^*(p) : b = 1, \ldots, B\}$. Thus, the bootstrap estimate $\sigma_1^{*2}(p)$ of $\sigma_1^2(p)$ is calculated as

$$\sigma_1^{*2}(p) = \frac{n}{B-1} \sum_{b=1}^{B} \left(\widetilde{R}_b^*(p) - \bar{R}^*(p) \right)^2,$$

where $\bar{R}^*(p) = \frac{1}{B} \sum_{b=1}^{B} \widetilde{R}_b^*(p)$.

Then an IHBEL confidence interval for $R(p)$ can be defined as follows:

$$CI_{2,\alpha} = \left\{ R(p) : c^*(p) l(R(p)) \leq \chi_1^2(1-\alpha) \right\}, \tag{2.18}$$

where $c^*(p) = \frac{\bar{R}^*(p)(1-\bar{R}^*(p))}{\sigma_1^{*2}(p)}$.

2.3 Simulation Studies

In this section simulation studies are conducted to evaluate the finite-sample performance of the proposed intervals (IPEL and IHBEL) for $R(p)$ at a given specificity p in terms of coverage probability. Here two typical settings of distributions are considered, one for symmetric distribution and one for asymmetric distribution:

(1) $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(1,1)$;

(2) $X \sim \exp(1)$ and $Y \sim \exp(2)$.

For each setting, 2000 random samples of incomplete data $(X_i, \delta_{X_i}), i = 1, \ldots, m$ and

$(Y_j, \delta_{Y_j}), j = 1, \ldots, n$ are generated from the underlying non-diseased distribution F and diseased distribution G, respectively. The sample size ranges from 50 to 150 with both $m = n$ and $m \neq n$ two cases for the two settings. We also consider different observation rates: $(\pi_1, \pi_2) = 90\%$ (high), 80% or 70% (moderate), and 60% (low) with $\pi_1 = \pi_2$ and $\pi_1 \neq \pi_2$. The full observation case with $(\pi_1, \pi_2) = (1, 1)$ is also included in studies as a comparison basis.

In Table 2.1-2.4, we present coverage probabilities of 90% and 95% confidence intervals for $R(p)$ when the specificities are fixed at $p = 0.7$, 0.8 and 0.9 based on the two proposed imputation-based empirical likelihood methods IPEL and IHBEL.

Simulation results in these tables suggest that the proposed methods work well generally, and to some extent, they are complementary to each other. The proposed methods under various missing settings could generate similar results with complete data case. Coverage probabilities of IPEL intervals are below the nominal levels when $p = 0.8$, 0.9 and sample sizes are small, but they are comparable to those with complete data cases. The performance of IPEL intervals is stable for all cases considered here. IHBEL works well except some cases with low observation rate $(\pi_1, \pi_2) = (60\%, 60\%)$. Both IPEL and IHBEL work well in moderate and large sample size $(n, m \geq 100)$ even in high missing rate setting.

Therefore, based on above observations, we suggest that: (i) when the sample size is large enough regardless of the missing rate, the IPEL is preferred because of its simplicity in calculation; (ii) when the sample size is small and missing rate is moderate, the IHBEL is preferred because it has better coverage probability, although the computation is a little extensive; (iii) when the sample size is small and the missing rate is high, the IPEL could be applied with a better density estimation method.

Also, coverage probabilities of 90% and 95% confidence intervals for $R(p)$ under the same model settings with observed data only are presented in Table 2.5-2.6. When sample sizes are moderate or large $(m, n \geq 100)$, the proposed methods perform similarly with the methods with observed data only. Therefore, results for small sample sizes $(m, n = 50, 80)$ with observed data only are presented in these two tables. Compared with results in Table 2.1-

2.4, we could find that with observed data only, coverage probabilities are slightly unstable and tend to be conservative. These observations make sense because missingness results in even smaller sample sizes. The proposed methods benefit from the imputation.

2.4 An Illustrate Example

In this section, we evaluate the diagnostic accuracy of the proposed methods by applying them to the data set of carbohydrate antigenic determinant CA19-9 in the detection of pancreatic cancer. For the purpose of comparison, we apply the proposed methods to the data set used in Qin, Davis and Jing [36], because the proposed methods will reduce to their methods when the data set is complete.

Pancreatic cancer is a disease in which the cancer cells progress in the tissues of pancreas. It is hard to diagnose the pancreatic cancer because this organ is hidden behind other organs. Furthermore, its early detection is poor or almost impossible. Therefore, the death rate of pancreatic cancer patients is extremely high. By the end of 2010 in the United States, it was estimated about 43,140 individuals would be diagnosed with this condition, and 36,800 would die from the disease.

The CA19-9 is a pancreatic cancer marker measured through a blood test. Based on the test result for CA19-9, the patient's status is classified into several levels: a high CA19-9 level indicates a progression of the disease; a low or stable CA19-9 level means improved prognosis [64]. Therefore, it is extremely important to estimate the sensitivity of CA19-9 and find its range at a fixed specificity of interest.

We apply the proposed IPEL and IHBEL methods to the data set studied by Wieand *et al.* [65] on the diagnostic accuracy of CA19-9 in detecting pancreatic cancer. The data set consists of 51 patients in the control group and 90 with pancreatic cancer. We estimated sensitivities as well as IPEL and IHBEL intervals for $R(p)$ with specificity p = 0.70, 0.80, 0.90, which are usually used. We simulated the missing mechanism MCAR to obtain missing data with different observation rates of (π_1, π_2), because the original data set is complete. The results are presented in Table 2.7. These intervals indicate that CA19-9 has moderate

to high sensitivity when the specificity is fixed at p = 0.70, 0.80, 0.90, respectively. Under different observation rates, $\widetilde{R}(p)$ is close to those with complete data, and all confidence intervals contain $\widetilde{R}(p)$ based on complete data.

Table 2.1 Model setting (1): Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 90%

		Observation rate (π_1, π_2)					
(m, n, p)	Methods	$(1, 1)$	$(0.9, 0.9)$	$(0.9, 0.8)$	$(0.8, 0.8)$	$(0.8, 0.7)$	$(0.6, 0.6)$
(50,50,0.9)	IPEL	0.8583	0.8560	0.8613	0.8553	0.8713	0.8717
	IHBEL	0.8965	0.8985	0.8900	0.8860	0.8865	0.8680
(50,50,0.8)	IPEL	0.8847	0.8747	0.8807	0.8807	0.8837	0.8890
	IHBEL	0.9055	0.9000	0.9030	0.8980	0.8885	0.8665
(50,50,0.7)	IPEL	0.9050	0.9070	0.9070	0.8990	0.9000	0.9007
	IHBEL	0.9095	0.8915	0.9095	0.9005	0.8955	0.8845
(80,50,0.9)	IPEL	0.8677	0.8713	0.8773	0.8800	0.8817	0.8813
	IHBEL	0.8885	0.8905	0.8920	0.8885	0.8785	0.8800
(80,50,0.8)	IPEL	0.8973	0.9023	0.9103	0.9093	0.9040	0.8987
	IHBEL	0.8965	0.8990	0.9045	0.9075	0.8970	0.8790
(80,50,0.7)	IPEL	0.8963	0.8940	0.9003	0.9043	0.9000	0.9040
	IHBEL	0.9220	0.9075	0.9040	0.9060	0.8975	0.8690
(80,80,0.9)	IPEL	0.8530	0.8627	0.8727	0.8627	0.8707	0.8717
	IHBEL	0.8815	0.8930	0.8915	0.8850	0.8880	0.8660
(80,80,0.8)	IPEL	0.8880	0.8853	0.8933	0.8857	0.8890	0.8910
	IHBEL	0.8995	0.9015	0.8990	0.8890	0.8815	0.8675
(80,80,0.7)	IPEL	0.8980	0.8890	0.8993	0.8977	0.9043	0.9007
	IHBEL	0.9015	0.9000	0.8995	0.8905	0.8870	0.8735
(100,100,0.9)	IPEL	0.8705	0.8685	0.8625	0.8580	0.8665	0.8685
	IHBEL	0.9005	0.8975	0.9040	0.8925	0.8920	0.8650
(100,100,0.8)	IPEL	0.8890	0.8805	0.8900	0.8830	0.8810	0.8710
	IHBEL	0.8975	0.8945	0.8970	0.8895	0.8780	0.8550
(100,100,0.7)	IPEL	0.8910	0.8990	0.9005	0.8990	0.9010	0.8930
	IHBEL	0.8950	0.9065	0.8995	0.8950	0.8950	0.8620
(150,100,0.9)	IPEL	0.8810	0.8955	0.8930	0.8885	0.8970	0.8740
	IHBEL	0.8950	0.8990	0.9020	0.8965	0.8945	0.8620
(150,100,0.8)	IPEL	0.8985	0.9005	0.9000	0.9075	0.9060	0.8945
	IHBEL	0.8985	0.8990	0.8895	0.8875	0.8880	0.8670
(150,100,0.7)	IPEL	0.8945	0.8935	0.9010	0.8925	0.8915	0.8900
	IHBEL	0.8990	0.8965	0.8985	0.8935	0.8825	0.8660

Table 2.2 Model setting (1): Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 95%

		Observation rate (π_1, π_2)					
(m, n, p)	Methods	$(1, 1)$	$(0.9, 0.9)$	$(0.9, 0.8)$	$(0.8, 0.8)$	$(0.8, 0.7)$	$(0.6, 0.6)$
(50,50,0.9)	IPEL	0.9095	0.9140	0.9165	0.9095	0.9205	0.9150
	IHBEL	0.9450	0.9455	0.9460	0.9375	0.9380	0.9305
(50,50,0.8)	IPEL	0.9365	0.9370	0.9335	0.9335	0.9380	0.9345
	IHBEL	0.9505	0.9465	0.9485	0.9425	0.9380	0.9225
(50,50,0.7)	IPEL	0.9475	0.9415	0.9460	0.9430	0.9495	0.9585
	IHBEL	0.9530	0.9595	0.9530	0.9505	0.9455	0.9335
(80,50,0.9)	IPEL	0.9200	0.9235	0.9255	0.9325	0.9340	0.9230
	IHBEL	0.9475	0.9465	0.9490	0.9510	0.9440	0.9425
(80,50,0.8)	IPEL	0.9425	0.9490	0.9520	0.9535	0.9565	0.9470
	IHBEL	0.9515	0.9515	0.9605	0.9600	0.9510	0.9390
(80,50,0.7)	IPEL	0.9565	0.9560	0.9560	0.9630	0.9640	0.9535
	IHBEL	0.9510	0.9590	0.9505	0.9540	0.9455	0.9290
(80,80,0.9)	IPEL	0.9160	0.9190	0.9260	0.9235	0.9310	0.9155
	IHBEL	0.9400	0.9405	0.9510	0.9455	0.9430	0.9325
(80,80,0.8)	IPEL	0.9425	0.9380	0.9385	0.9350	0.9415	0.9390
	IHBEL	0.9470	0.9455	0.9460	0.9415	0.9460	0.9310
(80,80,0.7)	IPEL	0.9585	0.9530	0.9535	0.9490	0.9505	0.9490
	IHBEL	0.9455	0.9505	0.9545	0.9525	0.9470	0.9330
(100,100,0.9)	IPEL	0.9205	0.9210	0.9255	0.9245	0.9265	0.9220
	IHBEL	0.9440	0.9550	0.9525	0.9445	0.9405	0.9325
(100,100,0.8)	IPEL	0.9430	0.9420	0.9450	0.9430	0.9435	0.9410
	IHBEL	0.9470	0.9425	0.9490	0.9380	0.9355	0.9220
(100,100,0.7)	IPEL	0.9445	0.9480	0.9520	0.9495	0.9565	0.9440
	IHBEL	0.9480	0.9470	0.9480	0.9470	0.9420	0.9265
(150,100,0.9)	IPEL	0.9395	0.9400	0.9425	0.9475	0.9445	0.9395
	IHBEL	0.9510	0.9480	0.9435	0.9440	0.9460	0.9240
(150,100,0.8)	IPEL	0.9490	0.9520	0.9595	0.9545	0.9540	0.9395
	IHBEL	0.9575	0.9525	0.9555	0.9560	0.9505	0.9200
(150,100,0.7)	IPEL	0.9505	0.9460	0.9470	0.9435	0.9535	0.9470
	IHBEL	0.9540	0.9510	0.9480	0.9450	0.9370	0.9205

Table 2.3 Model setting (2): Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 90%

		Observation rate (π_1, π_2)					
(m, n, p)	Methods	$(1, 1)$	$(0.9, 0.9)$	$(0.9, 0.8)$	$(0.8, 0.8)$	$(0.8, 0.7)$	$(0.6, 0.6)$
(50,50,0.9)	IPEL	0.8600	0.8515	0.8635	0.8490	0.8485	0.8405
	IHBEL	0.8900	0.8825	0.8860	0.8795	0.8820	0.8605
(50,50,0.8)	IPEL	0.8830	0.8855	0.8865	0.8715	0.8790	0.8785
	IHBEL	0.9015	0.8905	0.8940	0.8850	0.8835	0.8775
(50,50,0.7)	IPEL	0.8865	0.8875	0.8820	0.8785	0.8880	0.8660
	IHBEL	0.9070	0.9010	0.9070	0.8995	0.8890	0.8695
(80,50,0.9)	IPEL	0.8790	0.8840	0.8800	0.8730	0.8735	0.8655
	IHBEL	0.8970	0.9005	0.8975	0.8905	0.8975	0.8775
(80,50,0.8)	IPEL	0.8805	0.8890	0.8855	0.8820	0.8910	0.8810
	IHBEL	0.9105	0.9025	0.8995	0.8880	0.8920	0.8725
(80,50,0.7)	IPEL	0.8840	0.8925	0.8810	0.8785	0.8870	0.8880
	IHBEL	0.8925	0.8950	0.8900	0.8940	0.8885	0.8775
(80,80,0.9)	IPEL	0.8650	0.8775	0.8845	0.8635	0.8690	0.8540
	IHBEL	0.9030	0.9045	0.8895	0.9000	0.8945	0.8595
(80,80,0.8)	IPEL	0.8830	0.8925	0.8895	0.8760	0.8870	0.8665
	IHBEL	0.9020	0.9040	0.8920	0.8930	0.8855	0.8610
(80,80,0.7)	IPEL	0.8880	0.8935	0.8925	0.8940	0.8890	0.8820
	IHBEL	0.8920	0.8965	0.8975	0.8955	0.8815	0.8635
(100,100,0.9)	IPEL	0.8755	0.8705	0.8705	0.8640	0.8705	0.8575
	IHBEL	0.8955	0.9020	0.8925	0.8800	0.8845	0.8590
(100,100,0.8)	IPEL	0.8885	0.8950	0.9050	0.8910	0.8960	0.8890
	IHBEL	0.9080	0.9015	0.9045	0.8945	0.8975	0.8565
(100,100,0.7)	IPEL	0.8885	0.8925	0.9030	0.8990	0.9010	0.8760
	IHBEL	0.8915	0.8960	0.9020	0.9020	0.8980	0.8500
(150,100,0.9)	IPEL	0.8920	0.8860	0.8975	0.8935	0.8830	0.8785
	IHBEL	0.8995	0.8925	0.8900	0.8860	0.8765	0.8565
(150,100,0.8)	IPEL	0.8915	0.8880	0.8955	0.8875	0.8970	0.8940
	IHBEL	0.8930	0.8930	0.8935	0.8885	0.8795	0.8515
(150,100,0.7)	IPEL	0.8940	0.9040	0.8985	0.8895	0.9055	0.9025
	IHBEL	0.8990	0.8965	0.8880	0.8795	0.8825	0.8635

Table 2.4 Model setting (2): Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 95%

		Observation rate (π_1, π_2)					
(m, n, p)	Methods	$(1, 1)$	$(0.9, 0.9)$	$(0.9, 0.8)$	$(0.8, 0.8)$	$(0.8, 0.7)$	$(0.6, 0.6)$
(50,50,0.9)	IPEL	0.9105	0.9180	0.9165	0.9070	0.9140	0.8975
	IHBEL	0.9435	0.9395	0.9455	0.9415	0.9395	0.9375
(50,50,0.8)	IPEL	0.9335	0.9365	0.9345	0.9310	0.9335	0.9265
	IHBEL	0.9515	0.9475	0.9450	0.9395	0.9430	0.9450
(50,50,0.7)	IPEL	0.9355	0.9365	0.9435	0.9370	0.9300	0.9305
	IHBEL	0.9535	0.9495	0.9430	0.9465	0.9445	0.9455
(80,50,0.9)	IPEL	0.9320	0.9330	0.9355	0.9320	0.9310	0.9270
	IHBEL	0.9525	0.9550	0.9470	0.9455	0.9415	0.9500
(80,50,0.8)	IPEL	0.9360	0.9360	0.9345	0.9320	0.9425	0.9335
	IHBEL	0.9520	0.9500	0.9495	0.9480	0.9455	0.9430
(80,50,0.7)	IPEL	0.9390	0.9380	0.9370	0.9370	0.9310	0.9360
	IHBEL	0.9435	0.9445	0.9415	0.9415	0.9410	0.9305
(80,80,0.9)	IPEL	0.9320	0.9345	0.9290	0.9180	0.9230	0.9150
	IHBEL	0.9505	0.9535	0.9435	0.9475	0.9460	0.9385
(80,80,0.8)	IPEL	0.9340	0.9485	0.9455	0.9330	0.9385	0.9275
	IHBEL	0.9440	0.9500	0.9445	0.9435	0.9485	0.9335
(80,80,0.7)	IPEL	0.9490	0.9435	0.9435	0.9435	0.9405	0.9265
	IHBEL	0.9500	0.9530	0.9460	0.9480	0.9410	0.9235
(100,100,0.9)	IPEL	0.9260	0.9265	0.9360	0.9180	0.9320	0.9055
	IHBEL	0.9495	0.9475	0.9510	0.9385	0.9485	0.9305
(100,100,0.8)	IPEL	0.9420	0.9385	0.9440	0.9435	0.9450	0.9360
	IHBEL	0.9520	0.9505	0.9510	0.9505	0.9465	0.9265
(100,100,0.7)	IPEL	0.9355	0.9435	0.9490	0.9450	0.9475	0.9350
	IHBEL	0.9445	0.9465	0.9495	0.9480	0.9425	0.9340
(150,100,0.9)	IPEL	0.9465	0.9395	0.9435	0.9440	0.9430	0.9320
	IHBEL	0.9465	0.9495	0.9490	0.9450	0.9360	0.9240
(150,100,0.8)	IPEL	0.9470	0.9510	0.9460	0.9475	0.9500	0.9450
	IHBEL	0.9460	0.9460	0.9440	0.9475	0.9330	0.9215
(150,100,0.7)	IPEL	0.9490	0.9520	0.9440	0.9480	0.9490	0.9495
	IHBEL	0.9490	0.9470	0.9430	0.9425	0.9335	0.9260

Table 2.5 Model setting (1) with observed data only: Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 90% and 95%

		(π_1, π_2), 90%		(π_1, π_2), 95%	
(m, n, p)	Methods	$(0.8, 0.7)$	$(0.6, 0.6)$	$(0.8, 0.7)$	$(0.6, 0.6)$
(50,50,0.9)	IPEL	0.8795	0.8715	0.9295	0.9245
	IHBEL	0.9250	0.9110	0.9580	0.9520
(50,50,0.8)	IPEL	0.8940	0.8870	0.9455	0.9345
	IHBEL	0.9145	0.9145	0.9595	0.9505
(50,50,0.7)	IPEL	0.8980	0.9000	0.9515	0.9565
	IHBEL	0.9125	0.9135	0.9660	0.9610
(80,50,0.9)	IPEL	0.8715	0.8825	0.9325	0.9350
	IHBEL	0.9145	0.9085	0.9560	0.9535
(80,50,0.8)	IPEL	0.8980	0.9020	0.9550	0.9530
	IHBEL	0.9245	0.9135	0.9620	0.9650
(80,50,0.7)	IPEL	0.9150	0.9100	0.9560	0.9615
	IHBEL	0.9260	0.9130	0.9650	0.9630
(80,80,0.9)	IPEL	0.8670	0.8540	0.9260	0.9190
	IHBEL	0.9170	0.9010	0.9555	0.9460
(80,80,0.8)	IPEL	0.8950	0.8960	0.9395	0.9435
	IHBEL	0.9200	0.9035	0.9560	0.9515
(80,80,0.7)	IPEL	0.9090	0.9050	0.9525	0.9485
	IHBEL	0.9180	0.9050	0.9645	0.9505

Table 2.6 Model setting (2) with observed data only: Coverage probabilities of IPEL and IHBEL for $R(p)$ with nominal confidence level 90% and 95%

		(π_1, π_2), 90%		(π_1, π_2), 95%	
(m, n, p)	Methods	$(0.8, 0.7)$	$(0.6, 0.6)$	$(0.8, 0.7)$	$(0.6, 0.6)$
(50,50,0.9)	IPEL	0.8580	0.8640	0.9190	0.9205
	IHBEL	0.9105	0.9145	0.9580	0.9585
(50,50,0.8)	IPEL	0.8835	0.8785	0.9340	0.9300
	IHBEL	0.9165	0.9095	0.9615	0.9580
(50,50,0.7)	IPEL	0.8930	0.8895	0.9450	0.9420
	IHBEL	0.9170	0.9110	0.9590	0.9620
(80,50,0.9)	IPEL	0.8845	0.8920	0.9395	0.9390
	IHBEL	0.9170	0.9205	0.9590	0.9600
(80,50,0.8)	IPEL	0.8810	0.8875	0.9345	0.9435
	IHBEL	0.9045	0.9100	0.9535	0.9560
(80,50,0.7)	IPEL	0.8860	0.8870	0.9370	0.9350
	IHBEL	0.9110	0.9055	0.9500	0.9495
(80,80,0.9)	IPEL	0.8810	0.8830	0.9330	0.9280
	IHBEL	0.9190	0.9160	0.9605	0.9505
(80,80,0.8)	IPEL	0.8895	0.8840	0.9420	0.9390
	IHBEL	0.9100	0.9030	0.9560	0.9530
(80,80,0.7)	IPEL	0.8885	0.8845	0.9435	0.9360
	IHBEL	0.9135	0.9045	0.9560	0.9505

Table 2.7 A real example: 95% IPEL and IHBEL confidence intervals for $R(p)$ of CA19-9 at different levels of specificities with various observation rates.

		$p = 0.70$		$p = 0.80$		$p = 0.90$	
(π_1, π_2)	Methods	$\widetilde{R}(p)$	C.I.	$\widetilde{R}(p)$	C.I.	$\widetilde{R}(p)$	C.I.
(1.0,1.0)	IHBEL	0.822	(0.698, 0.897)	0.783	(0.667, 0.867)	0.740	(0.636, 0.853)
	IPEL	0.811	(0.722, 0.883)	0.778	(0.684, 0.856)	0.756	(0.660, 0.837)
(0.9,0.9)	IHBEL	0.797	(0.680, 0.892)	0.751	(0.629, 0.858)	0.700	(0.603, 0.841)
	IPEL	0.800	(0.698, 0.881)	0.756	(0.648, 0.845)	0.733	(0.624, 0.826)
(0.9,0.8)	IHBEL	0.805	(0.642, 0.882)	0.756	(0.599, 0.844)	0.708	(0.578, 0.841)
	IPEL	0.778	(0.662, 0.870)	0.733	(0.612, 0.835)	0.722	(0.600, 0.825)
(0.8,0.8)	IHBEL	0.837	(0.685, 0.905)	0.800	(0.688, 0.903)	0.743	(0.620, 0.880)
	IPEL	0.811	(0.702, 0.895)	0.811	(0.701, 0.896)	0.767	(0.652, 0.860)
(0.8,0.7)	IHBEL	0.688	(0.560, 0.836)	0.688	(0.560, 0.836)	0.688	(0.560, 0.836)
	IPEL	0.711	(0.569, 0.830)	0.711	(0.569, 0.830)	0.711	(0.569, 0.830)
(0.6,0.6)	IHBEL	0.778	(0.644, 0.911)	0.725	(0.562, 0.851)	0.677	(0.538, 0.817)
	IPEL	0.800	(0.655, 0.906)	0.722	(0.567, 0.848)	0.689	(0.532, 0.821)

CHAPTER 3

IMPUTATION-BASED EMPIRICAL LIKELIHOOD INFERENCE FOR THE AREA UNDER THE ROC CURVE WITH MISSING COMPLETELY AT RANDOM DATA

3.1 Introduction

In this chapter, we propose an imputation-based empirical likelihood method to construct confidence intervals for the AUC with missing completely at random (MCAR) type of data, which has not been considered in literature. The proposed method preserves the advantage of the method in Qin and Zhou [40], which has good small sample performance, and the advantage of the random hot deck imputation method, which preserves the distribution of item values whereas the deterministic imputation methods like the ratio imputation and the regression imputation do not have this appealing property [54].

The remainder of this chapter is organized as follows. Section 3.2 presents the imputation-based empirical likelihood method to construct confidence intervals for the AUC with missing data. In Section 3.3, we conduct simulation studies to evaluate the performance of the proposed method. In Section 3.4, we apply the new imputation-based empirical likelihood interval to a real example. All proofs are deferred until the Appendix B.

3.2 Imputation-based Empirical Likelihood for the AUC

In this section, we aim to construct empirical likelihood-based confidence intervals for the AUC with missing data. We first impute the missing data by the hot deck imputation technique, and then apply the empirical likelihood method to obtain confidence intervals for the AUC based on the imputed data. Finally we extend the proposed method to stratified random samples with missing data.

Proposition 1 in Chapter 2 has proved that based on the imputed data $\tilde{X}_i$'s and $\tilde{Y}_j$'s,

the empirical distributions $\widetilde{F}(x) = \frac{1}{m}\sum_{i=1}^{m} I(\tilde{X}_i \leq x)$, and $\widetilde{G}(y) = \frac{1}{n}\sum_{j=1}^{n} I(\tilde{Y}_j \leq y)$ are still consistent and asymptotically normal.

We define the imputed version of WMW estimator for the AUC as follows:

$$\widetilde{\delta} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} I(\tilde{Y}_j \geq \tilde{X}_i). \tag{3.1}$$

3.2.1 Empirical Likelihood for the AUC

In order to obtain better confidence intervals for the AUC, Qin and Zhou [40] proposed an empirical likelihood-based interval for the AUC. This interval has a good coverage accuracy for high values of the AUC when sample sizes for diseased and non-diseased subjects are small and unequal.

For a given test value Y from a diseased subject, let $U = 1 - F(Y)$. The value U can be interpreted as the proportion of the non-diseased population with test values greater than Y [66]. It is easy to obtain the following equality:

$$E(1-U) = E(F(Y)) = P(Y \geq X) = \delta.$$

Based on the relationship between δ and U, an empirical likelihood procedure for the inference of the AUC was derived by Qin and Zhou [40]. Let $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ be a probability vector, i.e., $\sum_{j=1}^{n} p_j = 1$ and $p_j \geq 0$ for all j. The empirical likelihood for the AUC, evaluated at the true value δ_0 of δ, is defined as follows:

$$L(\delta_0) = \sup\left\{\prod_{j=1}^{n} p_j : \sum_{j=1}^{n} p_j = 1, \sum_{j=1}^{n} p_j W_j(\delta_0) = 0\right\},$$

where $W_j(\delta_0) = 1 - U_j - \delta_0$ with $U_j = 1 - F(Y_j)$, $j = 1, 2, \ldots, n$. Since the unknown distribution function F of the non-diseased population can be replaced by its empirical distribution $F_m(x) = \frac{1}{m}\sum_{i=1}^{m} I(X_i \leq x)$, then a profile empirical likelihood (PEL) for δ_0 can

be given by

$$\widehat{L}(\delta_0) = \sup\left\{\prod_{j=1}^{n} p_j : \sum_{j=1}^{n} p_j = 1, \sum_{j=1}^{n} p_j \widehat{W}_j(\delta_0) = 0\right\},$$

where $\widehat{W}_j(\delta_0) = 1 - \widehat{U}_j - \delta_0$ with $\widehat{U}_j = 1 - F_m(Y_j)$, $j = 1, 2, \ldots, n$. By the standard procedure of empirical likelihood method, the empirical likelihood ratio for δ_0 can be defined as follows:

$$R(\delta_0) = \prod_{j=1}^{n}(np_j) = \prod_{j=1}^{n}\left\{1 + \widehat{\lambda}\widehat{W}_j(\delta_0)\right\}^{-1},$$

where $\widehat{\lambda}$ is the solution of

$$\frac{1}{n}\sum_{j=1}^{n}\frac{\widehat{W}_j(\delta_0)}{1 + \widehat{\lambda}\widehat{W}_j(\delta_0)} = 0. \tag{3.2}$$

Then the corresponding log-EL ratio is

$$\widehat{l}(\delta_0) \equiv -2\log R(\delta_0) = 2\sum_{j=1}^{n}\log\left\{1 + \widehat{\lambda}\widehat{W}_j(\delta_0)\right\}. \tag{3.3}$$

Qin and Zhou [40] proved that the limiting distribution of $\widehat{l}(\delta_0)$ is a scaled chi-square distribution.

3.2.2 Imputation-based Empirical Likelihood Interval for the AUC

Based on the imputed data $\tilde{X}_i$'s and $\tilde{Y}_j$'s, we could substitute all complete data X_i's and Y_j's in the previous part and obtain the similar log-EL ratio for δ_0 as follows:

$$\widetilde{l}(\delta_0) = 2\sum_{j=1}^{n}\log\left\{1 + \widetilde{\lambda}\widetilde{W}_j(\delta_0)\right\}. \tag{3.4}$$

where $\widetilde{W}_j(\delta_0) = 1 - \widetilde{U}_j - \delta_0$ with $\widetilde{U}_j = 1 - \widetilde{F}(\tilde{Y}_j)$, $j = 1, 2, \ldots, n$, and $\widetilde{\lambda}$ is the solution of

$$\frac{1}{n}\sum_{j=1}^{n} \frac{\widetilde{W}_j(\delta_0)}{1 + \widetilde{\lambda}\widetilde{W}_j(\delta_0)} = 0. \tag{3.5}$$

The following theorem establishes the asymptotic distribution of the imputation-based empirical log-likelihood ratio for the AUC.

Theorem 2 *Let δ_0 be the true value of the AUC. If $\lim_{m,n\to\infty} \frac{n}{m} = \kappa < \infty$, a fixed quantity, then the asymptotic distribution of $\widetilde{l}(\delta_0)$, defined by (3.4), is a scaled χ^2 distribution with degree of freedom one, i.e.,*

$$r(\delta_0)\widetilde{l}(\delta_0) \xrightarrow{d} \chi_1^2, \tag{3.6}$$

where the scale constant $r(\delta_0)$ is

$$r(\delta_0) = \frac{m}{m+n} \frac{\sum_{j=1}^{n} \widetilde{W}_j^2(\delta_0)}{nS^2}$$

with

$$\begin{aligned}
S^2 &= \frac{m(1 - \pi_2 + \pi_2^{-1})S_{01}^2 + n(1 - \pi_1 + \pi_1^{-1})S_{10}^2}{m+n}, \\
S_{10}^2 &= \frac{1}{(m-1)n^2}\left[\sum_{i=1}^{m}(R_i - i)^2 - m\left(\bar{R} - \frac{m+1}{2}\right)^2\right], \\
S_{01}^2 &= \frac{1}{(n-1)m^2}\left[\sum_{j=1}^{n}(S_j - j)^2 - n\left(\bar{S} - \frac{n+1}{2}\right)^2\right], \\
\bar{R} &= \frac{1}{m}\sum_{i=1}^{m} R_i, \quad \textit{and} \quad \bar{S} = \frac{1}{n}\sum_{j=1}^{n} S_j.
\end{aligned}$$

Here R_i is the rank of $\tilde{X}_{(i)}$ (the i-th ordered value among $\tilde{X}_i$'s) in the combined sample of $\tilde{X}_i$'s and $\tilde{Y}_j$'s, and S_j is the rank of $\tilde{Y}_{(j)}$ (the j-th ordered value among $\tilde{Y}_j$'s) in the combined sample of $\tilde{X}_i$'s and $\tilde{Y}_j$'s.

If only complete observations are used without applying the random hot deck impu-

tation, asymptotic distributions of empirical distributions with observed data only were obtained in Corollary 1. Define

$$\begin{aligned}\widetilde{F}^*(x) &= \frac{1}{r_X}\sum_{i\in S_{r_X}} I(X_i \le x)\\ \widetilde{G}^*(y) &= \frac{1}{r_Y}\sum_{j\in S_{r_Y}} I(Y_j \le y).\end{aligned}$$

Then we have that

$$\sqrt{m}(\widetilde{F}^*(x) - F(x)) \xrightarrow{d} \mathcal{N}(0, \sigma_X^{*2})$$

where $\sigma_X^{*2} = \pi_1^{-1}F(x)(1-F(x))$, and

$$\sqrt{n}(\widetilde{G}^*(y) - G(y)) \xrightarrow{d} \mathcal{N}(0, \sigma_Y^{*2})$$

where $\sigma_Y^{*2} = \pi_2^{-1}G(y)(1-G(y))$.

The above results for $\widetilde{F}^*(x)$ and $\widetilde{G}^*(y)$ are slightly different from Corollary 1. Without the random hot deck imputation, some terms in Corollary 1 are absent. Actually, it is equivalent to disregard missing data and apply the method based on complete data to the observed data only. When sample sizes are small and missing proportion is high, the performances of the method with observed data only may be unstable because missingness results in even smaller sample size. However, the proposed method will benefit from the imputation. Similar results were observed in simulation studies in Chapter 2.

The confidence interval for the AUC could be constructed based on Theorem 2. Intuitively, by plugging in the consistent estimates of all unknown quantities, we could get the plug-in form confidence interval. Let $\tilde{\pi}_1 = \frac{r_X}{m}$, $\tilde{\pi}_2 = \frac{r_Y}{n}$, and

$$\begin{aligned}\widetilde{S}^2 &= \frac{m(1-\tilde{\pi}_2+\tilde{\pi}_2^{-1})S_{01}^2 + n(1-\tilde{\pi}_1+\tilde{\pi}_1^{-1})S_{10}^2}{m+n},\\ r(\widetilde{\delta}) &= \frac{m}{m+n}\frac{\sum_{j=1}^n \widetilde{W}_j(\widetilde{\delta})}{n\widetilde{S}^2}.\end{aligned}$$

where $\widetilde{\delta}$ is defined by (3.1). Then a $(1-\alpha)100\%$ imputation-based profile empirical likelihood confidence interval for δ_0, denoted by IPEL interval, can be defined as follows:

$$R_\alpha(\delta) = \{\delta : r(\widetilde{\delta})\widetilde{l}(\delta) \leq \chi_1^2(1-\alpha)\}, \tag{3.7}$$

where $\chi_1^2(1-\alpha)$ is the $(1-\alpha)100\%$ quantile of the chi-square distribution with degree of freedom one.

3.2.3 Imputation-based EL Intervals for the AUC with Stratified Samples

In this section, we extend the IPEL method in the previous section to stratified samples. Suppose L institutions participate in a ROC study of continuous-scale diagnostic test, which are indexed by l. Let X_l and Y_l be the results of a continuous-scale test for a non-diseased and a diseased subject in the lth institution, and F_l and G_l be the corresponding distribution functions, respectively. Let $X_{l1}, \ldots, X_{lm_l}$ be the test results of a random sample of non-diseased patients, $Y_{l1}, \ldots, Y_{ln_l}$ be results of a random sample of diseased subjects in the l-th institution, and the observation rate pairs of each institution be $(\pi_{l1}, \pi_{l2}),\ 1 \leq l \leq L$. Based on the MCAR assumption and the random hot deck imputation technique, the imputed data $\tilde{X}_{l1}, \ldots, \tilde{X}_{lm_l}$ and $\tilde{Y}_{l1}, \ldots, \tilde{Y}_{ln_l}$ could be obtained for each institution.

Similar with Qin and Zhou [40], we do not assume that F_l's and G_l's are homogeneous institutions. Instead, we only assume $\delta_1 = \ldots = \delta_l = \delta$, where δ_l denotes the AUC for the l-th institution.

Let $\mathbf{p}_l = (p_{l1}, \ldots, p_{ln_l})$ be a probability vector for $l = 1, \ldots, L$. Similarly, the profile empirical likelihood for the common AUC, evaluated at the true value δ, is defined as follows:

$$\widetilde{L}(\delta) = \sup\left\{\prod_{l=1}^{L}\prod_{j=1}^{n_l} p_{lj} : \sum_{j=1}^{n_l} p_{lj} = 1, \sum_{j=1}^{n_l} p_{lj}\widetilde{W}_{lj}(\delta) = 0, l = 1, \ldots, L\right\},$$

where $\widetilde{W}_{lj}(\delta) = 1 - \widetilde{U}_{lj} - \delta$ with $\widetilde{U}_{lj} = 1 - \widetilde{F}_l(\widetilde{Y}_{lj}),\ l = 1, \ldots, \text{Ł},\ j = 1, 2, \ldots, n_l$, and the $\widetilde{F}_l$ is the imputation-based empirical distribution of F_l. Then, the corresponding empirical

log-likelihood ratio is

$$\widetilde{l}(\delta) = 2\sum_{l=1}^{L}\sum_{j=1}^{n_l}\log\left\{1+\widetilde{\lambda}_l\widetilde{W}_{lj}(\delta)\right\}, \tag{3.8}$$

where $\widetilde{\lambda}_l$ is the solution of

$$\frac{1}{n_l}\sum_{j=1}^{n_l}\frac{\widetilde{W}_{lj}(\delta)}{1+\widetilde{\lambda}_l\widetilde{W}_{lj}(\delta)} = 0, \quad l = 1,\ldots,L. \tag{3.9}$$

The following theorem establishes the asymptotic distribution of the imputation-based empirical log-likelihood ratio for the AUC with stratified samples.

Theorem 3 *Let δ_0 be the true value of the common AUC. If $\lim_{m_l,n_l\to\infty}\frac{n_l}{m_l} = \kappa_l < \infty$, a fixed quantity, for $l = 1,\ldots,L$, then the asymptotic distribution of $\widetilde{l}(\delta_0)$, defined by (3.8), is a weighted summation of independent χ^2 distribution with degree of freedom one, i.e.,*

$$\widetilde{l}(\delta_0) \xrightarrow{d} w_1\chi^2_{1,1} + \ldots + w_L\chi^2_{1,L}, \tag{3.10}$$

where $\chi^2_{1,l}, l = 1,\ldots,L$ are L independent chi-squared distributions with degree of freedom 1, and the weights $w_l = \lim_{m_l,n_l\to\infty}\widetilde{w}_l(\delta_0)$, $1 \le l \le L$, with

$$\begin{aligned}
\widetilde{w}_l(\delta_0) &= \frac{m_l+n_l}{m_l}\frac{n_l S_l^2}{\sum_{j=1}^{n_l}\widetilde{W}^2_{lj}(\delta_0)} \\
S_l^2 &= \frac{m_l(1-\pi_{l2}+\pi_{l2}^{-1})S_{01}(l)^2 + n_l(1-\pi_{l1}+\pi_{l1}^{-1})S_{10}^2(l)}{m_l+n_l}, \\
S_{10}^2(l) &= \frac{1}{(m_l-1)n_l^2}\left[\sum_{i=1}^{m_l}(R_i(l)-i)^2 - m_l\left(\bar{R}_l - \frac{m_l+1}{2}\right)^2\right], \\
S_{01}(l)^2 &= \frac{1}{(n_l-1)m_l^2}\left[\sum_{j=1}^{n_l}(S_j(l)-j)^2 - n_l\left(\bar{S}_l - \frac{n_l+1}{2}\right)^2\right], \\
& \bar{R}_l = \frac{1}{m_l}\sum_{i=1}^{m_l}R_i(l), \quad \textit{and} \quad \bar{S}_l = \frac{1}{n_l}\sum_{j=1}^{n_l}S_j(l).
\end{aligned}$$

Here $R_i(l)$ is the rank of $\tilde{X}_{l(i)}$ (the i-th ordered value among $\tilde{X}_{li}$'s) in the combined sample

of $\tilde{X}_{li}$'s and $\tilde{Y}_{lj}$'s, and $S_j(l)$ is the rank of $\tilde{Y}_{l(j)}$ (the j-th ordered value among $\tilde{Y}_{lj}$'s) in the combined sample of $\tilde{X}_{li}$'s and $\tilde{Y}_{lj}$'s.

Then the EL-based confidence interval for the common AUC can be constructed as follows:

$$R_\alpha(\delta) = \left\{ \delta : \widetilde{l}(\delta) \leq c_{1-\alpha} \right\}, \tag{3.11}$$

where $c_{1-\alpha}$ is the $(1-\alpha)100\%$th quantile of the weighted chi-square distribution $w_1\chi^2_{1,1} + \ldots + w_L\chi^2_{1,L}$. The quantile $c_{1-\alpha}$ could be calculated using a simple Monte Carlo simulation by plugging in consistent estimates of all unknown quantities. Therefore, $R_\alpha(\delta)$ defined by (3.11) offers an approximate confidence interval for the common AUC with asymptotically correct coverage probability $1-\alpha$.

3.3 Simulation Studies

In this section simulation studies are conducted to evaluate the finite-sample performance of the proposed IPEL intervals for the AUC in terms of coverage probability when the AUC is taken to be 0.8 (moderate accuracy), 0.9, and 0.95 (high accuracy). For simplicity, we take $L = 1$ in simulation studies. Here two typical settings of distribution are considered, one for symmetric distribution and the other for asymmetric distribution:

(1) $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(\sqrt{5}\Phi^{-1}(\delta), 2^2)$;

(2) $X \sim \exp(1)$ and $Y \sim \exp(\frac{\delta}{1-\delta})$.

Note that in the first simulation setting, δ is related to the mean and the standard deviation by the following relationship:

$$\delta = \Phi\left(\frac{\mu - \mu_0}{\sqrt{\sigma^2 + \sigma_0^2}}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution, if

$X \sim \mathcal{N}(\mu_0, \sigma_0^2)$ and $Y \sim \mathcal{N}(\mu, \sigma^2)$. Meanwhile, if $X \sim \exp(\theta_1)$ and $Y \sim \exp(\theta_2)$, then

$$\delta = \frac{\theta_2}{\theta_1 + \theta_2}.$$

For each setting, 2000 random samples of incomplete data $(X_i, \delta_{X_i}), i = 1, \ldots, m$ and $(Y_j, \delta_{Y_j}), j = 1, \ldots, n$ are generated from the underlying non-diseased distribution F and diseased distribution G, respectively. The sample size ranges from 50 to 200 with both $m = n$ and $m \neq n$ two cases for the two settings. We also consider different observation rate: $(\pi_1, \pi_2) = 90\%$ (high), 80% or 70% (moderate), and 60% (low) with $\pi_1 = \pi_2$ and $\pi_1 \neq \pi_2$. For comparison, the full observation case is also included in the study. Note that when $\pi_1 = \pi_2 = 1$, the proposed method will be reduced to the method developed by Qin and Zhou [40], which has been shown to have good finite sample performance.

In Table 3.1-3.4, we present the coverage probabilities of 90% and 95% IPEL intervals for various values of the AUC based on the proposed imputation-based empirical likelihood method under two model settings. The simulation results in these tables indicate that the proposed method works well in moderate accuracy cases even with small sample sizes (i.e., $m = n = 50$). In high accuracy cases, the proposed method seems to be conservative in small sample size case, and the performance improves as the sample size increases. Reasonably, the proposed method works better in symmetric distribution case. Also, the performance of the proposed method under missing data cases is comparable with that under the complete data cases in terms of coverage probability.

3.4 A Real Example

In this section, we evaluate the diagnostic accuracy of the proposed method by applying it to the data set of carbohydrate antigenic determinant CA19-9 in the detection of pancreatic cancer. This data set has already been introduced and analyzed in Chapter 2. In this chapter, we focus on the inference of the AUC.

We apply the newly proposed IPEL method to the data set studied by Wieand *et al.* [65]

on the diagnostic accuracy of CA19-9 in detecting pancreatic cancer. The data set consists of 51 patients in the control group and 90 patients with pancreatic cancer. We simulated the missing mechanism MCAR to obtain missing data with different observation rates of (π_1, π_2), because the original data set is complete. The WMW estimates and IPEL intervals for the AUC are calculated. The results are presented in Table 3.5. These intervals indicate that CA19-9 has moderate to high level of diagnostic accuracy in detecting patients with the pancreatic cancer. Under different observation rates, $\widetilde{\delta}$ is close to the one with complete data, and all confidence intervals contain $\widetilde{\delta}$ based on complete data.

Table 3.1 Model setting (1): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 90% and various observation rates (π_1, π_2)

		Observation rates (π_1, π_2)					
AUC	(m, n)	$(1, 1)$	$(0.9, 0.9)$	$(0.9, 0.8)$	$(0.8, 0.8)$	$(0.8, 0.7)$	$(0.6, 0.6)$
0.80	(50,50)	0.9159	0.9101	0.9050	0.9085	0.9078	0.9189
	(50,80)	0.9060	0.9084	0.9049	0.9003	0.9018	0.9120
	(80,80)	0.9040	0.9115	0.9090	0.9035	0.8958	0.9111
	(80,100)	0.8955	0.8960	0.8945	0.8985	0.8989	0.9068
	(100,100)	0.8955	0.8920	0.8955	0.8984	0.8965	0.8968
	(100,150)	0.9065	0.8925	0.9010	0.9000	0.9045	0.9010
	(200,200)	0.8890	0.8940	0.8980	0.9030	0.8939	0.9025
0.90	(50,50)	0.9315	0.9410	0.9367	0.9344	0.9319	0.9467
	(50,80)	0.9153	0.9252	0.9209	0.9229	0.9253	0.9432
	(80,80)	0.9061	0.9160	0.9148	0.9196	0.9201	0.9276
	(80,100)	0.8974	0.9022	0.8970	0.9048	0.9179	0.9152
	(100,100)	0.9018	0.9030	0.8995	0.9000	0.9056	0.9052
	(100,150)	0.9020	0.8933	0.9009	0.9009	0.9058	0.9091
	(200,200)	0.8904	0.8939	0.8979	0.8955	0.8958	0.9013
0.95	(50,50)	0.9498	0.9589	0.9521	0.9478	0.9463	0.9543
	(50,80)	0.9346	0.9455	0.9456	0.9465	0.9462	0.9617
	(80,80)	0.9256	0.9419	0.9479	0.9475	0.9479	0.9548
	(80,100)	0.9240	0.9333	0.9327	0.9297	0.9507	0.9459
	(100,100)	0.9072	0.9163	0.9299	0.9297	0.9354	0.9369
	(100,150)	0.9019	0.8951	0.9113	0.9140	0.9267	0.9333
	(200,200)	0.8882	0.9012	0.9003	0.9012	0.9065	0.9075

Table 3.2 Model setting (1): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 95% and various observation rates (π_1, π_2)

		Observation rates (π_1, π_2)					
AUC	(m, n)	$(1, 1)$	$(0.9, 0.9)$	$(0.9, 0.8)$	$(0.8, 0.8)$	$(0.8, 0.7)$	$(0.6, 0.6)$
0.80	(50,50)	0.9600	0.9603	0.9593	0.9582	0.9620	0.9742
	(50,80)	0.9510	0.9545	0.9494	0.9559	0.9544	0.9608
	(80,80)	0.9500	0.9550	0.9575	0.9640	0.9544	0.9623
	(80,100)	0.9490	0.9485	0.9560	0.9550	0.9540	0.9494
	(100,100)	0.9510	0.9515	0.9510	0.9530	0.9560	0.9464
	(100,150)	0.9535	0.9485	0.9530	0.9530	0.9585	0.9525
	(200,200)	0.9480	0.9445	0.9515	0.9525	0.9490	0.9490
0.90	(50,50)	0.9760	0.9754	0.9743	0.9747	0.9751	0.9739
	(50,80)	0.9617	0.9641	0.9660	0.9686	0.9675	0.9734
	(80,80)	0.9553	0.9572	0.9662	0.9611	0.9608	0.9704
	(80,100)	0.9540	0.9518	0.9543	0.9597	0.9650	0.9619
	(100,100)	0.9509	0.9515	0.9523	0.9533	0.9609	0.9577
	(100,150)	0.9505	0.9469	0.9530	0.9575	0.9579	0.9568
	(200,200)	0.9485	0.9460	0.9515	0.9545	0.9479	0.9519
0.95	(50,50)	0.9833	0.9757	0.9760	0.9712	0.9697	0.9729
	(50,80)	0.9761	0.9823	0.9733	0.9717	0.9763	0.9787
	(80,80)	0.9762	0.9751	0.9748	0.9748	0.9721	0.9761
	(80,100)	0.9640	0.9721	0.9723	0.9733	0.9748	0.9711
	(100,100)	0.9592	0.9635	0.9668	0.9672	0.9711	0.9642
	(100,150)	0.9527	0.9544	0.9661	0.9654	0.9680	0.9698
	(200,200)	0.9429	0.9531	0.9607	0.9567	0.9573	0.9673

Table 3.3 Model setting (2): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 90% and various observation rates (π_1, π_2)

		Observation rates (π_1, π_2)					
AUC	(m, n)	$(1, 1)$	$(0.9, 0.9)$	$(0.9, 0.8)$	$(0.8, 0.8)$	$(0.8, 0.7)$	$(0.6, 0.6)$
0.80	(50,50)	0.9137	0.9189	0.9296	0.9260	0.9319	0.9368
	(50,80)	0.9100	0.9058	0.9087	0.9147	0.9177	0.9078
	(80,80)	0.9065	0.9043	0.9094	0.9125	0.9129	0.9177
	(80,100)	0.9080	0.8974	0.9034	0.9014	0.9040	0.9134
	(100,100)	0.9085	0.9070	0.9130	0.9119	0.9209	0.9013
	(100,150)	0.9000	0.9080	0.9105	0.9140	0.9174	0.9199
	(200,200)	0.8950	0.8940	0.8990	0.9045	0.8995	0.8949
0.90	(50,50)	0.9260	0.9340	0.9366	0.9387	0.9455	0.9499
	(50,80)	0.9091	0.9181	0.9110	0.9214	0.9342	0.9318
	(80,80)	0.9008	0.9125	0.9102	0.9126	0.9234	0.9394
	(80,100)	0.9042	0.9031	0.9100	0.9093	0.9170	0.9256
	(100,100)	0.8998	0.9045	0.9134	0.9144	0.9177	0.9240
	(100,150)	0.9035	0.9084	0.9124	0.9194	0.9197	0.9323
	(200,200)	0.8940	0.8890	0.8984	0.8993	0.9018	0.8959
0.95	(50,50)	0.9434	0.9460	0.9494	0.9494	0.9489	0.9519
	(50,80)	0.9326	0.9361	0.9398	0.9436	0.9448	0.9559
	(80,80)	0.9204	0.9400	0.9435	0.9396	0.9530	0.9554
	(80,100)	0.9206	0.9267	0.9241	0.9295	0.9403	0.9494
	(100,100)	0.9109	0.9285	0.9350	0.9320	0.9450	0.9497
	(100,150)	0.9076	0.9109	0.9129	0.9184	0.9274	0.9414
	(200,200)	0.8954	0.8951	0.9060	0.9051	0.9059	0.9145

Table 3.4 Model setting (2): Coverage probabilities of the IPEL interval for the AUC with nominal confidence level 95% and various observation rates (π_1, π_2)

		Observation rates (π_1, π_2)					
AUC	(m, n)	$(1, 1)$	$(0.9, 0.9)$	$(0.9, 0.8)$	$(0.8, 0.8)$	$(0.8, 0.7)$	$(0.6, 0.6)$
0.80	(50,50)	0.9579	0.9657	0.9666	0.9645	0.9716	0.9771
	(50,80)	0.9565	0.9564	0.9549	0.9584	0.9634	0.9627
	(80,80)	0.9510	0.9564	0.9617	0.9568	0.9622	0.9687
	(80,100)	0.9560	0.9485	0.9509	0.9464	0.9530	0.9597
	(100,100)	0.9505	0.9565	0.9600	0.9585	0.9599	0.9592
	(100,150)	0.9565	0.9610	0.9530	0.9615	0.9635	0.9625
	(200,200)	0.9475	0.9520	0.9470	0.9455	0.9480	0.9460
0.90	(50,50)	0.9655	0.9714	0.9804	0.9748	0.9787	0.9791
	(50,80)	0.9593	0.9631	0.9670	0.9742	0.9750	0.9754
	(80,80)	0.9567	0.9631	0.9665	0.9660	0.9739	0.9759
	(80,100)	0.9534	0.9573	0.9527	0.9582	0.9651	0.9740
	(100,100)	0.9494	0.9568	0.9627	0.9678	0.9652	0.9709
	(100,150)	0.9550	0.9630	0.9604	0.9609	0.9679	0.9649
	(200,200)	0.9520	0.9485	0.9465	0.9454	0.9464	0.9527
0.95	(50,50)	0.9754	0.9727	0.9760	0.9710	0.9702	0.9727
	(50,80)	0.9717	0.9742	0.9720	0.9754	0.9737	0.9779
	(80,80)	0.9702	0.9752	0.9767	0.9761	0.9791	0.9800
	(80,100)	0.9674	0.9698	0.9711	0.9715	0.9725	0.9747
	(100,100)	0.9618	0.9722	0.9794	0.9758	0.9827	0.9809
	(100,150)	0.9593	0.9587	0.9676	0.9696	0.9739	0.9815
	(200,200)	0.9445	0.9478	0.9527	0.9548	0.9562	0.9624

Table 3.5 A real example: 95% IPEL confidence intervals for the AUC of CA19-9 with various observation rates.

(π_1, π_2)	$\widetilde{\delta}$	Confidence Interval	r_X	r_Y
(1.0,1.0)	0.862	(0.793, 0.913)	51	90
(0.9,0.9)	0.874	(0.803, 0.924)	46	86
(0.9,0.8)	0.873	(0.787, 0.931)	47	68
(0.8,0.8)	0.876	(0.793, 0.931)	42	72
(0.8,0.7)	0.811	(0.704, 0.891)	39	56
(0.6,0.6)	0.835	(0.717, 0.916)	27	48

CHAPTER 4

JOINT EMPIRICAL LIKELIHOOD CONFIDENCE REGIONS FOR THE EVALUATION OF CONTINUOUS-SCALE DIAGNOSTIC TESTS WITH MISSING COMPLETELY AT RANDOM DATA

4.1 Introduction

Recently, based on the empirical likelihood method, Adimari and Chiogna [4] considered joint inferences on both the (specificity, cut-off level) and the (sensitivity, cut-off level). Joint confidence regions depict the association of sensitivity, specificity and cut-off level for a continuous-scale test. By visually inspecting confidence regions, one can select a reasonable cut-off level τ in order to obtain a desirable sensitivity $\theta(\tau)$ and an acceptable specificity $\eta(\tau)$ simultaneously, because it is well known that there is a trade-off between the sensitivity and the specificity. Moreover, by constructing joint confidence regions, one can investigate the within-pair relationship of (θ, τ) or (η, τ), respectively. In diagnostic study, the AUC is a widely used summary index of diagnostic accuracy. However it can not be used to select a cut-off level because the AUC masks the effect of cut-off level.

The proposed empirical likelihood-based joint confidence regions provide a graphical tool to select a cut-off level which yields the desirable sensitivity and/or specificity by plotting joint confidence regions for (θ, τ) and (η, τ) in the same graph. Reasonable cut-off levels could be directly identified from the overlapping part of the two regions. Such visual tool is straightforward and easy to implement in practice. It is necessary to point out that the proposed confidence regions preserve many good properties of empirical likelihood method, such as good small sample performance, data determined confidence regions and range-respecting, which could be a problem for normal-approximation based confidence regions.

In this chapter, motivated by the work of Adimari and Chiogna [4] and the missing data problem in practice, we extend the unified framework of building bivariate confidence

regions for the pair (cut-off level, sensitivity) at a fixed value of specificity, the pair (cut-off level, specificity) at a fixed value of sensitivity or the pair (specificity, sensitivity) at a fixed cut-off value by applying the empirical likelihood method, to the missing data case, especially when data are missing completely at random (MCAR). The new confidence regions preserve the good small sample performance of the empirical likelihood method, and they are computationally simple and easy to implement in practice.

This chapter is organized as follows. Section 4.2 develops imputation-based bivariate empirical likelihood confidence regions with MCAR data. Some simulation studies are presented in Section 4.3 to illustrate the finite sample performance of the proposed methods. The proof is deferred in the Appendix C.

4.2 Imputation-based Bivariate Empirical Likelihood Confidence Regions with MCAR Data

In this section, we extend the bivariate empirical likelihood confidence regions developed by Adimari and Chiogna [4] to MCAR data case by using the random hot deck imputation technique, and we obtain imputation-based bivariate empirical likelihood confidence regions with MCAR data.

4.2.1 Bivariate Nonparametric Confidence Regions with Complete Data

In order to obtain bivariate nonparametric confidence regions for the evaluation of continuous-scale diagnostic tests, Adimari and Ciogna [4] have proposed empirical likelihood-based confidence regions for any two quantities of the sensitivity, the specificity and the cut-off value. The proposed method works under very weak assumptions and easy to implement. Also, it has been shown to deserve good performance when the sample size is moderate or high.

Let $X_1, \ldots, X_m$ be a random sample from X with a distribution function $F(x)$, i.e., the test results from m non-diseased patients, and be $Y_1, \ldots, Y_n$ a random sample from Y with a distribution function $G(y)$, i.e., the test results from n diseased patients. Moreover, let

$F_m(x) = \frac{1}{m}\sum_{i=1}^{m} I(X_i \leq x)$ denote the empirical distribution function based on $X_1, \ldots, X_m$, and let $G_n(y) = \frac{1}{n}\sum_{j=1}^{n} I(Y_j \leq y)$ denote the empirical distribution function based on $Y_1, \ldots, Y_n$.

Consider the empirical likelihood function based on the two independent samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, i.e.,

$$L(\mathbf{p}, \mathbf{q}) = \prod_{i=1}^{m} p_i \prod_{j=1}^{n} q_j,$$

where, $\mathbf{p} = (p_1, \ldots, p_m)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ are probability vectors, ie.e, $\sum_{i=1}^{m} p_i = 1, p_i \geq 0$ and $\sum_{j=1}^{n} q_j = 1, q_j \geq 0$, representing multinomial distributions on $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, respectively.

Then, one could maximize $L(\mathbf{p}, \mathbf{q})$ subject to the constraints

$$\sum_{i=1}^{m} p_i I(X_i \leq \tau) = \eta, \quad \sum_{j=1}^{n} q_j I(Y_j \leq \tau) = 1 - \theta.$$

The constrained maximum is given by the product

$$\left(\sup_{\mathbf{p}:\sum_{i=1}^{m} p_i I(X_i \leq \tau) = \eta} \prod_{i=1}^{m} p_i \right) \times \left(\sup_{\mathbf{q}:\sum_{j=1}^{n} q_j I(Y_j \leq \tau) = 1-\theta} \prod_{j=1}^{n} q_j \right). \tag{4.1}$$

By applying the Lagrange multiplier method and the regular method of empirical likelihood, it follows that the empirical likelihood ratio statistic (i.e., minus twice the maximum empirical log-likelihood ratio) corresponding to the maximization of $L(\mathbf{p}, \mathbf{q})$ subject to constrains reduces to

$$\begin{aligned} l(\theta, \eta, \tau) &= 2m \left\{ F_m(\tau) \log \frac{F_m(\tau)}{\eta} + [1 - F_m(\tau)] \log \frac{1 - F_m(\tau)}{1 - \eta} \right\} \\ &+ 2n \left\{ G_n(\tau) \log \frac{G_n(\tau)}{1 - \theta} + [1 - G_n(\tau)] \log \frac{1 - G_n(\tau)}{\theta} \right\}, \end{aligned} \tag{4.2}$$

for $\eta \in (0, 1), \theta \in (0, 1), \tau \in \mathscr{T}$, where $\mathscr{T} = [x_{(1)}, x_{(m)}) \cap [y_{(1)}, y_{(n)})$.

Adimari and Chiogna [4] have proved that the asymptotic distribution of $l(\theta, \eta, \tau)$ eval-

uated at the true values $(\theta_0, \eta_0, \tau_0)$ is a chi-squared distribution with 2 degrees of freedom.

4.2.2 Imputation-based Bivariate Empirical Likelihood Confidence Regions

Based on the imputed data $\tilde{X}_i$'s and $\tilde{Y}_j$'s from the random hot deck imputation method [54], we could substitute all complete data X_i's and Y_j's in the previous part and obtain the similar log-EL ratio for $(\theta_0, \eta_0, \tau_0)$ as follows:

$$\begin{aligned} \tilde{l}(\theta, \eta, \tau) &= 2m \left\{ \widetilde{F}(\tau) \log \frac{\widetilde{F}(\tau)}{\eta} + [1 - \widetilde{F}(\tau)] \log \frac{1 - \widetilde{F}(\tau)}{1 - \eta} \right\} \\ &+ 2n \left\{ \widetilde{G}(\tau) \log \frac{\widetilde{G}(\tau)}{1 - \theta} + [1 - \widetilde{G}(\tau)] \log \frac{1 - \widetilde{G}(\tau)}{\theta} \right\}, \end{aligned} \tag{4.3}$$

for $\theta \in (0, 1), \eta \in (0, 1), \tau \in \mathscr{T}$, where $\mathscr{T} = [X_{(1)}, X_{(m)}) \cap [Y_{(1)}, Y_{(n)})$, and $\widetilde{F}(x)$ and $\widetilde{G}(y)$ are defined by (2.1) and (2.2).

Then the following result holds.

Theorem 4 *Let $F(x)$ and $G(y)$ be continuous and strictly increasing in a neighborhood of true cut-off value τ_0. Let $\theta_0 = 1 - G(\tau_0)$ and $\eta_0 = F(\tau_0)$ be the true sensitivity and the specificity levels corresponding to the threshold τ_0. Then, when $\min\{m, n\} \to +\infty$, the asymptotic distribution of $\tilde{l}(\theta_0, \eta_0, \tau_0)$ is a weighted summation of independent χ^2 distributions with degree of freedom 1, i.e.*

$$\tilde{l}(\theta_0, \eta_0, \tau_0) \xrightarrow{d} (1 - \pi_1 + \pi_1^{-1})\chi^2_{1,1} + (1 - \pi_2 + \pi_2^{-1})\chi^2_{1,2}, \tag{4.4}$$

where $\chi^2_{1,1}$ and $\chi^2_{1,2}$ are two independent chi-squared distributions with degree of freedom 1.

Theorem 4 provides $\tilde{l}(\theta, \eta, \tau)$ as an asymptotic pivotal with MCAR data for the inference of the parameter pairs (θ_0, τ_0), (η_0, τ_0) or (θ_0, η_0), given a fixed value of the third remaining parameter. But π_1 and π_2 are still unknown. Intuitively, by plugging in the consistent estimates of the two unknown quantities, we could get the plug-in form confidence interval. Let $\tilde{\pi}_1 = \frac{r_X}{m}$, $\tilde{\pi}_2 = \frac{r_Y}{n}$. Then three $(1-\alpha)100\%$ imputation-based profile empirical likelihood confidence regions for different pairs could be defined as follows:

- $\mathscr{R}^1_{\alpha,MCAR}(\theta,\tau) = \{(\theta,\tau) : \tilde{l}(\theta,\eta_0,\tau) \leq c_{\alpha,MCAR}\}$;
- $\mathscr{R}^2_{\alpha,MCAR}(\eta,\tau) = \{(\eta,\tau) : \tilde{l}(\theta_0,\eta,\tau) \leq c_{\alpha,MCAR}\}$;
- $\mathscr{R}^3_{\alpha,MCAR}(\theta,\eta) = \{(\theta,\eta) : \tilde{l}(\theta,\eta,\tau_0) \leq c_{\alpha,MCAR}\}$;

where $\alpha \in (0,1)$ and $c_{\alpha,MCAR}$ is the upper α quantile of the weighted chi-square distribution $(1-\tilde{\pi}_1+\tilde{\pi}_1^{-1})\chi^2_{1,1}+(1-\tilde{\pi}_2+\tilde{\pi}_2^{-1})\chi^2_{1,2}$, i.e. $P\left((1-\tilde{\pi}_1+\tilde{\pi}_1^{-1})\chi^2_{1,1}+(1-\tilde{\pi}_2+\tilde{\pi}_2^{-1})\chi^2_{1,2} > c_{\alpha,MCAR}\right) = \alpha$, which could be obtained by the Monte Carlo method. Also, bootstrap method could be employed to find quantiles.

4.3 Simulation Studies

In this section, simulation studies are conducted to evaluate the finite-sample performance of proposed confidence regions for every parameter pair with MCAR data in terms of coverage probability. For the purpose of comparison, similar settings of distribution with those in Adimari and Chiogna [4] are considered, one for symmetric distribution and one for asymmetric distribution:

(1) Gaussian models $\mathcal{N}(\mu,\sigma)$: $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(\mu,1/2)$;

(2) Exponential models $\text{Exp}(\gamma)$: $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(\gamma)$.

The unknown values of μ in scenario (1) and γ in scenario (2) as well as the cut-off value τ depend on the choice of reference values (θ_0,η_0) for the true pair of (Sensitivity, Specificity). The relationship has been illustrated by Adimari and Chiogna [4].

For each scenario, 4000 random samples of incomplete data $(X_i,\delta_{X_i}), i = 1,\ldots,m$ and $(Y_j,\delta_{Y_j}), j = 1,\ldots,n$ are generated from the underlying non-diseased distribution F and diseased distribution G, respectively. The sample size ranges from 20 to 100 with both $m = n$ and $m \neq n$ two cases for the two settings. We also consider different observation rate: $(\pi_1,\pi_2) = 90\%$ (high), 80% or 70% (moderate), and 60% (low) with $\pi_1 = \pi_2$ and $\pi_1 \neq \pi_2$. The full observation case with $(\pi_1,\pi_2) = (1,1)$ is also included in studies as a comparison basis. Note that when $\pi_1 = \pi_2 = 1$, the proposed method will reduce to the

method developed by Adimari and Chiogna [4], which has been shown to have good finite sample performance.

In Table 4.1-4.4, we present the coverage probabilities of 90% and 95% confidence regions for various values of the pair (θ, η) at different cut-off levels τ based on the proposed imputation-based empirical likelihood method under two model settings. Simulation results in these tables indicate that the proposed method works well in moderate accuracy cases even with small sample sizes (i.e., $m = n = 30$). In high accuracy cases, the proposed method seems to be conservative in small sample size cases, and the performance improves as the sample size increases. Reasonably, the proposed method works better in symmetric distribution case. Also, the performance of the proposed method under missing data cases is comparable with under complete data cases in terms of coverage probability.

Table 4.1 Model setting (1): Coverage probabilities of the confidence regions obtained by the empirical likelihood statistic $\tilde{l}(\theta, \eta, \tau)$ with nominal confidence level 90% at various observation rates (π_1, π_2) in the presence of the MCAR data.

						Observation rates (π_1, π_2)					
τ_0	μ	η_0	θ_0	m	n	$(1,1)$	$(0.9,0.9)$	$(0.9,0.8)$	$(0.8,0.8)$	$(0.8,0.7)$	$(0.6,0.6)$
0.842	1.262	0.8	0.80	20	20	0.929	0.927	0.933	0.940	0.936	0.943
				30	30	0.897	0.898	0.897	0.902	0.899	0.919
				50	20	0.897	0.898	0.897	0.902	0.899	0.919
				50	30	0.892	0.892	0.895	0.890	0.894	0.908
				50	50	0.892	0.889	0.896	0.898	0.897	0.897
				100	100	0.895	0.895	0.892	0.902	0.895	0.899
	1.482		0.90	20	20	0.945	0.946	0.945	0.946	0.946	0.944
				30	30	0.912	0.927	0.929	0.927	0.935	0.944
				50	20	0.928	0.927	0.924	0.924	0.924	0.917
				50	30	0.921	0.916	0.925	0.917	0.926	0.921
				50	50	0.886	0.896	0.912	0.915	0.926	0.918
				100	100	0.894	0.897	0.897	0.890	0.887	0.898
	1.664		0.95	20	20	0.942	0.936	0.932	0.939	0.925	0.926
				30	30	0.931	0.929	0.926	0.925	0.926	0.938
				50	20	0.917	0.918	0.918	0.909	0.902	0.885
				50	30	0.919	0.914	0.922	0.918	0.923	0.913
				50	50	0.915	0.922	0.927	0.928	0.928	0.922
				100	100	0.898	0.904	0.913	0.913	0.919	0.923
1.282	1.702	0.9	0.80	20	20	0.944	0.943	0.943	0.943	0.948	0.942
				30	30	0.915	0.917	0.912	0.920	0.922	0.939
				50	20	0.900	0.914	0.920	0.923	0.926	0.942
				50	30	0.882	0.900	0.902	0.909	0.913	0.929
				50	50	0.897	0.901	0.906	0.913	0.908	0.919
				100	100	0.891	0.893	0.897	0.900	0.894	0.896
	1.922		0.90	20	20	0.964	0.948	0.944	0.950	0.951	0.943
				30	30	0.947	0.949	0.949	0.953	0.950	0.952
				50	20	0.925	0.929	0.932	0.935	0.932	0.939
				50	30	0.917	0.928	0.935	0.933	0.941	0.942
				50	50	0.881	0.910	0.922	0.925	0.935	0.944
				100	100	0.895	0.896	0.895	0.896	0.890	0.896
	1.482		0.95	20	20	0.957	0.942	0.940	0.945	0.932	0.927
				30	30	0.950	0.946	0.943	0.947	0.943	0.944
				50	20	0.914	0.921	0.920	0.920	0.911	0.910
				50	30	0.915	0.921	0.931	0.935	0.932	0.928
				50	50	0.920	0.927	0.934	0.945	0.950	0.948
				100	100	0.896	0.904	0.918	0.916	0.924	0.922

Table 4.2 Model setting (1): Coverage probabilities of the confidence regions obtained by the empirical likelihood statistic $\tilde{l}(\theta, \eta, \tau)$ with nominal confidence level 95% at various observation rates (π_1, π_2) in the presence of the MCAR data.

						Observation rates (π_1, π_2)					
τ_0	μ	η_0	θ_0	m	n	$(1,1)$	$(0.9,0.9)$	$(0.9,0.8)$	$(0.8,0.8)$	$(0.8,0.7)$	$(0.6,0.6)$
0.842	1.262	0.8	0.80	20	20	0.963	0.968	0.964	0.970	0.973	0.973
				30	30	0.944	0.947	0.953	0.959	0.965	0.968
				50	20	0.951	0.955	0.954	0.952	0.957	0.957
				50	30	0.946	0.941	0.951	0.946	0.954	0.957
				50	50	0.948	0.947	0.947	0.947	0.948	0.948
				100	100	0.947	0.941	0.945	0.949	0.952	0.949
	1.482		0.90	20	20	0.971	0.975	0.971	0.974	0.972	0.970
				30	30	0.958	0.964	0.966	0.965	0.971	0.975
				50	20	0.960	0.963	0.960	0.956	0.959	0.962
				50	30	0.959	0.960	0.962	0.957	0.965	0.961
				50	50	0.945	0.952	0.957	0.956	0.968	0.960
				100	100	0.946	0.942	0.943	0.943	0.944	0.948
	1.664		0.95	20	20	0.968	0.969	0.962	0.967	0.960	0.960
				30	30	0.964	0.965	0.965	0.969	0.966	0.971
				50	20	0.955	0.959	0.957	0.951	0.946	0.938
				50	30	0.961	0.959	0.958	0.962	0.965	0.955
				50	50	0.958	0.962	0.963	0.960	0.966	0.961
				100	100	0.952	0.953	0.961	0.962	0.962	0.959
1.282	1.702	0.9	0.80	20	20	0.968	0.973	0.968	0.971	0.975	0.971
				30	30	0.955	0.959	0.962	0.970	0.970	0.972
				50	20	0.950	0.963	0.963	0.966	0.970	0.974
				50	30	0.949	0.952	0.959	0.959	0.964	0.969
				50	50	0.948	0.955	0.958	0.959	0.961	0.967
				100	100	0.944	0.943	0.944	0.945	0.943	0.950
	1.922		0.90	20	20	0.970	0.976	0.975	0.973	0.972	0.967
				30	30	0.975	0.974	0.975	0.977	0.978	0.978
				50	20	0.965	0.973	0.968	0.969	0.968	0.973
				50	30	0.965	0.973	0.973	0.973	0.973	0.972
				50	50	0.953	0.961	0.969	0.973	0.978	0.977
				100	100	0.948	0.944	0.948	0.949	0.943	0.953
	1.482		0.95	20	20	0.978	0.975	0.967	0.969	0.967	0.961
				30	30	0.979	0.973	0.968	0.972	0.970	0.974
				50	20	0.964	0.963	0.957	0.959	0.953	0.950
				50	30	0.963	0.970	0.970	0.970	0.971	0.964
				50	50	0.961	0.973	0.971	0.972	0.975	0.975
				100	100	0.954	0.952	0.959	0.960	0.963	0.966

Table 4.3 Model setting (2): Coverage probabilities of the confidence regions obtained by the empirical likelihood statistic $\tilde{l}(\theta, \eta, \tau)$ with nominal confidence level 90% at various observation rates (π_1, π_2) in the presence of the MCAR data.

						Observation rates (π_1, π_2)					
τ_0	γ	η_0	θ_0	m	n	$(1,1)$	$(0.9,0.9)$	$(0.9,0.8)$	$(0.8,0.8)$	$(0.8,0.7)$	$(0.6,0.6)$
1.609	0.139	0.8	0.80	20	20	0.926	0.935	0.937	0.938	0.943	0.946
				30	30	0.906	0.910	0.905	0.907	0.908	0.926
				50	20	0.909	0.916	0.918	0.916	0.923	0.923
				50	30	0.896	0.895	0.901	0.899	0.907	0.913
				50	50	0.905	0.895	0.905	0.903	0.895	0.899
				100	100	0.905	0.899	0.903	0.907	0.902	0.908
	0.065		0.90	20	20	0.943	0.946	0.944	0.947	0.945	0.939
				30	30	0.925	0.932	0.936	0.935	0.941	0.945
				50	20	0.934	0.927	0.936	0.932	0.929	0.929
				50	30	0.922	0.922	0.931	0.925	0.930	0.921
				50	50	0.895	0.898	0.915	0.913	0.918	0.921
				100	100	0.899	0.897	0.906	0.902	0.899	0.904
	0.032		0.95	20	20	0.940	0.935	0.925	0.927	0.926	0.931
				30	30	0.940	0.932	0.935	0.929	0.929	0.935
				50	20	0.929	0.908	0.910	0.917	0.914	0.904
				50	30	0.931	0.919	0.931	0.926	0.924	0.913
				50	50	0.930	0.925	0.930	0.927	0.932	0.930
				100	100	0.894	0.899	0.914	0.914	0.921	0.923
2.303	0.097	0.9	0.80	20	20	0.954	0.950	0.951	0.948	0.948	0.940
				30	30	0.925	0.935	0.927	0.930	0.927	0.934
				50	20	0.904	0.914	0.923	0.931	0.940	0.943
				50	30	0.889	0.911	0.910	0.911	0.913	0.938
				50	50	0.895	0.896	0.905	0.914	0.908	0.924
				100	100	0.891	0.899	0.894	0.892	0.896	0.900
	0.046		0.90	20	20	0.968	0.951	0.946	0.951	0.948	0.938
				30	30	0.952	0.957	0.957	0.955	0.960	0.951
				50	20	0.929	0.927	0.934	0.943	0.943	0.946
				50	30	0.920	0.938	0.943	0.938	0.943	0.946
				50	50	0.889	0.907	0.921	0.925	0.930	0.943
				100	100	0.890	0.892	0.896	0.891	0.891	0.892
	0.022		0.95	20	20	0.958	0.938	0.931	0.936	0.932	0.923
				30	30	0.949	0.955	0.955	0.948	0.944	0.943
				50	20	0.924	0.917	0.920	0.933	0.930	0.929
				50	30	0.922	0.935	0.948	0.946	0.937	0.944
				50	50	0.930	0.932	0.936	0.945	0.948	0.946
				100	100	0.884	0.894	0.910	0.909	0.912	0.915

Table 4.4 Model setting (2): Coverage probabilities of the confidence regions obtained by the empirical likelihood statistic $\tilde{l}(\theta, \eta, \tau)$ with nominal confidence level 95% at various observation rates (π_1, π_2) in the presence of the MCAR data.

						Observation rates (π_1, π_2)					
τ_0	γ	η_0	θ_0	m	n	$(1,1)$	$(0.9,0.9)$	$(0.9,0.8)$	$(0.8,0.8)$	$(0.8,0.7)$	$(0.6,0.6)$
1.609	0.139	0.8	0.80	20	20	0.965	0.972	0.973	0.974	0.974	0.975
				30	30	0.950	0.951	0.954	0.962	0.963	0.972
				50	20	0.957	0.959	0.961	0.959	0.962	0.962
				50	30	0.949	0.950	0.950	0.953	0.956	0.956
				50	50	0.951	0.958	0.961	0.960	0.962	0.967
				100	100	0.952	0.949	0.952	0.950	0.948	0.949
	0.065		0.90	20	20	0.972	0.977	0.978	0.978	0.975	0.973
				30	30	0.966	0.968	0.972	0.972	0.976	0.979
				50	20	0.965	0.970	0.968	0.964	0.962	0.965
				50	30	0.969	0.965	0.970	0.962	0.964	0.958
				50	50	0.951	0.958	0.961	0.960	0.962	0.967
				100	100	0.952	0.949	0.952	0.950	0.948	0.949
	0.032		0.95	20	20	0.969	0.973	0.969	0.971	0.965	0.969
				30	30	0.970	0.964	0.970	0.970	0.968	0.970
				50	20	0.965	0.957	0.961	0.959	0.957	0.950
				50	30	0.967	0.962	0.965	0.963	0.965	0.955
				50	50	0.967	0.964	0.967	0.966	0.968	0.964
				100	100	0.954	0.954	0.964	0.961	0.968	0.958
2.303	0.097	0.9	0.80	20	20	0.975	0.973	0.978	0.974	0.974	0.975
				30	30	0.964	0.965	0.967	0.971	0.970	0.971
				50	20	0.958	0.965	0.965	0.969	0.973	0.971
				50	30	0.955	0.956	0.962	0.964	0.965	0.972
				50	50	0.956	0.954	0.953	0.961	0.956	0.964
				100	100	0.948	0.946	0.951	0.947	0.944	0.952
	0.046		0.90	20	20	0.977	0.982	0.980	0.979	0.976	0.971
				30	30	0.981	0.980	0.979	0.979	0.980	0.975
				50	20	0.969	0.974	0.970	0.974	0.969	0.975
				50	30	0.966	0.975	0.978	0.974	0.972	0.977
				50	50	0.954	0.965	0.967	0.970	0.974	0.974
				100	100	0.949	0.944	0.952	0.948	0.942	0.947
	0.022		0.95	20	20	0.969	0.980	0.971	0.968	0.964	0.961
				30	30	0.985	0.977	0.975	0.975	0.973	0.970
				50	20	0.974	0.965	0.960	0.966	0.963	0.964
				50	30	0.975	0.977	0.976	0.974	0.975	0.976
				50	50	0.968	0.986	0.972	0.976	0.977	0.975
				100	100	0.948	0.954	0.962	0.955	0.957	0.958

CHAPTER 5

EMPIRICAL LIKELIHOOD CONFIDENCE REGIONS FOR THE EVALUATION OF CONTINUOUS SCALE DIAGNOSTIC TEST IN THE PRESENCE OF VERIFICATION BIAS

5.1 Introduction

In this chapter, based on the IPW, FI, MSI and SPE estimates for the sensitivity and the specificity, we propose various bias-corrected joint empirical likelihood confidence regions for the pairs of (sensitivity, cut-off level), (specificity, cut-off level), and (sensitivity, specificity). Furthermore, we provide a general framework that combines the empirical likelihood and general estimation equations with nuisance parameters. Comparative studies are conducted to evaluate these confidence regions and the normal approximation-based confidence regions. Misspecified models are also employed to show the double robustness of the SPE joint empirical likelihood confidence regions. The proposed empirical likelihood-based joint confidence regions provide a graphical tool to select a cut-off level which yields the desirable sensitivity and/or specificity by plotting joint confidence regions for (τ, θ) and (τ, η) in the same graph. Reasonable cut-off levels could be directly identified from the overlapping part of the two regions. Such visual tool is straightforward and easy to implement. The proposed confidence regions preserve many good properties of empirical likelihood method, such as good small sample performance, data determined confidence regions and range-respecting, which could be a problem for normal-approximation based confidence regions.

This chapter is organized as follows. In Section 5.2, we propose various bias-corrected joint confidence regions. Some simulation studies are presented in Section 5.3 to evaluate the finite sample performance of the proposed methods as well as their robustness to model misspecification. Real data analysis is used to compare proposed methods in Section 5.4. Proofs are included in the Appendix D.

5.2 Bias-corrected Empirical Likelihood Confidence Regions

In this section, we will develop joint empirical likelihood confidence regions in the presence of verification bias, which is assumed to be MAR.

5.2.1 Empirical Likelihood and General Estimation Equations with Nuisance Parameters

In order to develop joint empirical likelihood confidence regions in the presence of verification bias, a general framework combining empirical likelihood and general estimation equations with nuisance parameters is needed for our research. Qin and Lawless [43] linked estimating equations and empirical likelihood, and developed methods of combining information about parameters. Wang and Chen [51] applied empirical likelihood to estimating equations with missing data based on a nonparametric imputation of missing values from a kernel estimator of the conditional distribution of the missing variable given the always observable variables. Qin *et al.* [52] proposed a unified empirical likelihood approach to missing data problems and explored the use of the empirical likelihood to effectively combine unbiased estimating equations when the number of estimating equations is greater than the number of unknown parameters. Hjort *et al.* [44] extended the scope of general empirical likelihood methodology by introducing plug-in estimates of nuisance parameters in estimating equations. But there are no explicit asymptotic results on empirical likelihood defined by general estimation equations with nuisance parameters, which are estimated by another set of estimating equations.

Motivated by the common estimating equation framework to derive asymptotic properties of the two-phase disease prevalence estimators proposed by Alonzo *et al.* [57], we combine the empirical likelihood and generalized estimating equations (GEE) with nuisance parameters, and derive the asymptotic distribution of the empirical log-likelihood ratio statistic for the inference of the main parameters. Our method is different from those developed by Qin *et al.* [52] because their methods are under the regression model setting. Instead of

obtaining the explicit empirical likelihood estimators, our methods focus on the confidence interval/region estimation by using the limiting distribution of the empirical log-likelihood ratio statistic.

Let $W_1, \ldots, W_n$ be an i.i.d. sample from a d-dimensional random vector W with an unknown distribution function F_W, and a p-dimensional parameter ψ of interest and a q-dimensional nuisance parameter β are associated with F_W. We assume that the information about ψ and F_W is available in the form of p functionally independent unbiased estimating functions, i.e., $u_j(W, \psi, \beta)$, such that $E[u_j(W, \psi, \beta)] = 0$, $j = 1, \ldots, p$. By using vector notation, we have

$$U(W, \psi, \beta) = (u_1(W, \psi, \beta), \ldots, u_p(W, \psi, \beta))', \tag{5.1}$$

such that $E[U(W, \psi, \beta)] = 0$. We also assume that the information about the nuisance parameter β is available through q functionally independent unbiased estimating functions that are not involved in ψ, i.e., $v_k(W, \beta)$, such that $E[v_k(W, \beta)] = 0$, $k = 1, \ldots, q$. In vector form, we have

$$V(W, \beta) = (v_1(W, \beta), \ldots, v_q(W, \beta))', \tag{5.2}$$

such that $E[V(W, \beta)] = 0$.

Then the empirical likelihood for (ψ, β) can be defined as follows:

$$L(\psi, \beta) = \sup \left\{ \prod_{i=1}^{n} p_i : \text{each } p_i > 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i U(W_i, \psi, \beta) = 0 \right\}. \tag{5.3}$$

The nuisance parameter β in $L(\psi, \beta)$ could be consistently estimated by $\widehat{\beta}$ which is the solution to the estimating equation: $\frac{1}{n}\sum_{i=1}^{n} V(W_i, \beta) = 0$. By plugging $\widehat{\beta}$ into (5.3), we obtain a profile empirical likelihood for ψ:

$$\widehat{L}(\psi) = \sup \left\{ \prod_{i=1}^{n} p_i : \text{each } p_i > 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i U(W_i, \psi, \widehat{\beta}) = 0 \right\}. \tag{5.4}$$

By regular methods of EL, such as Qin and Lawless [43], the corresponding empirical likelihood ratio for ψ is:

$$\widehat{R}(\psi) = \prod_{i=1}^{n}(np_i) = \prod_{i=1}^{n}\left\{1 + \widehat{t}'U(W_i, \psi, \widehat{\beta})\right\}^{-1},$$

where $\widehat{t}$ is the solution of

$$\frac{1}{n}\sum_{i=1}^{n}\frac{U(W_i, \psi, \widehat{\beta})}{1 + \widehat{t}'U(W_i, \psi, \widehat{\beta})} = 0. \tag{5.5}$$

Then the empirical log-likelihood ratio for ψ is given by

$$\widehat{l}(\psi) \equiv -2\log\widehat{R}(\psi) = 2\sum_{i=1}^{n}\log\left\{1 + \widehat{t}'U(W_i, \psi, \widehat{\beta})\right\}. \tag{5.6}$$

Let

$$Q_{1n}(t, \psi, \beta) \equiv \frac{1}{n}\sum_{i=1}^{n}Q_1(W_i, t, \psi, \beta) \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{U(W_i, \psi, \beta)}{1 + t'U(W_i, \psi, \beta)}, \tag{5.7}$$

$$Q_{2n}(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n}V(W_i, \beta). \tag{5.8}$$

Then, $Q_{1n}(0, \psi, \beta) = \frac{1}{n}\sum_{i=1}^{n}U(W_i, \psi, \beta)$, which will be used in the Taylor expansion at $t = 0$.

Under mild regularity conditions, we obtain the asymptotic distribution of the empirical log-likelihood ratio for ψ, which is a weighted sum of independent chi-squared distributions, in the following theorem.

Theorem 5 *Assume that* (ψ_0, β_0) *is the true value of* (ψ, β), $E\left[\frac{\partial U(W,\psi_0,\beta_0)}{\partial\psi'}\right]$ *and* $E\left[\frac{\partial V(W,\beta_0)}{\partial\beta'}\right]$ *are negatively definite. Then,*

$$\widehat{l}(\psi_0) \xrightarrow{d} r_1\chi^2_{1,1} + \cdots + r_p\chi^2_{1,p}, \tag{5.9}$$

where $\chi^2_{1,j}, j = 1, \ldots, p$ *are independent chi-squared random variables with one degree of freedom, and the weights* $r_1, \ldots, r_p$ *are* p *none-zero eigenvalues of* Λ*, defined as follows:*

$$\begin{aligned}
\Lambda &= (S^*)^{\frac{1}{2}} \begin{pmatrix} I \\ (-S_{22})^{-1} S_{12}' \end{pmatrix} (-S_{11})^{-1} \begin{pmatrix} I & S_{12}(-S_{22})^{-1} \end{pmatrix} (S^*)^{\frac{1}{2}}, \\
S^* &= Cov[(U'(W, \psi_0, \beta_0), V'(W, \beta_0))'], \\
S_{11} &= E\left[\frac{\partial Q_1(W, 0, \psi_0, \beta_0)}{\partial t'}\right], \\
S_{12} &= E\left[\frac{\partial Q_1(W, 0, \psi_0, \beta_0)}{\partial \beta'}\right], \\
S_{22} &= E\left[\frac{\partial V(W, \beta_0)}{\partial \beta'}\right].
\end{aligned}$$

Remark 1: The eigenvalues of Λ could be calculated by solving the eigenvalues of

$$\Lambda^* = (-S_{11})^{-1} \begin{pmatrix} I & S_{12}(-S_{22})^{-1} \end{pmatrix} S^* \begin{pmatrix} I \\ (-S_{22})^{-1} S_{12}' \end{pmatrix}.$$

Remark 2: This theorem could be treated as a special case of Theorem 2.1 provided by Hjort *et al.* [44]. But we proved it with a usual EL approach. In our framework, nuisance parameters are estimated from another set of estimating equations. All conditions from (A0) to (A3) in Theorem 2.1 could be verified directly. Also, explicit formulas of V_1 and V_2 in Theorem 2.1 are derived, whereas they are conditions in Theorem 2.1. By using notations in Theorem 2.1, $U \sim \mathcal{N}_p(0, V_1)$ and $V_2 = -S_{11}$, where

$$V_1 = \begin{pmatrix} I & S_{12}(-S_{22})^{-1} \end{pmatrix} S^* \begin{pmatrix} I \\ (-S_{22})^{-1} S_{12}' \end{pmatrix}.$$

It is easy to see Λ^* and $V_2^{-1} V_1$ share same none-zero eigenvalues. Therefore, these two theorems coincide.

When a consistent estimate of Λ is available, Theorem 5 can be used to make inference

for the parameters of interest.

5.2.2 Bias-corrected Empirical Likelihood Confidence Regions

In this part, we try to derive various bias-corrected joint empirical likelihood confidence regions for the sensitivity and the specificity as well as the cut-off level based on the FI, MSI, IPW and SPE methods.

Let T_i denote the continuous test result from a screening test, and let D_i denote the binary disease status without measurement error, $i = 1, \ldots, n$, where $D_i = 1$ indicates the ith patient is diseased and $D_i = 0$ indicates the ith patient is free of disease. Due to various causes, such as cost limits and privacy security, only a subset of patients have their disease statuses verified; let V_i denote the binary verification status of the ith patient, with $V_i = 1$ if the ith patient has the true disease status verified, and $V_i = 0$ if otherwise. In practice, some covariate information, other than the results from the screening test, can be obtained. Let A_i be a vector of observed covariates for the ith patient that may be associated with both D_i and V_i.

When all patients are verified, i.e., $V_i = 1, i = 1, \ldots, n$, a complete data set is obtained. In this case, for any cut-off level τ, the sensitivity $\theta(\tau)$, and the specificity $\eta(\tau)$ can be estimated by

$$\widehat{\theta}_{Full}(\tau) = \frac{\sum_{i=1}^{n} I(T_i > \tau) D_i}{\sum_{i=1}^{n} D_i}, \quad \widehat{\eta}_{Full}(\tau) = \frac{\sum_{i=1}^{n} I(T_i \leq \tau)(1 - D_i)}{\sum_{i=1}^{n} (1 - D_i)}. \tag{5.10}$$

Obviously, $\widehat{\theta}_{Full}(\tau)$ and $\widehat{\eta}_{Full}(\tau)$ are unbiased estimators for θ and η respectively.

Many current studies center on the MAR assumption because it is manageable in practice. Under this assumption, whether one subject has his or her disease status verified is conditionally independent of the true disease status given test results and observed covariates, i.e., $V \perp D|T, A$ or $P(V|D, T, A) = P(V|T, A)$ or $P(D|V, T, A) = P(D|T, A)$. In other words, the decision to verify the patient's true disease status only depends on T and A regardless of the true disease status D. All following methods are based on this assumption.

Assume that $S_i = (T_i, A_i, V_i, D_i), i = 1, \ldots, n$ is an i.i.d. sample of $S = (T, A, V, D)$, and that V_i and D_i are conditionally independent on T_i and A_i. Let $\pi_i = P(V_i = 1|T_i, A_i)$ and $\rho_i = P(D_i = 1|T_i, A_i)$. Motivated by He *et al.* [27], based on the IPW estimator, we observed that

$$E\left[\sum_{i=1}^{n} \pi_i^{-1} V_i D_i (I(T_i \leq \tau) - (1-\theta))\right] = 0, \tag{5.11}$$

$$E\left[\sum_{i=1}^{n} \pi_i^{-1} V_i (1-D_i)(I(T_i \leq \tau) - \eta)\right] = 0. \tag{5.12}$$

Actually, with the FI, MSI and SPE methods, we could obtain similar estimating functions. Let's take the SPE method as an example, and the other two are straightforward. When misspecification of either verification model or disease model occurs, the semiparametric efficient (SPE) estimators for sensitivity and specificity were shown to be "doubly robust" ([57], [58], [28]). Hopefully, bias-corrected joint confidence regions with the SPE method would retain such a good property. Under the MAR assumption, we observe that

$$\begin{aligned}
& E[I(T_i \leq \tau)\{V_i D_i/\pi_i - (V_i - \pi_i)\rho_i/\pi_i\}] \\
= \; & E\big[E[I(T_i \leq \tau)\{V_i D_i/\pi_i - (V_i - \pi_i)\rho_i/\pi_i\}]|T_i, A_i\big] \\
= \; & E\big[I(T_i \leq \tau)\frac{1}{\pi_i} E V_i D_i - \frac{\rho_i}{\pi_i}(E V_i - \pi_i)|T_i, A_i\big] \\
= \; & E\big[I(T_i \leq \tau)\frac{1}{\pi_i} E V_i E D_i - \frac{\rho_i}{\pi_i}(E V_i - \pi_i)|T_i, A_i\big] \\
= \; & E\big[I(T_i \leq \tau)\rho_i\big] \\
= \; & E\big[(1-\theta)\{V_i D_i/\pi_i - (V_i - \pi_i)\rho_i/\pi_i\}\big].
\end{aligned}$$

Thus,

$$E\left[\sum_{i=1}^{n}\{V_i D_i/\pi_i - (V_i - \pi_i)\rho_i/\pi_i\}\big(I(T_i \leq \tau) - (1-\theta)\big)\right] = 0. \tag{5.13}$$

Similarly, we have

$$E\left[\sum_{i=1}^{n}\{V_i(1-D_i)/\pi_i-(V_i-\pi_i)(1-\rho_i)/\pi_i\}\big(I(T_i\le\tau)-\eta\big)\right]=0. \tag{5.14}$$

Therefore, motivated by the IPW, FI, MSI and SPE methods, we can obtain following four pairs of estimating functions for $\vartheta=(\theta,\eta,\tau)'$:

$$\text{IPW:}\quad \begin{cases} g_{\text{IPW},1}(S_i,\vartheta,\pi_i)=\pi_i^{-1}V_iD_i(I(T_i\le\tau)-(1-\theta)) \\ g_{\text{IPW},2}(S_i,\vartheta,\pi_i)=\pi_i^{-1}V_i(1-D_i)(I(T_i\le\tau)-\eta) \end{cases}$$

$$\text{FI:}\quad \begin{cases} g_{\text{FI},1}(S_i,\vartheta,\rho_i)=\rho_i(I(T_i\le\tau)-(1-\theta)) \\ g_{\text{FI},2}(S_i,\vartheta,\rho_i)=(1-\rho_i)(I(T_i\le\tau)-\eta) \end{cases}$$

$$\text{MSI:}\quad \begin{cases} g_{\text{MSI},1}(S_i,\vartheta,\rho_i)=(V_iD_i+(1-V_i)\rho_i)(I(T_i\le\tau)-(1-\theta)) \\ g_{\text{MSI},2}(S_i,\vartheta,\rho_i)=(V_i(1-D_i)+(1-V_i)(1-\rho_i))(I(T_i\le\tau)-\eta) \end{cases}$$

$$\text{SPE:}\quad \begin{cases} g_{\text{SPE},1}(S_i,\vartheta,\pi_i,\rho_i)=(V_iD_i/\pi_i-(V_i-\pi_i)\rho_i/\pi_i)(I(T_i\le\tau)-(1-\theta)) \\ g_{\text{SPE},2}(S_i,\vartheta,\pi_i,\rho_i)=(V_i(1-D_i)/\pi_i-(V_i-\pi_i)(1-\rho_i)/\pi_i)(I(T_i\le\tau)-\eta) \end{cases}$$

Let

$$g_{\text{IPW}}(S_i,\vartheta,\pi_i) = (g_{\text{IPW},1}(S_i,\vartheta,\pi_i),g_{\text{IPW},2}(S_i,\vartheta,\pi_i))', \tag{5.15}$$

$$g_{\text{FI}}(S_i,\vartheta,\rho_i) = (g_{\text{FI},1}(S_i,\vartheta,\rho_i),g_{\text{FI},2}(S_i,\vartheta,\rho_i))', \tag{5.16}$$

$$g_{\text{MSI}}(S_i,\vartheta,\rho_i) = (g_{\text{MSI},1}(S_i,\vartheta,\rho_i),g_{\text{MSI},2}(S_i,\vartheta,\rho_i))', \tag{5.17}$$

$$g_{\text{SPE}}(S_i,\vartheta,\pi_i,\rho_i) = (g_{\text{SPE},1}(S_i,\vartheta,\pi_i,\rho_i),g_{\text{SPE},2}(S_i,\vartheta,\pi_i,\rho_i))'. \tag{5.18}$$

Because π_i's and ρ_i's are often unknown in practice, one way to profile them is to replace them by their consistent estimators. In general, with binary essentials of both D_i's and V_i's, $\pi_i=P(V_i=1|T_i,A_i)$ and $\rho_i=P(D_i=1|T_i,A_i)$ could be modeled by employing parametric models such as logistic regression or probit model. For illustration, we use probit model to model ρ_i's and logistic regressions to model π_i's as follows, which will be used in the

simulation studies:

$$\Phi^{-1}(\rho_i) = Z_i'\boldsymbol{\alpha}, \qquad \log\frac{\pi_i}{1-\pi_i} = Z_i'\boldsymbol{\beta},$$

where $Z_i = (1, T_i, A_i)', \boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)', \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$, and $\Phi(\cdot)$ is the standard normal distribution function. These two models could include different covariates. Here, we use same covariates for notation simplicity. Thus, $\rho_i = \Phi(Z_i'\boldsymbol{\alpha})$ and $\pi_i = \frac{1}{1+e^{-Z_i'\boldsymbol{\beta}}}$.

Actually, both logistic regressions and probit models could be included in a general estimation equation framework. Assume that $\rho_i = \rho(Z_i, \boldsymbol{\alpha})$ and $\pi_i = \pi(Z_i, \boldsymbol{\beta})$ with a known function ρ and π. Then the unknown parameter $\boldsymbol{\beta}$ could be estimated by the estimating equation $Q_{2n,2}(\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{i=1}^n Q_{2,2}(S_i, \boldsymbol{\beta}) = 0$, where $Q_{2,2}(S, \boldsymbol{\beta})$ is an estimating function satisfying $E[Q_{2,2}(S, \boldsymbol{\beta})] = 0$. In the logistic regression setting,

$$Q_{2n,2}(\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{i=1}^n Q_{2,2}(S_i, \boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{i=1}^n (V_i - \pi_i)Z_i. \tag{5.19}$$

For D_i's, exact disease statuses are only available for verified subjects with $V_i = 1$. Therefore, only verified observations could be employed to obtain an estimate of $\boldsymbol{\alpha}$. Then the unknown parameters $\boldsymbol{\alpha}$ could be estimated by the estimating equation $Q_{2n,1}(\boldsymbol{\alpha}) \equiv \frac{1}{n}\sum_i^n V_i Q_{2,1}(S_i, \boldsymbol{\alpha}) = 0$, where $Q_{2,1}(S, \boldsymbol{\alpha})$ is an estimating function satisfying $E[Q_{2,1}(S, \boldsymbol{\alpha})] = 0$. In the probit model setting,

$$Q_{2n,1}(\boldsymbol{\alpha}) \equiv \frac{1}{n}\sum_{i=1}^n V_i Q_{2,1}(S_i, \boldsymbol{\alpha}) \equiv \frac{1}{n}\sum_{i=1}^n \frac{V_i[D_i - \Phi(Z_i'\boldsymbol{\alpha})]\phi(Z_i'\boldsymbol{\alpha})}{\Phi(Z_i'\boldsymbol{\alpha})(1-\Phi(Z_i'\boldsymbol{\alpha}))} Z_i, \tag{5.20}$$

where $\phi(\cdot)$ is the standard normal density function.

After obtaining consistent estimators $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$ and $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, we can plug in consistent estimators $\widehat{\rho}_i = \rho(Z_i, \widehat{\boldsymbol{\alpha}})$ of ρ_i and $\widehat{\pi}_i = \pi(Z_i, \widehat{\boldsymbol{\beta}})$ of π_i into (5.15)-(5.18), and then obtain the profile empirical likelihood for ϑ:

$$\widehat{L}_P(\vartheta) = \sup\left\{\prod_{i=1}^n p_i : \text{each } p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{j=1}^n p_i g(S_i, \vartheta, \widehat{\psi}_i) = 0\right\}, \tag{5.21}$$

where $\widehat{\psi}_i$ could be $\widehat{\pi}_i$, $\widehat{\rho}_i$, or $(\widehat{\pi}_i, \widehat{\rho}_i)$. In the simulation studies in Section 5.3, we use estimates $\widehat{\rho}_i = \Phi(Z_i'\widehat{\boldsymbol{\alpha}})$ and $\widehat{\pi}_i = 1/(1 + e^{-Z_i'\widehat{\boldsymbol{\beta}}})$.

Then, the profile empirical log-likelihood ratio for ϑ is given by

$$\widehat{l}_P(\vartheta) = 2\sum_{i=1}^{n} \log(1 + \widehat{t}' g(S_i, \vartheta, \widehat{\psi}_i)), \tag{5.22}$$

where $\widehat{t}$ is the solution of the following equation:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{g(S_i, \vartheta, \widehat{\psi}_i)}{1 + \widehat{t}' g(S_i, \vartheta, \widehat{\psi}_i)} = 0.$$

Let

$$Q_{1n}(t, \vartheta, \boldsymbol{\psi}) \equiv \frac{1}{n}\sum_{i=1}^{n} Q_1(S_i, t, \vartheta, \boldsymbol{\psi}) \equiv \frac{1}{n}\sum_{i=1}^{n} \frac{g(S_i, \vartheta, \psi_i)}{1 + t' g(S_i, \vartheta, \psi_i)}, \tag{5.23}$$

where ψ_i could be π_i, ρ_i, or (π_i, ρ_i), depending on the selection of $g(S_i, \vartheta, \psi_i)$. Then, $Q_{1n}(0, \vartheta, \boldsymbol{\psi}) = \frac{1}{n}\sum_{i=1}^{n} g(S_i, \vartheta, \psi_i)$, and $Q_{1n}(0, \vartheta, \widehat{\boldsymbol{\psi}}) = \frac{1}{n}\sum_{i=1}^{n} g(S_i, \vartheta, \widehat{\psi}_i)$ where $\widehat{\psi}_i$ could be $\widehat{\pi}_i$, $\widehat{\rho}_i$, or $(\widehat{\pi}_i, \widehat{\rho}_i)$.

In the previous section, we provide a general framework combining the empirical likelihood and general estimation equations with nuisance parameters, which explicitly offers the limiting distribution, a weighted chi-squared distribution, of the empirical likelihood ratio statistic. By applying this framework to estimation functions (5.15)-(5.18), we can obtain the following results.

Theorem 6 *Assume that $(\vartheta_0, \boldsymbol{\beta}_0)$ is the true value of $(\vartheta, \boldsymbol{\beta})$. Then,*

$$\widehat{l}_{IPW}(\vartheta_0) = 2\sum_{i=1}^{n} \log(1 + \widehat{t}' g_{IPW}(S_i, \vartheta_0, \widehat{\pi}_i)) \xrightarrow{d} r_1\chi^2_{1,1} + r_2\chi^2_{1,2} \tag{5.24}$$

where $\{\chi^2_{1,j}, j = 1, 2\}$ are independent chi-squared random variables with one degree of free-

dom, and the weights r_1, r_2 are the two none-zero eigenvalues of Λ, defined as follows:

$$\begin{aligned}
\Lambda &= I - (-S_{11})^{-1} S_{12} (-S_{22})^{-1} S_{12}' \\
S_{11} &= E\left[\frac{\partial Q_1(S_i, 0, \vartheta_0, \boldsymbol{\beta}_0)}{\partial t'}\right] = -E[g_{IPW}(S_i, \vartheta_0, \pi_i) g_{IPW}'(S_i, \vartheta_0, \pi_i)] = -Cov(Q_1(S_i, 0, \vartheta_0, \boldsymbol{\beta}_0)), \\
S_{12} &= E\left[\frac{\partial Q_1(S_i, 0, \vartheta_0, \boldsymbol{\beta}_0)}{\partial \beta'}\right] = -Cov\left(Q_1(S_i, 0, \vartheta_0, \boldsymbol{\beta}_0), Q_2(S_i, \boldsymbol{\beta}_0)\right), \\
S_{22} &= E\left[\frac{\partial Q_2(S_i, \boldsymbol{\beta}_0)}{\partial \beta'}\right].
\end{aligned}$$

Additionally, if $Q_2(Z_i, \boldsymbol{\beta})$ is a score function, S_{22} has the form

$$S_{22} = -Cov[Q_2(S_i, \boldsymbol{\beta}_0)].$$

Theorem 7 *Assume that $(\vartheta_0, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ is the true value of $(\vartheta, \boldsymbol{\alpha}, \boldsymbol{\beta})$. Then,*

$$\left.\begin{aligned}
\widehat{l}_{FI}(\vartheta_0) = 2\textstyle\sum_{i=1}^n \log(1 + \widehat{t}' g_{FI}(S_i, \vartheta_0, \widehat{\rho}_i)) \\
\widehat{l}_{MSI}(\vartheta_0) = 2\textstyle\sum_{i=1}^n \log(1 + \widehat{t}' g_{MSI}(S_i, \vartheta_0, \widehat{\rho}_i))
\end{aligned}\right\} \xrightarrow{d} r_1 \chi_{1,1}^2 + r_2 \chi_{1,2}^2, \tag{5.25}$$

where $\{\chi_{1,j}^2, j = 1, 2\}$ are independent chi-squared random variables with one degree of freedom, and the weights r_1, r_2 are the two none-zero eigenvalues of Λ, defined as follows:

$$\begin{aligned}
\Lambda &= (S^*)^{\frac{1}{2}} \begin{pmatrix} I \\ (-S_{22})^{-1} S_{12}' \end{pmatrix} (-S_{11})^{-1} \begin{pmatrix} I & S_{12}(-S_{22})^{-1} \end{pmatrix} (S^*)^{\frac{1}{2}}, \\
S^* &= Cov\left[(g'(S_i, \vartheta_0, \psi_i), V_i Q_{2,1}'(S_i, \boldsymbol{\alpha}_0))'\right], \\
S_{11} &= E\left[\frac{\partial Q_1(S_i, 0, \vartheta_0, \boldsymbol{\alpha}_0)}{\partial t'}\right] = -E[g(S_i, \vartheta_0, \psi_i) g'(S_i, \vartheta_0, \psi_i)], \\
S_{12} &= E\left[\frac{\partial Q_1(S_i, 0, \vartheta_0, \boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}'}\right], \\
S_{22} &= E\left[\frac{V_i \partial Q_{2,1}(S_i, \boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}'}\right],
\end{aligned}$$

where $g(S_i, \vartheta_0, \psi_i)$ is $g_{FI}(S_i, \vartheta_0, \rho_i)$ for $\widehat{l}_{FI}(\vartheta_0)$, or $g_{MSI}(S_i, \vartheta, \rho_i)$ for $\widehat{l}_{MSI}(\vartheta_0)$.

Theorem 8 *Assume that* $(\vartheta_0, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ *is the true value of* $(\vartheta, \boldsymbol{\alpha}, \boldsymbol{\beta})$*. Then,*

$$\widehat{l}_{SPE}(\vartheta_0) = 2\sum_{i=1}^{n} \log(1 + \widehat{t}' g_{SPE}(S_i, \vartheta_0, \widehat{\pi}_i, \widehat{\rho}_i)) \xrightarrow{d} r_1\chi^2_{1,1} + r_2\chi^2_{1,2}, \tag{5.26}$$

where $\{\chi^2_{1,j}, j = 1, 2\}$ *are independent chi-squared random variables with one degree of freedom, and the weights* r_1, r_2 *are the two none-zero eigenvalues of* Λ*, defined as follows:*

$$\begin{aligned}
\Lambda &= (S^*)^{\frac{1}{2}} \begin{pmatrix} I \\ (-S_{22})^{-1}S'_{12} \end{pmatrix} (-S_{11})^{-1} \begin{pmatrix} I & S_{12}(-S_{22})^{-1} \end{pmatrix} (S^*)^{\frac{1}{2}}, \\
S^* &= Cov\left[(g'_{SPE}(S_i, \vartheta_0, \pi_i, \rho_i), V_i Q'_{2,1}(S_i, \boldsymbol{\alpha}_0), Q'_{2,2}(S_i, \boldsymbol{\beta}_0))'\right], \\
S_{11} &= E\left[\frac{\partial Q_1(S_i, 0, \vartheta_0, (\boldsymbol{\alpha}'_0, \boldsymbol{\beta}'_0)')}{\partial t'}\right] = -E[g_{SPE}(S_i, \vartheta_0, \pi_i, \rho_i) g'_{SPE}(S_i, \vartheta_0, \pi_i, \rho_i)], \\
S_{12} &= E\left[\left(\frac{\partial Q_1(S_i, 0, \vartheta_0, (\boldsymbol{\alpha}'_0, \boldsymbol{\beta}'_0)')}{\partial \boldsymbol{\alpha}'}, \frac{\partial Q_1(S_i, 0, \vartheta_0, (\boldsymbol{\alpha}'_0, \boldsymbol{\beta}'_0)')}{\partial \boldsymbol{\beta}'}\right)\right], \\
S_{22} &= \begin{pmatrix} E\left[\frac{V_i \partial Q_{2,1}(S_i, \boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}'}\right] & 0 \\ 0 & E\left[\frac{\partial Q_{2,2}(S_i, \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}'}\right] \end{pmatrix}.
\end{aligned}$$

Theorems 6-8 provide $\widehat{l}_{\mathrm{IPW}}(\vartheta) = \widehat{l}_{\mathrm{IPW}}(\theta, \eta, \tau)$, $\widehat{l}_{\mathrm{FI}}(\vartheta) = \widehat{l}_{\mathrm{FI}}(\theta, \eta, \tau)$, $\widehat{l}_{\mathrm{MSI}}(\vartheta) = \widehat{l}_{\mathrm{MSI}}(\theta, \eta, \tau)$ and $\widehat{l}_{\mathrm{SPE}}(\vartheta) = \widehat{l}_{\mathrm{SPE}}(\theta, \eta, \tau)$ as asymptotic pivots in the presence of verification bias for the inference of parameter pairs (θ, τ), (η, τ) or (θ, η), given a fixed value of the third remaining parameter. Matrices Λ in Theorem 6-8 are still unknown, but they can be consistently estimated by their empirical counterparts which can be obtained by plugging in Λ the bias-corrected estimators $\widehat{\theta}$ and $\widehat{\eta}$ (obtained from (1.6), (1.7) or (1.9)) at a pre-determined cut-off level τ. Taking Theorem 7 as an example, Λ can be estimated by

$$\widehat{\Lambda} = (\widehat{S}^*)^{\frac{1}{2}} \begin{pmatrix} I \\ (-\widehat{S}_{22})^{-1}\widehat{S}'_{12} \end{pmatrix} (-\widehat{S}_{11})^{-1} \begin{pmatrix} I & \widehat{S}_{12}(-\widehat{S}_{22})^{-1} \end{pmatrix} (\widehat{S}^*)^{\frac{1}{2}},$$

where

$$
\begin{aligned}
\widehat{S}^* &= \frac{1}{n}\sum_{i=1}^{n}\left[(g'_{\mathrm{SPE}}(S_i,\widehat{\vartheta},\widehat{\pi}_i,\widehat{\rho}_i), V_i Q'_{2,1}(S_i,\widehat{\boldsymbol{\alpha}}), Q'_{2,2}(S_i,\widehat{\boldsymbol{\beta}}))'\right],\\
\widehat{S}_{11} &= -\frac{1}{n}\sum_{i=1}^{n} g_{\mathrm{SPE}}(S_i,\widehat{\vartheta},\widehat{\pi}_i,\widehat{\rho}_i) g'_{\mathrm{SPE}}(S_i,\widehat{\vartheta},\widehat{\pi}_i,\widehat{\rho}_i),\\
\widehat{S}_{12} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial Q_1(S_i,0,\widehat{\vartheta},(\widehat{\boldsymbol{\alpha}}',\widehat{\boldsymbol{\beta}}')')}{\partial\boldsymbol{\alpha}'}, \frac{\partial Q_1(S_i,0,\widehat{\vartheta},(\widehat{\boldsymbol{\alpha}}',\widehat{\boldsymbol{\beta}}')')}{\partial\boldsymbol{\beta}'}\right),\\
\widehat{S}_{22} &= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\frac{V_i\partial Q_{2,1}(S_i,\widehat{\boldsymbol{\alpha}})}{\partial\boldsymbol{\alpha}'} & 0\\ 0 & \frac{1}{n}\sum_{i=1}^{n}\frac{\partial Q_{2,2}(S_i,\widehat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}'}\end{pmatrix}.
\end{aligned}
$$

To our surprise, by taking into account the variation in estimating π_i's, the eigenvalues of Λ in Theorem 6 would be smaller than 1, which means the weighted chi-squared distribution is less variant than the chi-squared distribution with degrees of freedom 2. This is probably because S_{12} is the negative of $\mathrm{Cov}\,(Q_1(S_i,0,\vartheta_0,\boldsymbol{\beta}_0), Q_2(S_i,\boldsymbol{\beta}_0))$ in this special model. Also, it is pointed out by Robins *et al.* [67] that the inverse weighted estimating equations with known propensity have larger variation than those with unknown propensity, and this is part of the reason that the eigenvalues of Λ are smaller than 1. Our observation will be verified in simulation studies.

However, verification probabilities are known in some studies. In these cases, the above result could be reduced when nuisance parameters ρ_i's and π_i's are known, and the result is given in the following corollary which is consistent with that obtained by Qin and Lawless [43] or Adimari and Guolo [68].

Corollary 2 *If ρ_i's and π_i's are known, then S_{12} in Theorems 6-8 is equal to 0. Furthermore,*

$$l_P(\vartheta_0) \xrightarrow{d} \chi^2_2,$$

and $l_P(\vartheta_0) = 2\sum_{i=1}^{n}\log(1+\widehat{t}'g(S_i,\vartheta_0,\psi_i))$, where $\widehat{t}$ is the solution to $\frac{1}{n}\sum_{i=1}^{n}\frac{g(S_i,\vartheta,\psi_i)}{1+\widehat{t}'g(S_i,\vartheta,\psi_i)} = 0$ where ψ_i could be π_i, ρ_i, or (π_i,ρ_i).

By utilizing Theorems 6-8, at the level of $(1-\alpha)100\%$, we can construct four types,

i.e., IPW, FI, MSI and SPE, of profile empirical likelihood confidence regions (called EL(F) regions) for three pairs of parameters as follows:

- $\mathscr{R}_{\alpha,1}(\theta,\tau) = \{(\theta,\tau) : \widehat{l}_P(\theta,\eta_0,\tau) \leq c_\alpha\}$;
- $\mathscr{R}_{\alpha,2}(\eta,\tau) = \{(\eta,\tau) : \widehat{l}_P(\theta_0,\eta,\tau) \leq c_\alpha\}$;
- $\mathscr{R}_{\alpha,3}(\theta,\eta) = \{(\theta,\eta) : \widehat{l}_P(\theta,\eta,\tau_0) \leq c_\alpha\}$;

where $\alpha \in (0,1)$, $\widehat{l}_P(\cdot)$ could be any one of $\widehat{l}_{\text{IPW}}(\cdot)$, $\widehat{l}_{\text{FI}}(\cdot)$, $\widehat{l}_{\text{MSI}}(\cdot)$ and $\widehat{l}_{\text{SPE}}(\cdot)$, and c_α is the $(1-\alpha)$-th quantile of the distributions $r_1\chi^2_{1,1} + r_2\chi^2_{1,2}$ for $\widehat{l}_{\text{IPW}}(\vartheta_0)$, $\widehat{l}_{\text{FI}}(\vartheta_0)$, $\widehat{l}_{\text{MSI}}(\vartheta_0)$, and $\widehat{l}_{\text{SPE}}(\vartheta_0)$. Note that the Monte Carlo simulation is needed to calculate the critical value c_α. This can be done by first finding consistent estimates $\widehat{r}_i$'s of r_i's and generating a large number of copies of $\widehat{r}_1\chi^2_{1,1} + \widehat{r}_2\chi^2_{1,2}$, and then taking c_α to be the $(1-\alpha)$-*th* sample quantile of these copies. In our simulation studies, we generate 40000 random copies from a standard chi-square distribution with degree of freedom 1 to obtain the $(1-\alpha)$-th quantile of this weighted chi-square distribution, and the computation is fast. When π_i's and ρ_i's are known, $c_\alpha = \chi^2_{2,1-\alpha}$ is the $(1-\alpha)$-th quantile of the chi-squared distribution with degrees of freedom 2. We can get the reduced profile empirical likelihood confidence regions (called EL(R) regions) by using Corollary 2.

These confidence regions could provide a good solution to the problem of selecting a reasonable cut-off point for a continuous-scale diagnostic test in a flexible manner. Depending on the availability of disease models or verification models, we can apply bias-corrected confidence regions correspondingly. If only disease models are available, we can apply FI or MSI regions; if only verification models are available, we can apply IPW regions; if both models are available or only one of them is correctly specified, SPE regions could be applied. To find the EL-based confidence region for sensitivity and specificity, firstly, at a desirable sensitivity value θ_0 and specificity value η_0, we construct the corresponding EL(R) regions $\mathscr{R}_{\alpha,2}(\eta,\tau)$ and $\mathscr{R}_{\alpha,1}(\theta,\tau)$ respectively. Secondly, by plotting these two regions on the same graph along with cut-off level as the horizontal axis, a reasonable cut-off level τ_0 could be identified from the overlapping part of these two regions. Finally, the EL(F) region $\mathscr{R}_{\alpha,3}(\theta,\eta)$

for the sensitivity and the specificity can be constructed at the selected cut-off level τ_0. Later in the real data analysis, we will show the entire process for the diagnosis of depression in elderly patients.

5.3 Simulation Studies

In this section, we conduct simulation studies to compare the finite sample performance and robustness of the proposed bias-corrected empirical likelihood confidence regions.

Simulation settings in the presence of verification bias are similar to those in Alonzo and Pepe [24] and He *et al.* [27]. Firstly, two independent underlying continuous disease processes are generated, saying $Z_1 \sim \mathcal{N}(0, 0.5)$ and $Z_2 \sim \mathcal{N}(0, 0.5)$. The disease status indicator random variable D is generated as a binary variable indicating whether a random variable $Z = Z_1 + Z_2 \sim \mathcal{N}(0, 1)$ exceeds a certain threshold h, which determines the disease prevalence. The continuous diagnostic test result T and the auxiliary covariate A are generated to be related to D through Z_1 and Z_2: $T = \nu_1 Z_1 + \kappa_1 Z_2 + \varepsilon_1$ and $A = \nu_2 Z_1 + \kappa_2 Z_2 + \varepsilon_2$, where $\varepsilon_1 \sim \mathcal{N}(0, 0.25)$ and $\varepsilon_2 \sim \mathcal{N}(0, 0.25)$ are independent. The extent to which the test result T and the covariate A are correlated with each other, and the effect of different components of the underlying disease process on the test result vary as one changes ν_1, ν_2, κ_1 and κ_2. The explanations of different values of ν_1, ν_2, κ_1 and κ_2 were discussed by Alonzo and Pepe [24]. Under this model setting, we could obtain the joint distribution of $(Z, T, A)'$, which follows a multivariate normal distribution:

$$\begin{pmatrix} Z \\ T \\ A \end{pmatrix} \sim \mathcal{N}_3 \left(\mathbf{0}, \begin{pmatrix} 1 & 0.5\nu_1^2 + 0.5\kappa_1^2 & 0.5\nu_2^2 + 0.5\kappa_2^2 \\ 0.5\nu_1^2 + 0.5\kappa_1^2 & 0.5\nu_1^2 + 0.5\kappa_1^2 + 0.25 & 0.5\nu_1\nu_2 + 0.5\kappa_1\kappa_2 \\ 0.5\nu_2^2 + 0.5\kappa_2^2 & 0.5\nu_1\nu_2 + 0.5\kappa_1\kappa_2 & 0.5\nu_2^2 + 0.5\kappa_2^2 + 0.25 \end{pmatrix} \right).$$

This joint distribution could be used to get the true values of the specificity η and the sensitivity θ, given the cut-off level τ and the disease prevalence h. This joint distribution could be used to get the true values of the specificity η and the sensitivity θ, given the cut-off level τ and the disease prevalence h.

5.3.1 Correct Models

In simulation studies, the verification probability $\pi(Z, \boldsymbol{\beta})$ is chosen to be a specified function of $Z = (1, T, A)'$ to match the MAR assumption, and the parameter $\boldsymbol{\beta}$ could be estimated by an estimating equation. Here, we set $\log(\frac{\pi}{1-\pi}) = -0.7 + T + A$ with $\pi = P(V = 1|T, A)$, and D is assumed to be missing for those subjects with $V = 0$. Thus roughly 40% subjects will have their disease statuses verified. More specifically, 20%-30% nondiseased subjects will have their disease statuses verified, and roughly 70%-80% of the diseased subjects will have their disease statuses verified in different settings. Additionally, FI, MSI and SPE methods require a parametric model for the probability, ρ_i's, of getting a disease. It was shown in Alonzo and Pepe [24] that a probit model that was linear in T and A was correct under this simulation setting.

To evaluate the performance of the proposed various EL(F) and EL(R) confidence regions in terms of coverage probability, 4000 random samples are generated from the underlying distributions with sample sizes $n = 200, 400$ and 500 respectively. In this part, we set $\nu_2 = \kappa_2 = 1$, and select h such that the prevalence of disease equals 0.3 and 0.5. Different pairs of (ν_1, κ_1) are selected to make comparison. For the purpose of comparison, confidence regions based on the normal approximation (denoted by NA) of GEE estimators with IPW, FI, MSI and SPE methods proposed by Alonzo and Pepe [24] are also included in the study.

In Table 5.1 and 5.2, we present coverage probabilities of IPW, FI, MSI and SPE based bias-corrected confidence regions with nominal levels 90% and 95% for various values of the pair (θ, η) at different levels of disease prevalence, ν_1, κ_1 and τ, in the presence of verification bias under the above model setting. Simulation results in these tables indicate that the proposed four bias-corrected EL(F) confidence regions work well with moderate sample size cases ($n \geq 400$). From these tables, we also observe that only if (ν_1, κ_1) is comparable with (ν_2, κ_2) ($\nu_1 = \kappa_1 = 1$ and $\nu_2 = \kappa_2 = 1$), better sensitivity and better specificity could be achieved simultaneously by carefully selecting the cut-off level τ. The IPW bias-corrected EL(R) regions are more conservative than the EL(F) regions. In contrast, the FI and MSI bias-corrected EL(R) regions severely under-cover (θ, η) compared with the

corresponding EL(F) regions. The SPE bias-corrected EL(R) regions perform similarly with the EL(F) regions, and this probably results from the property of double robustness of the SPE method. Thus the reduced method in Corollary 2 must be employed cautiously.

If both verification model and disease model are correctly specified, any one of IPW, FI, MSI and SPE based bias-corrected EL(F) regions is valid, but MSI-based EL(F) region is preferred. From Table 5.1 and 5.2, when sample sizes are relatively small ($n = 200$), IPW, FI and SPE based EL(F) regions slightly under-cover (θ, η). On the contrary, MSI-based EL(F) regions work well in most settings. Three possible reasons for this preference are that (a) MSI method is relatively simple to apply than SPE method, (b) only one part of D_i's are required to be imputed compared with the FI method, and (c) the maximum likelihood estimators in the logistic regression model with the IPW method are unstable and biased when the sample size is not large enough [69]. When one does not have a reasonable cut-off level, one can use the SPE-based EL(R) regions $\mathscr{R}_{\alpha,1}$ and $\mathscr{R}_{\alpha,2}$ to identify a reasonable cut-off level. Compared with all proposed methods, the IPW, FI and MSI based NA regions under-cover true parameters, and their performance improves slowly when the sample size increases as expected. The SPE-based NA regions could perform well in some settings especially when the sensitivity and the specificity are close to each other, but the overall performances are not stable. Additionally, GEE estimates involve the selection of initial values, which will influence the convergence of the algorithm. Also, Alonzo and Pepe [24] set the sample size to 5000 with prevalence 0.1 in the simulation, and this large sample size setting may not be applicable in practice. Our proposed methods work well in much smaller sample size cases. Thus they are more applicable in real settings.

5.3.2 Misspecified Models

Till now, disease models and verification models are correctly specified in the simulation. But in practice, misspecification of underlying models is possible. In the following, we will discuss the robustness of proposed bias-corrected joint empirical likelihood confidence regions.

To introduce misspecification, similar to those used in Alonzo and Pepe [24], V is generated from a Bernoulli random variable with $P(V = 1) = 1$ for subjects with $T > t^{(0.8)}$ and $P(V = 1) = 0.2$ for others, where $t^{(0.8)}$ is the 80-th quantile of the distribution of T. However, we still apply logistic regression to model V, which results in the misspecification of the verification model.

In correct models, the disease status is generated as $D = I(Z_1 + Z_2 > h)$, and T and A are generated from linear combinations of Z_1 and Z_2. Following the discussion in [24], with $\nu_1 = 1$ and $\tau = 0$ for T and $\nu_2 = 0$ and $\tau_2 = 1$ for A, the disease model $P(D|T)$ that is linear in T is misspecified. Here we apply probit models to D on T for verified subjects to simulate misspecification. The SPE estimator was shown to be doubly robust in [57] and [58]. It is expected that bias-corrected joint empirical likelihood confidence regions based on $\widehat{l}_{\text{SPE}}(\vartheta_0)$ are still doubly robust.

When either disease models or verification models are misspecified, our simulation results (not reported here) indicate that all other methods perform poorly except SPE-based regions. Thus only results for SPE-based regions are presented here. Similar with observations in correct models, Tables 5.3 and 5.4 indicate that the proposed SPE-based bias-corrected EL(F) confidence regions work well with moderate sample size cases ($n \geq 400$) when misspecified disease models and verification models are present, respectively. But in these cases, EL(R) regions break down due to the misspecification. SPE-based NA regions could still perform well in some settings, but their performances are not stable. Probably, larger sample size (e.g., sample size set to be 5000 in [24] is required to obtain good performance for SPE-based NA regions.

5.4 Study of Depression in Elderly Patients Recruited from Primary-care Practices

We apply proposed various bias-corrected joint empirical likelihood confidence regions to the data set from a longitudinal study of depression in elderly patients (age ≥ 65) recruited from primary-care practices in Monroe County, New York. This data set is provided by

He *et al.* [27] who directly estimated the area under the ROC curve in the presence of verification bias. The purpose of this analysis is to run a comparative study of choosing a reasonable cut-off point for the Hamilton Depression Rating Scale (HAM-D) and to evaluate the accuracy of the HAM-D in diagnosing depression in terms of specificity and sensitivity. Because the full data set is available, we could compare all proposed bias-corrected regions with regions from the full data.

The HAM-D, a 24-item observer-rated scale designed to measure the severity of depression, is treated as a screening marker for the diagnosis of depression. The HAM-D takes much shorter time, approximately 15-20 minutes, to administer, compared with 1-3 hours to administer the Structured Clinical Interview for DSM-IV (SCID), an intensive examiner-based assessment that could be used as a practical gold standard for this medical diagnosis [70]. During the collection process, 708 patients were recruited, and they were evaluated by a comprehensive diagnostic assessment for depression using the SCID. Based on the SCID, 249 patients were diagnosed as having depression and 459 patients were diagnosed as being free of depression. Other auxiliary information was also collected, including the HAM-D, age, gender, years of education, and the Cumulative Illness Rating Scale (CIRS). The CIRS is a reliable and valid measure of medical burden that quantifies the amount of pathology in each organ system [71].

Data for both SCID and HAM-D were collected from recruited patients in the data set. To go through the whole procedure of evaluating the diagnostic accuracy of HAM-D in the presence of verification bias, a verified subset of the data should be obtained to simulate the process of a two-phase design. In this subset, HAM-D results were obtained for all patients, but SCID results were only available for certain patients selected according to the following mechanism:

$$\log \frac{P(\text{SCID available})}{1 - P(\text{SCID available})} = -1 + 5I(\text{HAM-D} > 7) + 4I(\text{CIRS} > 7)I(\text{Age} < 75).$$

Therefore, similar to the verification mechanism in [27], patients who had a HAM-D score

> 7 or patients under the age of 75 with a relatively high cumulative illness burden are more likely to be selected by our verification mechanism in the log-odds-ratio way. Our verification mechanism selected 477 patients of the 708 patients (67.4%) to have their depression statuses to be verified. Because of the availability of the full data, for patients with disease, the verification rate is 86.7%; for patients without disease, the verification rate is 56.9%.

In this example, D=SCID diagnosis, T=HAM-D, and other factors, including age, gender, years of education, and CIRS score were treated as covariates A. In order to apply proposed methods in the presence of verification bias, a model for both $P(V = 1|T, A)$ and $P(D = 1|T, A)$ are required. Here, we use a logistic model with $I(\text{HAM-D} > 7)$ and $I(\text{CIRS} > 7)I(\text{Age} < 75)$ as covariates to V. For disease status D if available, we tried both logistic regression and probit models with $I(\text{HAM-D} > 7)$ and $I(\text{CIRS} > 7)I(\text{Age} < 75)$ as covariates, and there is no big difference. Thus we only use probit models here to model D. It is noted that the disease model is unavailable in this study, and probit models of D may be misspecified. Then it is of great interest to apply SPE-based confidence regions to get robust results. Also we provide MSI-based confidence regions and confidence regions from the full data generated by the method in [4] for comparison purpose.

To show how proposed empirical likelihood confidence regions could be used to identify a reasonable cut-off point that results in both higher sensitivity and higher specificity, we borrow ideas from the real data analysis part in [4]. Firstly, proper joint inferences on the pair (sensitivity, cut-off level) at a fixed specificity value and on the pair (specificity, cut-off level) at a fixed sensitivity value are required to investigate the relationship between the sensitivity/specificity and the cut-off level, when a certain specificity/sensitivity (e.g., 0.7, 0.8) is required. Due to the less powerful essential of HAM-D, the sensitivity and the specificity of the HAM-D are fixed at a moderate level of 0.75 for both cases after a series of comparison among possible ranges. Since the sensitivity θ and specificity η of the HAM-D depend on the cut-off level τ, EL(R) regions are adopted to choose τ. This is not a big problem for SPE-based regions because of their robustness reported in our simulation studies. Contour curves are employed to show confidence regions by calculating $\widehat{l}_P(\theta, \eta_0, \tau)$

and $\widehat{l}_P(\theta_0, \eta, \tau)$ values on a fine grid and connecting these values according to $(1-\alpha)$-th quantiles.

The left panel in Figure 5.1 shows the contour curves of $\widehat{l}_{SPE}(\theta, 0.75, \tau)$, giving confidence regions for the pair (sensitivity, cut-off level) at different nominal confidence levels 90%, 95% and 99%. The plot shows that, at the nominal 95% level, a variety of pairs of values for (θ, τ) are compatible with the target specificity level of 0.75. Due to integer HAM-D scores, the valid values of τ are limited to 8. The right panel in Figure 5.1 shows the contour curves of $\widehat{l}_{SPE}(0.75, \eta, \tau)$, giving confidence regions for the pair (specificity, cut-off level) at different nominal levels 90%, 95% and 99%. This plot also shows that, at the nominal 95% level, a variety of pairs of values for (η, τ) are compatible with the target sensitivity level of 0.75, and the valid values of τ are limited from 8 to 9. Similarly we have Figure 5.2 for $\widehat{l}_{MSI}(\theta, 0.75, \tau)$ and $\widehat{l}_{MSI}(0.75, \eta, \tau)$, and Figure 5.3 for confidence regions from full data, respectively. All three figures offer similar shapes of confidence regions at nominal levels 90% and 95%. For 99% confidence regions, they differ due to the introduction of verification bias. Also, probit model assumption for D seems to be reasonable by comparing the first two figures with Figure 5.3.

In order to simultaneously obtain reasonably good sensitivity and specificity, a reasonable cut-off level should be carefully selected. The joint confidence regions offer a good chance to search for cut-off levels compatible with both specificity and sensitivity at 0.75. Figure 5.4-5.6 shows both the 95% confidence region for the pair (sensitivity, cut-off level) at the fixed specificity level of 0.75 and the 95% confidence region for the pair (specificity, cut-off level) at the fixed sensitivity level of 0.75, provided by SPE-based regions, MSI-based regions and regions from full data, respectively. The cut-off level 8 could be identified as a reasonable value.

Figure 5.7 to 5.9 show the contour curves of $\widehat{l}_{SPE}(\theta, \eta, 8)$, $\widehat{l}_{MSI}(\theta, \eta, 8)$ and $\widehat{l}(\theta, \eta, 8)$ with full data, respectively, indicating the joint 95% EL(F) confidence regions for the pair (specificity, sensitivity) with the cut-off level fixed at 8. "True" sensitivity and specificity of HAM-D could be obtained nonparametrically because the full data set is available. When

the cut-off level is 8, $\eta = 0.752$ and $\theta = 0.767$, marked as points in all three figures. It is clear that "true" values are covered by their corresponding 95% confidence regions. When the cut-off level is fixed at 8, the specificity and the sensitivity never fall below 0.6 and 0.5, respectively. MSI-based regions are relatively more conservative. Generally speaking, the performance of the HAM-D, treated as the screening test, is not satisfactory enough. Thus the gold standard SCID is required to verify the depression status.

Table 5.1 Correct models: Coverage probabilities of various joint empirical likelihood confidence regions with nominal confidence level 90% in the presence of verification bias. Pre. means disease prevalence; EL(F) stands for the proposed method based on weighted chi-squared method; EL(R) stands for the reduced method from Corollary 2; NA means the normality approximation method.

							$n = 200$			$n = 400$			$n = 500$		
Type	Pre.	τ_0	ν_1	κ_1	η_0	θ_0	EL(F)	EL(R)	NA	EL(F)	EL(R)	NA	EL(F)	EL(R)	NA
IPW	0.3	0.2	1	1	0.783	0.924	0.887	0.925	0.912	0.904	0.946	0.836	0.904	0.944	0.848
		0.4	1	1	0.855	0.864	0.871	0.914	0.882	0.886	0.927	0.874	0.892	0.932	0.884
		0.15	1	0	0.690	0.715	0.880	0.930	0.907	0.901	0.952	0.875	0.892	0.942	0.877
		-0.2	1	0	0.520	0.850	0.865	0.908	0.840	0.895	0.934	0.846	0.895	0.892	0.829
		0.15	0	1	0.690	0.715	0.889	0.929	0.911	0.894	0.942	0.884	0.899	0.940	0.886
		-0.2	0	1	0.520	0.850	0.872	0.917	0.838	0.903	0.940	0.843	0.886	0.933	0.838
	0.5	0	1	1	0.852	0.852	0.873	0.911	0.850	0.896	0.922	0.879	0.899	0.922	0.882
		-0.2	1	1	0.771	0.913	0.876	0.908	0.867	0.910	0.931	0.838	0.898	0.932	0.826
		0	1	0	0.696	0.696	0.887	0.928	0.906	0.902	0.952	0.877	0.897	0.941	0.885
		-0.4	1	0	0.495	0.851	0.840	0.877	0.807	0.899	0.931	0.825	0.892	0.932	0.843
		0	0	1	0.696	0.696	0.873	0.921	0.902	0.907	0.956	0.887	0.902	0.950	0.884
		-0.4	0	1	0.495	0.851	0.846	0.885	0.799	0.902	0.938	0.837	0.908	0.940	0.828
FI	0.3	0.2	1	1	0.783	0.924	0.884	0.617	0.825	0.891	0.611	0.856	0.899	0.622	0.873
		0.4	1	1	0.855	0.864	0.884	0.631	0.836	0.894	0.624	0.861	0.908	0.654	0.883
		0.15	1	0	0.690	0.715	0.891	0.733	0.867	0.892	0.737	0.884	0.900	0.753	0.891
		-0.2	1	0	0.520	0.850	0.891	0.735	0.859	0.891	0.739	0.868	0.900	0.745	0.880
		0.15	0	1	0.690	0.715	0.882	0.731	0.870	0.892	0.735	0.884	0.903	0.747	0.896
		-0.2	0	1	0.520	0.850	0.886	0.740	0.855	0.894	0.742	0.873	0.896	0.738	0.882
	0.5	0	1	1	0.852	0.852	0.870	0.530	0.807	0.902	0.540	0.846	0.901	0.545	0.870
		-0.2	1	1	0.771	0.913	0.868	0.528	0.798	0.896	0.519	0.839	0.897	0.523	0.861
		0	1	0	0.696	0.696	0.882	0.688	0.865	0.893	0.692	0.879	0.896	0.707	0.893
		-0.4	1	0	0.495	0.851	0.880	0.696	0.855	0.888	0.701	0.876	0.889	0.707	0.883
		0	0	1	0.696	0.696	0.883	0.702	0.872	0.890	0.696	0.879	0.892	0.702	0.890
		-0.4	0	1	0.495	0.851	0.880	0.702	0.863	0.892	0.703	0.873	0.898	0.703	0.885
MSI	0.3	0.2	1	1	0.783	0.924	0.890	0.768	0.836	0.897	0.740	0.846	0.901	0.770	0.870
		0.4	1	1	0.855	0.864	0.895	0.760	0.847	0.897	0.747	0.864	0.910	0.780	0.896
		0.15	1	0	0.690	0.715	0.896	0.796	0.873	0.895	0.795	0.887	0.900	0.811	0.893
		-0.2	1	0	0.520	0.850	0.886	0.784	0.859	0.893	0.791	0.873	0.899	0.798	0.883
		0.15	0	1	0.690	0.715	0.891	0.793	0.874	0.898	0.799	0.888	0.902	0.803	0.893
		-0.2	0	1	0.520	0.850	0.892	0.795	0.861	0.896	0.793	0.873	0.890	0.792	0.878
	0.5	0	1	1	0.852	0.852	0.885	0.617	0.822	0.890	0.616	0.856	0.894	0.625	0.876
		-0.2	1	1	0.771	0.913	0.889	0.615	0.803	0.891	0.598	0.842	0.904	0.605	0.871
		0	1	0	0.696	0.696	0.887	0.734	0.867	0.891	0.732	0.877	0.898	0.746	0.896
		-0.4	1	0	0.495	0.851	0.883	0.726	0.851	0.891	0.732	0.881	0.894	0.734	0.885
		0	0	1	0.696	0.696	0.892	0.744	0.878	0.891	0.740	0.884	0.897	0.752	0.889
		-0.4	0	1	0.495	0.851	0.884	0.736	0.860	0.890	0.739	0.875	0.901	0.742	0.889
SPE	0.3	0.2	1	1	0.783	0.924	0.874	0.874	0.945	0.892	0.892	0.937	0.895	0.893	0.942
		0.4	1	1	0.855	0.864	0.886	0.887	0.946	0.892	0.892	0.939	0.903	0.901	0.952
		0.15	1	0	0.690	0.715	0.882	0.882	0.912	0.892	0.891	0.902	0.902	0.900	0.908
		-0.2	1	0	0.520	0.850	0.866	0.866	0.929	0.893	0.893	0.907	0.890	0.889	0.905
		0.15	0	1	0.690	0.715	0.888	0.888	0.917	0.895	0.895	0.907	0.901	0.899	0.907
		-0.2	0	1	0.520	0.850	0.885	0.885	0.919	0.890	0.889	0.898	0.893	0.892	0.898
	0.5	0	1	1	0.852	0.852	0.879	0.881	0.842	0.885	0.885	0.847	0.893	0.892	0.846
		-0.2	1	1	0.771	0.913	0.871	0.873	0.869	0.892	0.891	0.937	0.894	0.893	0.854
		0	1	0	0.696	0.696	0.877	0.877	0.895	0.893	0.892	0.840	0.897	0.896	0.849
		-0.4	1	0	0.495	0.851	0.874	0.875	0.914	0.891	0.890	0.870	0.890	0.889	0.859
		0	0	1	0.696	0.696	0.888	0.888	0.920	0.892	0.892	0.848	0.897	0.895	0.861
		-0.4	0	1	0.495	0.851	0.879	0.879	0.883	0.888	0.886	0.858	0.892	0.890	0.855

Table 5.2 Correct models: Coverage probabilities of various joint empirical likelihood confidence regions with nominal confidence level 95% in the presence of verification bias. Pre. means disease prevalence; EL(F) stands for the proposed method based on weighted chi-squared method; EL(R) stands for the reduced method from Corollary 2; NA means the normality approximation method.

							$n = 200$			$n = 400$			$n = 500$		
Type	Pre.	τ_0	ν_1	κ_1	η_0	θ_0	EL(F)	EL(R)	NA	EL(F)	EL(R)	NA	EL(F)	EL(R)	NA
IPW	0.3	0.2	1	1	0.783	0.924	0.938	0.966	0.956	0.961	0.980	0.890	0.962	0.981	0.901
		0.4	1	1	0.855	0.864	0.935	0.961	0.927	0.948	0.976	0.929	0.951	0.974	0.935
		0.15	1	0	0.690	0.715	0.940	0.964	0.952	0.960	0.984	0.930	0.952	0.975	0.934
		-0.2	1	0	0.520	0.850	0.926	0.954	0.894	0.949	0.972	0.896	0.947	0.945	0.882
		0.15	0	1	0.690	0.715	0.936	0.967	0.953	0.949	0.977	0.938	0.947	0.978	0.936
		-0.2	0	1	0.520	0.850	0.935	0.963	0.890	0.953	0.975	0.906	0.951	0.974	0.901
	0.5	0	1	1	0.852	0.852	0.941	0.965	0.901	0.941	0.958	0.934	0.944	0.962	0.933
		-0.2	1	1	0.771	0.913	0.941	0.960	0.914	0.952	0.963	0.891	0.958	0.970	0.868
		0	1	0	0.696	0.696	0.939	0.964	0.951	0.963	0.986	0.936	0.953	0.979	0.941
		-0.4	1	0	0.495	0.851	0.907	0.933	0.864	0.950	0.969	0.889	0.949	0.948	0.912
		0	0	1	0.696	0.696	0.932	0.961	0.951	0.963	0.987	0.940	0.962	0.984	0.938
		-0.4	0	1	0.495	0.851	0.911	0.940	0.859	0.961	0.976	0.895	0.960	0.977	0.888
FI	0.3	0.2	1	1	0.783	0.924	0.937	0.708	0.874	0.948	0.705	0.906	0.946	0.717	0.923
		0.4	1	1	0.855	0.864	0.939	0.729	0.891	0.947	0.723	0.917	0.953	0.744	0.935
		0.15	1	0	0.690	0.715	0.942	0.823	0.928	0.943	0.825	0.938	0.950	0.830	0.946
		-0.2	1	0	0.520	0.850	0.946	0.819	0.912	0.942	0.818	0.925	0.949	0.831	0.931
		0.15	0	1	0.690	0.715	0.940	0.814	0.830	0.946	0.817	0.942	0.948	0.828	0.944
		-0.2	0	1	0.520	0.850	0.939	0.817	0.911	0.946	0.820	0.928	0.945	0.823	0.932
	0.5	0	1	1	0.852	0.852	0.930	0.628	0.857	0.959	0.632	0.903	0.952	0.638	0.915
		-0.2	1	1	0.771	0.913	0.931	0.610	0.856	0.947	0.615	0.896	0.949	0.617	0.916
		0	1	0	0.696	0.696	0.936	0.774	0.924	0.948	0.778	0.935	0.946	0.786	0.948
		-0.4	1	0	0.495	0.851	0.941	0.776	0.907	0.946	0.787	0.933	0.942	0.789	0.935
		0	0	1	0.696	0.696	0.939	0.782	0.921	0.942	0.787	0.932	0.942	0.782	0.939
		-0.4	0	1	0.495	0.851	0.940	0.788	0.919	0.938	0.790	0.931	0.948	0.787	0.940
MSI	0.3	0.2	1	1	0.783	0.924	0.931	0.848	0.883	0.948	0.828	0.900	0.948	0.846	0.916
		0.4	1	1	0.855	0.864	0.947	0.847	0.902	0.950	0.840	0.923	0.953	0.863	0.943
		0.15	1	0	0.690	0.715	0.944	0.875	0.931	0.947	0.875	0.941	0.948	0.877	0.946
		-0.2	1	0	0.520	0.850	0.941	0.862	0.913	0.947	0.864	0.928	0.950	0.875	0.929
		0.15	0	1	0.690	0.715	0.943	0.869	0.933	0.947	0.880	0.941	0.952	0.878	0.945
		-0.2	0	1	0.520	0.850	0.945	0.870	0.915	0.949	0.867	0.929	0.942	0.865	0.931
	0.5	0	1	1	0.852	0.852	0.941	0.713	0.873	0.940	0.709	0.914	0.942	0.724	0.921
		-0.2	1	1	0.771	0.913	0.931	0.699	0.859	0.939	0.687	0.903	0.953	0.703	0.922
		0	1	0	0.696	0.696	0.942	0.822	0.926	0.945	0.820	0.935	0.946	0.829	0.946
		-0.4	1	0	0.495	0.851	0.938	0.811	0.907	0.940	0.819	0.929	0.943	0.823	0.937
		0	0	1	0.696	0.696	0.943	0.827	0.926	0.944	0.823	0.935	0.949	0.821	0.944
		-0.4	0	1	0.495	0.851	0.940	0.815	0.916	0.941	0.823	0.931	0.948	0.823	0.937
SPE	0.3	0.2	1	1	0.783	0.924	0.923	0.922	0.974	0.950	0.950	0.969	0.945	0.946	0.974
		0.4	1	1	0.855	0.864	0.939	0.939	0.973	0.940	0.940	0.972	0.948	0.948	0.977
		0.15	1	0	0.690	0.715	0.937	0.937	0.956	0.945	0.945	0.950	0.947	0.947	0.956
		-0.2	1	0	0.520	0.850	0.924	0.924	0.965	0.947	0.947	0.952	0.940	0.940	0.953
		0.15	0	1	0.690	0.715	0.941	0.941	0.960	0.950	0.950	0.954	0.952	0.952	0.953
		-0.2	0	1	0.520	0.850	0.941	0.941	0.958	0.945	0.945	0.951	0.945	0.945	0.945
	0.5	0	1	1	0.852	0.852	0.940	0.940	0.909	0.946	0.946	0.910	0.946	0.946	0.909
		-0.2	1	1	0.771	0.913	0.933	0.934	0.932	0.955	0.955	0.904	0.944	0.945	0.917
		0	1	0	0.696	0.696	0.940	0.940	0.939	0.948	0.948	0.905	0.946	0.946	0.916
		-0.4	1	0	0.495	0.851	0.941	0.941	0.954	0.949	0.949	0.922	0.941	0.941	0.918
		0	0	1	0.696	0.696	0.946	0.945	0.959	0.941	0.941	0.915	0.943	0.944	0.916
		-0.4	0	1	0.495	0.851	0.942	0.942	0.935	0.950	0.950	0.915	0.947	0.947	0.925

Table 5.3 Misspecified disease models: Coverage probabilities of SPE-based joint empirical likelihood confidence regions with nominal confidence levels 90% and 95% in the presence of verification bias. Pre. means disease prevalence; EL(F) stands for the proposed method based on weighted chi-squared method; EL(R) stands for the reduced method from Corollary 2; NA means the normality approximation method.

					$n = 300$			$n = 400$			$n = 500$		
Level	Pre.	τ_0	η_0	θ_0	EL(F)	EL(R)	NA	EL(F)	EL(R)	NA	EL(F)	EL(R)	NA
90%	0.3	0.0	0.620	0.779	0.892	0.914	0.832	0.893	0.915	0.823	0.901	0.923	0.811
		-0.2	0.520	0.850	0.892	0.906	0.854	0.892	0.915	0.961	0.894	0.914	0.827
	0.5	-0.2	0.599	0.781	0.892	0.913	0.944	0.890	0.910	0.933	0.890	0.910	0.913
		-0.4	0.495	0.851	0.887	0.907	0.950	0.893	0.915	0.868	0.893	0.913	0.858
95%	0.3	0.0	0.620	0.779	0.944	0.959	0.895	0.949	0.963	0.889	0.949	0.961	0.879
		-0.2	0.520	0.850	0.940	0.956	0.916	0.948	0.959	0.917	0.945	0.956	0.886
	0.5	-0.2	0.599	0.781	0.941	0.954	0.961	0.943	0.957	0.956	0.943	0.955	0.949
		-0.4	0.495	0.851	0.936	0.953	0.966	0.948	0.959	0.914	0.946	0.958	0.910

Table 5.4 Misspecified verification models: Coverage probabilities of SPE-based joint empirical likelihood confidence regions with nominal confidence levels 90% and 95% in the presence of verification bias. Pre. means disease prevalence; EL(F) stands for the proposed method based on weighted chi-squared method; EL(R) stands for the reduced method from Corollary 1; NA means the normality approximation method.

							$n = 300$			$n = 400$			$n = 500$		
Level	Pre.	τ_0	ν_1	κ_1	η_0	θ_0	EL(F)	EL(R)	NA	EL(F)	EL(R)	NA	EL(F)	EL(R)	NA
90%	0.3	0.2	1	1	0.783	0.924	0.888	0.839	0.939	0.904	0.854	0.888	0.909	0.860	0.882
		0.4	1	1	0.855	0.864	0.889	0.828	0.921	0.896	0.822	0.901	0.891	0.815	0.874
		0.15	1	0	0.690	0.715	0.891	0.927	0.968	0.904	0.939	0.963	0.890	0.921	0.966
		-0.2	1	0	0.520	0.850	0.889	0.920	0.983	0.885	0.917	0.962	0.888	0.917	0.983
		0.15	0	1	0.690	0.715	0.882	0.917	0.967	0.890	0.922	0.966	0.907	0.930	0.942
		-0.2	0	1	0.520	0.850	0.888	0.924	0.975	0.891	0.917	0.967	0.900	0.920	0.953
	0.5	0	1	1	0.852	0.852	0.890	0.837	0.866	0.893	0.825	0.858	0.896	0.832	0.845
		-0.2	1	1	0.771	0.913	0.908	0.875	0.887	0.898	0.849	0.870	0.895	0.854	0.860
		0	1	0	0.696	0.696	0.883	0.918	0.963	0.890	0.928	0.954	0.897	0.933	0.958
		-0.4	1	0	0.495	0.851	0.888	0.914	0.966	0.898	0.937	0.957	0.891	0.928	0.945
		0	0	1	0.696	0.696	0.882	0.926	0.965	0.890	0.926	0.958	0.895	0.938	0.960
		-0.4	0	1	0.495	0.851	0.881	0.913	0.962	0.903	0.942	0.956	0.895	0.941	0.958
95%	0.3	0.2	1	1	0.783	0.924	0.943	0.908	0.969	0.953	0.912	0.945	0.961	0.921	0.933
		0.4	1	1	0.855	0.864	0.945	0.894	0.960	0.956	0.894	0.948	0.946	0.890	0.930
		0.15	1	0	0.690	0.715	0.958	0.977	0.975	0.964	0.979	0.974	0.947	0.963	0.977
		-0.2	1	0	0.520	0.850	0.948	0.965	0.986	0.946	0.967	0.970	0.944	0.962	0.987
		0.15	0	1	0.690	0.715	0.943	0.967	0.975	0.952	0.972	0.974	0.957	0.970	0.962
		-0.2	0	1	0.520	0.850	0.949	0.968	0.981	0.947	0.965	0.973	0.955	0.971	0.961
	0.5	0	1	1	0.852	0.852	0.957	0.913	0.923	0.949	0.901	0.923	0.955	0.904	0.910
		-0.2	1	1	0.771	0.913	0.960	0.935	0.939	0.955	0.925	0.924	0.948	0.915	0.922
		0	1	0	0.696	0.696	0.942	0.964	0.973	0.949	0.975	0.973	0.953	0.976	0.972
		-0.4	1	0	0.495	0.851	0.944	0.955	0.977	0.954	0.976	0.971	0.950	0.970	0.959
		0	0	1	0.696	0.696	0.942	0.965	0.976	0.954	0.974	0.974	0.952	0.973	0.973
		-0.4	0	1	0.495	0.851	0.944	0.966	0.972	0.961	0.977	0.972	0.959	0.978	0.966

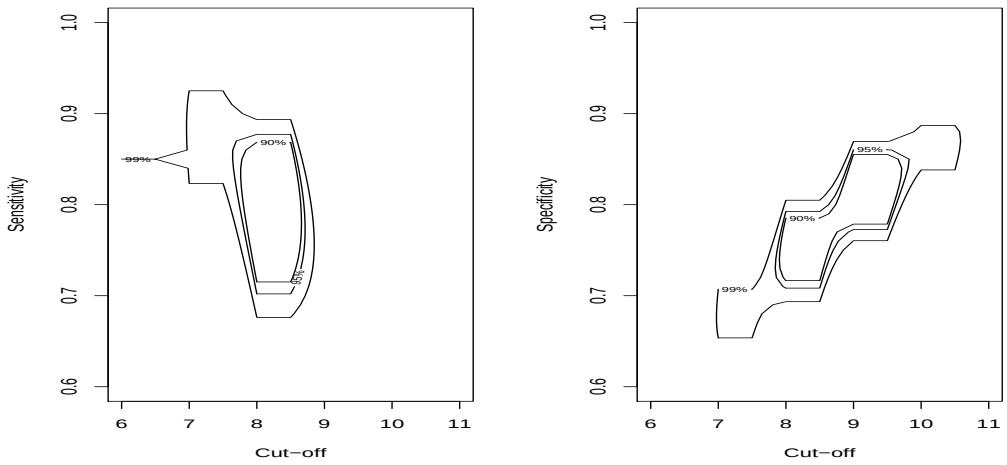

Figure 5.1 Left panel: contour curves of $\widehat{l}_{SPE}(\theta, 0.75, \tau)$, giving confidence regions for the pair (cut-off level, sensitivity) at the fixed specificity level of 0.75. Right panel: contour curves of $\widehat{l}_{SPE}(0.75, \eta, \tau)$, giving confidence regions for the pair (cut-off level, specificity) at the fixed sensitivity level of 0.75. Contours in both panels correspond to nominal confidence levels 90%, 95% and 99%.

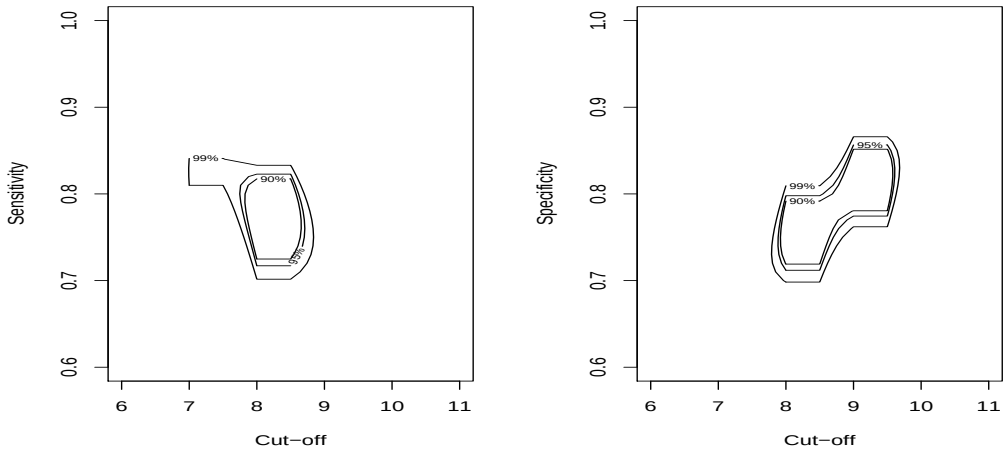

Figure 5.2 Left panel: contour curves of $\widehat{l}_{MSI}(\theta, 0.75, \tau)$, giving confidence regions for the pair (cut-off level, sensitivity) at the fixed specificity level of 0.75. Right panel: contour curves of $\widehat{l}_{MSI}(0.75, \eta, \tau)$, giving confidence regions for the pair (cut-off level, specificity) at the fixed sensitivity level of 0.75. Contours in both panels correspond to nominal confidence levels 90%, 95% and 99%.

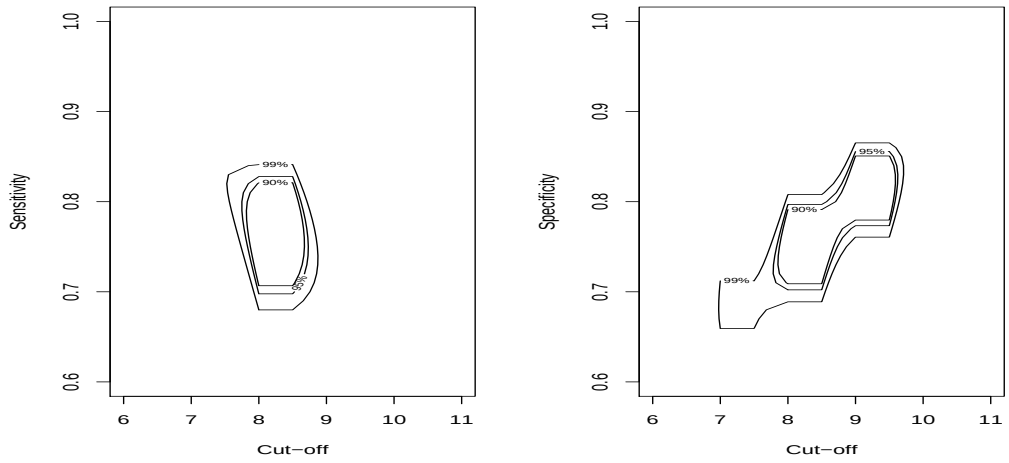

Figure 5.3 Left panel: contour curves of $\widehat{l}(\theta, 0.75, \tau)$ from full data, giving confidence regions for the pair (cut-off level, sensitivity) at the fixed specificity level of 0.75. Right panel: contour curves of $\widehat{l}(0.75, \eta, \tau)$ from full data, giving confidence regions for the pair (cut-off level, specificity) at the fixed sensitivity level of 0.75. Contours in both panels correspond to nominal confidence levels 90%, 95% and 99%.

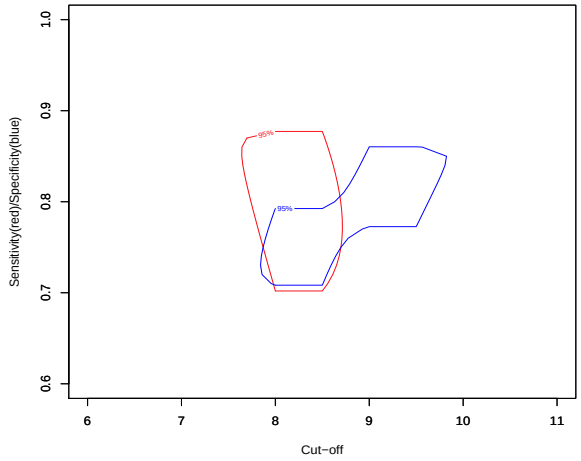

Figure 5.4 In red, the 95% SPE-based confidence region for the pair (cut-off level, sensitivity) at the fixed 0.75 level of specificity; In blue, the 95% SPE-based confidence region for the pair (cut-off level, specificity,) at the fixed 0.75 level of sensitivity.

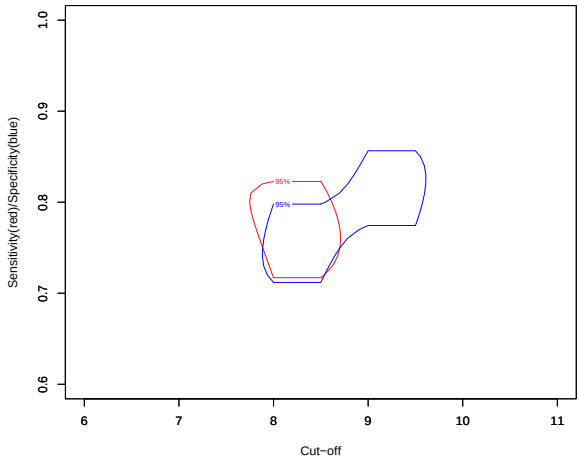

Figure 5.5 In red, the 95% MSI-based confidence region for the pair (cut-off level, sensitivity) at the fixed 0.75 level of specificity; In blue, the 95% MSI-based confidence region for the pair (cut-off level, specificity,) at the fixed 0.75 level of sensitivity.

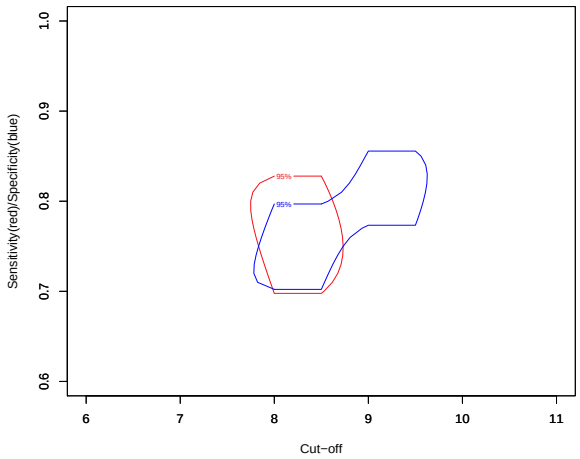

Figure 5.6 In red, the 95% confidence region for the pair (cut-off level, sensitivity) at the fixed 0.75 level of specificity with full data; In blue, the 95% confidence region for the pair (cut-off level, specificity,) at the fixed 0.75 level of sensitivity with full data.

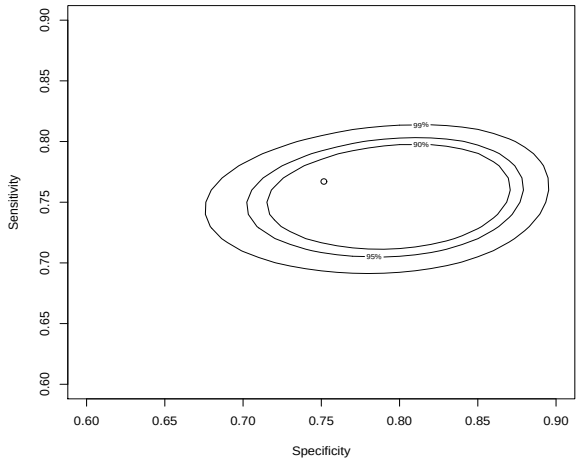

Figure 5.7 Contour curves of $\widehat{l}_{SPE}(\theta, \eta, 8)$, offering the confidence regions for the pair (specificity, sensitivity), at different nominal coverage levels 90%, 95% and 99%.

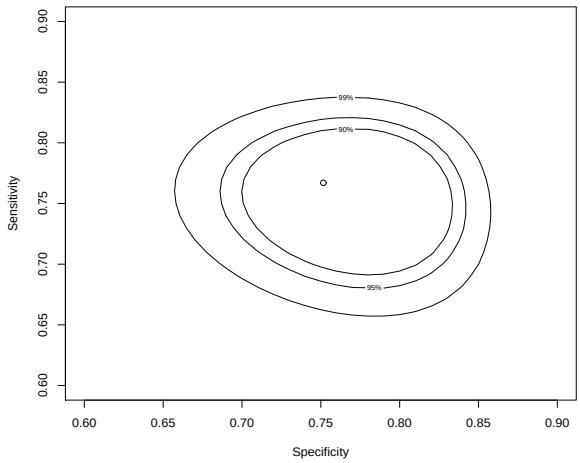

Figure 5.8 Contour curves of $\widehat{l}_{MSI}(\theta, \eta, 8)$, offering the confidence regions for the pair (specificity, sensitivity), at different nominal coverage levels 90%, 95% and 99%.

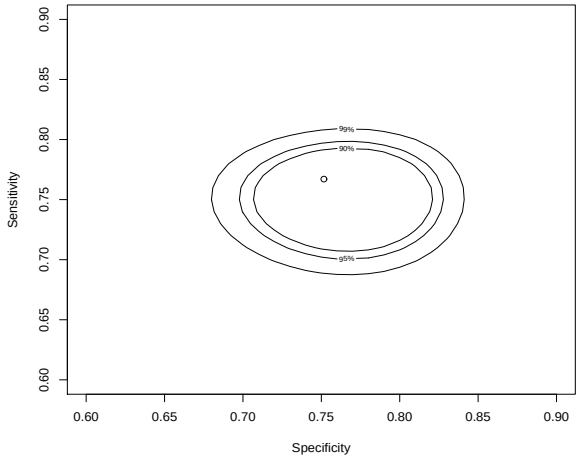

Figure 5.9 Contour curves of $\widehat{l}(\theta, \eta, 8)$ from full data, offering the confidence regions for the pair (specificity, sensitivity), at different nominal coverage levels 90%, 95% and 99%.

CHAPTER 6

JACKKNIFE EMPIRICAL LIKELIHOOD CONFIDENCE REGIONS FOR THE EVALUATION OF CONTINUOUS SCALE DIAGNOSTIC TEST WITH VERIFICATION BIAS

6.1 Introduction

In Chapter 5, motivated by Adimari and Chiogna [4], we developed various bias-corrected joint empirical likelihood confidence regions for continuous-scale tests. Their methods are based on the inverse probability weighting (IPW) estimator, the full imputation (FI) method, the mean score imputation (MSI) method, and the semi-parametric efficient estimator (SPE) for the sensitivity and the specificity [24]. We constructed confidence regions for the pairs of (sensitivity, cut-off level), (specificity, cut-off level), and (sensitivity, specificity) based on limiting weighted chi-squared distributions of empirical log-likelihood ratio statistics for the sensitivity and the specificity as well as the cut-off level. We also point out the necessity of constructing joint confidence regions. Our EL-based joint confidence regions provide visual tools to choose cut-off levels for desirable sensitivity and specificity by drawing contours of joint confidence regions for (θ, τ) and (η, τ) in the same graph. Such visual tools are easy to implement in practice. Additionally, those proposed confidence regions inherit merits from the empirical likelihood method, such as good small sample performance, data determined confidence regions and range-respecting.

The jackknife EL method, proposed by Jing *et al.* [41], is a powerful EL-based method to overcome the computational difficulty for dealing with nonlinear functionals with a particular application to U-statistics. Li *et al.* [45] proposed a jackknife EL method to construct confidence regions for parameters of interest in the presence of nuisance parameters being simply replaced by some estimators under the general estimating equation framework. With jackknife pseudo samples, the resulting jackknife EL method retains the attractive property

of standard chi-squared limiting distributions. In order to reduce the computation in the jackknife empirical likelihood method when explicit estimators of nuisance parameters are not available, Peng [46] proposed an approximate jackknife empirical likelihood method.

In this chapter, we would apply jackknife empirical likelihood methods proposed in [45] to the framework proposed in the precious chapter to construct jackknife empirical likelihood confidence regions for the evaluation of continuous-scale diagnostic tests in the presence of verification bias. Since the proposed jackknife empirical log-likelihood ratio statistics have standard chi-squared distributions as their limiting distributions, it is easy to do joint inferences for the sensitivity and the specificity as well as the cut-off level in practice. Most previous works, like Alonzo and Pepe [24], require estimates of complicate variance-covariance matrices based on normal approximation theory in order to do inferences on joint confidence regions. If explicit formulas are not available, bootstrap methods are required at a price of computation burden. Also, Alonzo and Pepe [24] used a large sample size, $n = 5000$, in their simulation studies, which are often unavailable in practice. Our methods persist the attractive property of the empirical likelihood method, which has standard chi-squared distribution as the asymptotic distribution of the EL ratio statistic and is free of the estimation of any variance-covariance matrix. Additionally, our methods work well in moderate sample size $n \geq 300$ shown in simulation studies, and small sample size in practice means a great save of money and other resources. These are contributions of our work.

We organize this chapter as follows. In Section 6.2, we apply the jackknife empirical likelihood method to construct joint empirical likelihood confidence regions in the presence of verification bias under estimating equation framework. Section 6.3 presents some simulation studies to evaluate the finite sample performance and robustness of proposed methods. Section 6.4 presents a real data analysis with one data set.

6.2 Jackknife Empirical Likelihood Confidence Regions in the Presence of Verification Bias

Some standard notations are necessary to derive following results. We use T_i to denote the continuous-scale test result from a screening test, and D_i to denote the dichotomous indicator of disease status without measurement error, for $i = 1, \ldots, n$, where $D_i = 1$ means the ith patient is diseased and $D_i = 0$ means the ith patient is free of disease. Due to many reasons, such as budget limits and privacy security, only a part of patients decide to have their disease statuses verified by gold standard tests. Let V_i denote the dichotomous indicator of the verification status of the ith patient, with $V_i = 1$ if the ith patient has his or her true disease status verified, and $V_i = 0$ if otherwise. Usually, during diagnostic process, some covariates, like demographic variables, other than results from the screening test, are collected. By incorporating this information, it is possible to model the verification mechanism and the disease status, and we are interested in such kind of cases. Let A_i denote a vector of observed covariates for the ith patient that may be associated with both D_i and V_i.

Similar with Chapter 5, under the missing at random (MAR) assumption, we assume that the verification of disease status is conditionally independent of the true disease status given test results and observed covariates, i.e., $V \perp D|T, A$. That is to say, whether or not a patient has his or her true disease status verified only depends on T and A regardless of the true disease status D, part of which are missing. Throughout this chapter, results are based on this assumption.

With an i.i.d. sample $S_i = (T_i, A_i, V_i, D_i), i = 1, \ldots, n$, conditionally independent V_i and D_i given T_i and A_i, motivated by the IPW, FI, MSI and SPE method-based bias-corrected estimators of the sensitivity and the specificity provided in [24], Chapter 5 provides following

four pairs of estimating functions for $\vartheta = (\theta, \eta, \tau)'$:

$$\text{IPW:} \quad \begin{cases} g_{\text{IPW},1}(S_i, \vartheta, \pi_i) = \pi_i^{-1} V_i D_i (I(T_i \le \tau) - (1-\theta)) \\ g_{\text{IPW},2}(S_i, \vartheta, \pi_i) = \pi_i^{-1} V_i (1-D_i)(I(T_i \le \tau) - \eta) \end{cases}$$

$$\text{FI:} \quad \begin{cases} g_{\text{FI},1}(S_i, \vartheta, \rho_i) = \rho_i (I(T_i \le \tau) - (1-\theta)) \\ g_{\text{FI},2}(S_i, \vartheta, \rho_i) = (1-\rho_i)(I(T_i \le \tau) - \eta) \end{cases}$$

$$\text{MSI:} \quad \begin{cases} g_{\text{MSI},1}(S_i, \vartheta, \rho_i) = (V_i D_i + (1-V_i)\rho_i)(I(T_i \le \tau) - (1-\theta)) \\ g_{\text{MSI},2}(S_i, \vartheta, \rho_i) = (V_i(1-D_i) + (1-V_i)(1-\rho_i))(I(T_i \le \tau) - \eta) \end{cases}$$

$$\text{SPE:} \quad \begin{cases} g_{\text{SPE},1}(S_i, \vartheta, \pi_i, \rho_i) = (V_i D_i/\pi_i - (V_i - \pi_i)\rho_i/\pi_i)(I(T_i \le \tau) - (1-\theta)) \\ g_{\text{SPE},2}(S_i, \vartheta, \pi_i, \rho_i) = (V_i(1-D_i)/\pi_i - (V_i - \pi_i)(1-\rho_i)/\pi_i)(I(T_i \le \tau) - \eta) \end{cases}$$

where $\pi_i = P(V_i = 1|T_i, A_i)$, and $\rho_i = P(D_i = 1|T_i, A_i)$. Let

$$g_{\text{IPW}}(S_i, \vartheta, \pi_i) = (g_{\text{IPW},1}(S_i, \vartheta, \pi_i), g_{\text{IPW},2}(S_i, \vartheta, \pi_i))', \tag{6.1}$$

$$g_{\text{FI}}(S_i, \vartheta, \rho_i) = (g_{\text{FI},1}(S_i, \vartheta, \rho_i), g_{\text{FI},2}(S_i, \vartheta, \rho_i))', \tag{6.2}$$

$$g_{\text{MSI}}(S_i, \vartheta, \rho_i) = (g_{\text{MSI},1}(S_i, \vartheta, \rho_i), g_{\text{MSI},2}(S_i, \vartheta, \rho_i))', \tag{6.3}$$

$$g_{\text{SPE}}(S_i, \vartheta, \pi_i, \rho_i) = (g_{\text{SPE},1}(S_i, \vartheta, \pi_i, \rho_i), g_{\text{SPE},2}(S_i, \vartheta, \pi_i, \rho_i))'. \tag{6.4}$$

π_i's and ρ_i's are sometimes unknown in practice. But one can replace them by their consistent estimators from parametric models on observed information, i.e., $\rho_i = \rho(S_i, \boldsymbol{\alpha})$ and $\pi_i = \pi(S_i, \boldsymbol{\beta})$. Usually, logistic regressions and probit models are employed to model binary outcomes. For illustration, probit models are used to model ρ_i's and logistic regressions are used to model π_i's by incorporating covariates. Let $Z_i = (1, T_i, A_i)', \boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)', \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$, and $\Phi(\cdot)$ is the standard normal distribution function. Then, $\rho_i = \Phi(Z_i'\boldsymbol{\alpha})$ and $\pi_i = \frac{1}{1+e^{-Z_i'\boldsymbol{\beta}}}$. In the estimating equation framework, for probit models, we have:

$$Q_{2n,1}(\boldsymbol{\alpha}) \equiv \frac{1}{n}\sum_{i=1}^{n} V_i Q_{2,1}(S_i, \boldsymbol{\alpha}) \equiv \frac{1}{n}\sum_{i=1}^{n} \frac{V_i[D_i - \Phi(Z_i'\boldsymbol{\alpha})]\phi(Z_i'\boldsymbol{\alpha})}{\Phi(Z_i'\boldsymbol{\alpha})(1-\Phi(Z_i'\boldsymbol{\alpha}))} Z_i, \tag{6.5}$$

where exact disease statuses are only available for verified subjects with $V_i = 1$ and $\phi(\cdot)$ is the standard normal density function. For logistic regressions, we have:

$$Q_{2n,2}(\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{i=1}^{n} Q_{2,2}(S_i, \boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{i=1}^{n}(V_i - \pi_i)Z_i. \tag{6.6}$$

If above models are correctly specified, we can obtain consistent estimates $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$ and $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$. Then profiled estimating equations are given by plugging in resulting consistent estimates $\widehat{\rho}_i = \rho(Z_i, \widehat{\boldsymbol{\alpha}})$ of ρ_i and $\widehat{\pi}_i = \pi(Z_i, \widehat{\boldsymbol{\beta}})$ of π_i into (6.1)-(6.4). These plug-in estimating equations are easy to obtain, but they are not independent anymore. Therefore, Chapter 5 defines a plug-in EL for $\vartheta = (\theta, \eta, \tau)'$, and proves that the corresponding profile empirical likelihood ratio statistic has a weighted chi-squared distributions as its limiting distribution. But the weights are required to be estimated before constructing confidence regions.

Motivated by [45], the jackknife technique could be employed to obtain asymptotically independent pseudo values. Let $\widehat{\boldsymbol{\alpha}}_{-i}$ denote the solution to the equations

$$Q_{2n,-i,1}(\boldsymbol{\alpha}) \equiv \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} \frac{V_j[D_j - \Phi(Z_j'\boldsymbol{\alpha})]\phi(Z_j'\boldsymbol{\alpha})}{\Phi(Z_j'\boldsymbol{\alpha})(1-\Phi(Z_j'\boldsymbol{\alpha}))} Z_j, \tag{6.7}$$

and $\widehat{\boldsymbol{\beta}}_{-i}$ denote the solution to the equations

$$Q_{2n,-i,2}(\boldsymbol{\beta}) \equiv \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} (V_j - \pi_j)Z_j. \tag{6.8}$$

Set

$$
\begin{aligned}
T_{n,\mathrm{IPW}}(\vartheta) &= \sum_{i=1}^{n} g_{\mathrm{IPW}}(S_i, \vartheta, \widehat{\pi}_i), \\
T_{n,\mathrm{FI}}(\vartheta) &= \sum_{i=1}^{n} g_{\mathrm{FI}}(S_i, \vartheta, \widehat{\rho}_i), \\
T_{n,\mathrm{MSI}}(\vartheta) &= \sum_{i=1}^{n} g_{\mathrm{MSI}}(S_i, \vartheta, \widehat{\rho}_i), \\
T_{n,\mathrm{SPE}}(\vartheta) &= \sum_{i=1}^{n} g_{\mathrm{SPE}}(S_i, \vartheta, \widehat{\pi}_i, \widehat{\rho}_i).
\end{aligned}
$$

And let $\widehat{\rho}_{j,-i} = \Phi(Z_j'\widehat{\boldsymbol{\alpha}}_{-i})$ and $\widehat{\pi}_{j,-i} = \frac{1}{1+e^{-Z_j'\widehat{\boldsymbol{\beta}}_{-i}}}$, if $j \neq i$. Similarly, define

$$
\begin{aligned}
T_{n,-i,\mathrm{IPW}}(\vartheta) &= \sum_{j=1, j\neq i}^{n} g_{\mathrm{IPW}}(S_j, \vartheta, \widehat{\pi}_{j,-i}), \\
T_{n,-i,\mathrm{FI}}(\vartheta) &= \sum_{j=1, j\neq i}^{n} g_{\mathrm{FI}}(S_j, \vartheta, \widehat{\rho}_{j,-i}), \\
T_{n,-i,\mathrm{MSI}}(\vartheta) &= \sum_{j=1, j\neq i}^{n} g_{\mathrm{MSI}}(S_j, \vartheta, \widehat{\rho}_{j,-i}), \\
T_{n,-i,\mathrm{SPE}}(\vartheta) &= \sum_{j=1, j\neq i}^{n} g_{\mathrm{SPE}}(S_j, \vartheta, \widehat{\pi}_{j,-i}, \widehat{\rho}_{j,-i}).
\end{aligned}
$$

Therefore, jackknife pseudo samples are defined as

$$
Y_{i,\mathrm{IPW}}(\vartheta) = nT_{n,\mathrm{IPW}}(\vartheta) - (n-1)T_{n,-i,\mathrm{IPW}}(\vartheta) \tag{6.9}
$$

$$
Y_{i,\mathrm{FI}}(\vartheta) = nT_{n,\mathrm{FI}}(\vartheta) - (n-1)T_{n,-i,\mathrm{FI}}(\vartheta) \tag{6.10}
$$

$$
Y_{i,\mathrm{MSI}}(\vartheta) = nT_{n,\mathrm{MSI}}(\vartheta) - (n-1)T_{n,-i,\mathrm{MSI}}(\vartheta) \tag{6.11}
$$

$$
Y_{i,\mathrm{SPE}}(\vartheta) = nT_{n,\mathrm{SPE}}(\vartheta) - (n-1)T_{n,-i,\mathrm{SPE}}(\vartheta). \tag{6.12}
$$

With similar argument in Tukey [72], the above $Y_{i,\mathrm{IPW}}(\vartheta)$'s, $Y_{i,\mathrm{FI}}(\vartheta)$'s, $Y_{i,\mathrm{MSI}}(\vartheta)$'s and $Y_{i,\mathrm{SPE}}(\vartheta)$'s, $i = 1, \ldots, n$, are expected to be asymptotically independent, respectively. Then standard empirical likelihood methods could be applied to above jackknife samples for con-

structing joint empirical likelihood confidence regions for ϑ. We define the jackknife EL function as follows:

$$L^J(\vartheta) = \sup\left\{\prod_{i=1}^{n} p_i : \text{each } p_i > 0, \sum_{i=1}^{n} p_i = 1, \sum_{j=1}^{n} p_i Y_i(\vartheta) = 0\right\}, \tag{6.13}$$

where $Y_i(\vartheta)$ could be any one of $Y_{i,\text{IPW}}(\vartheta)$, $Y_{i,\text{FI}}(\vartheta)$, $Y_{i,\text{MSI}}(\vartheta)$ and $Y_{i,\text{SPE}}(\vartheta)$. With the Lagrange multiplier method, the jackknife log empirical likelihood ratio for ϑ is given by

$$l^J(\vartheta) = 2\sum_{i=1}^{n} \log(1 + \widehat{t}' Y_i(\vartheta)), \tag{6.14}$$

where $\widehat{t}$ is the solution of the following equation:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{Y_i(\vartheta)}{1 + t' Y_i(\vartheta)} = 0.$$

By applying Theorem 1 in [45] to these special cases with all conditions there satisfied, we obtain the following theorem.

Theorem 9 *Assume that ϑ_0 is the true value of ϑ. Then,*

$$l^J(\vartheta_0) \xrightarrow{d} \chi^2_2, \tag{6.15}$$

where χ^2_2 is a chi-squared random variable with two degrees of freedom.

Remark 1: Theorem 9 provides $l^J(\vartheta) = l^J(\theta, \eta, \tau)$ as a very good asymptotic pivotal for the inference of parameter pairs (θ_0, τ_0), (η_0, τ_0) or (θ_0, η_0) with verification bias, when the third remaining parameter is fixed at a given value. It is clear that based on Theorem 9, we do not need to estimate any complicate variance-covariance matrices, which are necessary for the normal approximation-based inferences. These complicate expressions of such kind of matrices limit the application of normal approximation-based methods, and bootstrap methods are required to alleviate these problems at prices of computation burden and time consumption. Alonzo and Pepe [24] sketched the proof of asymptotic results in the framework

of general estimating equations, and explicit formulas are complicate. Therefore, they used bootstrap resampling in simulation studies.

Remark 2: Theorem 9 provides four types of bias-corrected joint confidence regions, which are flexible to accommodate many situations. One could use one or several of them according to the specific situation. Also, simulation studies show that our methods work well in moderate sample size, $n \geq 300$, compared with $n = 5000$ in [24]. Good performance in smaller sample size situations implies a great save of money and other resources, often crucial in practical applications.

Based on above remarks, we could say our methods are best ones so far in constructing joint confidence regions in the presence of verification bias under the MAR assumption.

The proof of Theorem 1 directly follows from the theory in [45]. Only two points should be pointed out here. In our case, estimating equations $Q_{2n,1}(\boldsymbol{\alpha})$ for $\boldsymbol{\alpha}$ and $Q_{2n,2}(\boldsymbol{\beta})$ for $\boldsymbol{\beta}$ do not involve ϑ. Thus their estimates are not functions of ϑ. Additionally, the number of total parameters equal the number of estimating functions in our case, thus as mentioned in [45], $l^J(\widehat{\vartheta}) = 0$.

In practice, the verification mechanism is known sometimes. If this is true, it is unnecessary to apply IPW-based jackknife empirical likelihood confidence regions for ϑ. One can use the reduced empirical likelihood confidence regions for ϑ provided in Chapter 5 because the reduced empirical log-likelihood ratio statistic for ϑ is still a standard chi-square distribution with 2 degrees of freedom.

In this chapter, we consider three types of $(1-\alpha)100\%$ jackknife empirical likelihood confidence regions as follows:

- $\mathscr{R}_{\alpha,1}(\theta,\tau) = \{(\theta,\tau) : l^J(\theta,\eta_0,\tau) \leq \chi^2_{\alpha,2}\}$;
- $\mathscr{R}_{\alpha,2}(\eta,\tau) = \{(\eta,\tau) : l^J(\theta_0,\eta,\tau) \leq \chi^2_{\alpha,2}\}$;
- $\mathscr{R}_{\alpha,3}(\theta,\eta) = \{(\theta,\eta) : l^J(\theta,\eta,\tau_0) \leq \chi^2_{\alpha,2}\}$;

where $\alpha \in (0,1)$, and $\chi^2_{\alpha,2}$ is the $(1-\alpha)$-th quantile of the chi-square distribution with 2 degrees of freedom. Compared with the methods proposed in Chapter 5, we do not need to

estimate quantiles of weighted chi-square distribution. Methods proposed in this paper merit a lot from the application of the jackknife technique. Thus, they bring much convenience in practice.

The above three types of confidence regions will be utilized to select a reasonable cut-off level for a continuous-scale screening test. With the selected cut-off level, joint confidence regions of the sensitivity and the specificity can be constructed. The whole procedure is similar to that in Chapter 5. In this chapter, it will be illustrated by another real data set.

6.3 Simulation Studies

In this section, simulation studies are conducted to evaluate the finite sample performance and robustness of proposed various bias-corrected jackknife empirical likelihood confidence regions.

Same model settings with those in Chapter 5 are utilized here for the purpose of comparison. Firstly, two independent underlying continuous disease processes are generated, denoted by $Z_1 \sim \mathcal{N}(0, 0.5)$ and $Z_2 \sim \mathcal{N}(0, 0.5)$. The disease status indicator D is generated as a binary variable: if a random variable $Z = Z_1 + Z_2 \sim \mathcal{N}(0, 1)$ exceeds a certain threshold h, then $D = 1$, indicating the patient is diseased; otherwise, $D = 0$. Thus h determines the disease prevalence. Continuous screening test results T and the auxiliary covariates A are generated through Z_1 and Z_2: $T = \nu_1 Z_1 + \kappa_1 Z_2 + \varepsilon_1$ and $A = \nu_2 Z_1 + \kappa_2 Z_2 + \varepsilon_2$, where $\varepsilon_1 \sim \mathcal{N}(0, 0.25)$ and $\varepsilon_2 \sim \mathcal{N}(0, 0.25)$ are independent. It is clear that ν_1, ν_2, κ_1 and κ_2 determine the strength of correlation between T and A. Additionally, T and A are related to D. For detailed explanation, one may refer to [24]. Under this model setting, it can be shown that D conditional on T and A follows a probit model.

6.3.1 Correct Models

Under the MAR assumption, the verification probability $\pi(Z, \boldsymbol{\beta})$ is specified as a function of $Z = (1, T, A)'$, and the parameter $\boldsymbol{\beta}$ could be estimated from an estimating equation regardless of ϑ of major interest. In this section, we set $\log(\frac{\pi}{1-\pi}) = -0.7 + T + A$ with

$\pi = P(V = 1|T, A)$. In the presence of verification bias, D is only available for those patients with $V = 1$. Therefore, disease status results are available for roughly 40% of patients. More exactly, 20%-30% of non-diseased patients have their disease status verified, compared with roughly 70%-80% for diseased patients under model settings. In order to apply FI, MSI and SPE methods, a parametric model for probabilities, ρ_i's, are required to be specified. It was shown in [24] that a probit model that was linear in T and A was a true model under above settings.

4000 random samples are drawn from underlying distributions with sample sizes $n = 200, 300, 400$ and 500 respectively to evaluate the performance of proposed various jackknife empirical likelihood confidence regions in terms of coverage probability at nominal levels 90% and 95%. At this moment, we fix $\nu_2 = \kappa_2 = 1$, and select h to make the prevalence of disease equals 0.3 and 0.5. Different values of (ν_1, κ_1) are selected to generate balanced and unbalanced specificity and sensitivity. For the purpose of comparison, one may refer to Chapter 5 for results of profile empirical likelihood methods and normal approximation methods. The proposed jackknife EL confidence regions clearly perform much better than normal approximation-based methods.

In Table 6.1, coverage probabilities of IPW, FI, MSI and SPE based jackknife EL confidence regions are presented with nominal levels 90% and 95% under various scenarios in the presence of verification bias. It is clear that the proposed four bias-corrected jackknife EL confidence regions generally work pretty well in moderate sample size cases ($n \geq 300$). Also, if (ν_1, κ_1) is comparable with (ν_2, κ_2) (i.e. $\nu_1 = \kappa_1 = 1$ and $\nu_2 = \kappa_2 = 1$), better sensitivity and better specificity could be achieved simultaneously with carefully selected cut-off level τ. When sample size is small, $n = 200$, all four types of confidence regions are not satisfactory in many cases, especially for unbalanced sensitivity and specificity case. This observation is reasonable because jackknife technique relies on the asymptotic independence of pseudo samples. With smaller sample sizes, pseudo samples may not be "adequately independent".

With correctly specified verification models and disease models, IPW, FI, MSI and SPE

based bias-corrected jackknife EL regions are competent with each other. But when the disease prevalence is lower, 0.3, $\tau_0 = 0.2$, and sensitivity is much greater than specificity, coverage probabilities are not stable enough with smaller sample sizes ($n \leq 400$). Our observation could still be explained by asymptotic independence of jackknife pseudo samples. According to our experience, we have several suggestions on the application of these methods. If the verification mechanism is more likely to be correctly specified, IPW-based regions are preferred. Alternatively, if the prevalence of disease is more likely to be modeled correctly, MSI-based regions are more preferred based on the following considerations: only part of D_i's for unverified patients are required to be imputed compared with the FI method, resulting in less variation. If either verification mechanism or prevalence of disease is correctly specified, SPE-based confidence regions are employed, because it is doubly robust. This means if one of these two models is misspecified, resulting estimates are still consistent ([26], [58], [57]).

When facing the problem of selecting a reasonable cut-off level for a screening test, one can use jackknife EL confidence regions $\mathscr{R}_{\alpha,1}$ and $\mathscr{R}_{\alpha,2}$ to identify a reasonable cut-off level. The entire procedure will be illustrated in the section of real case study.

Compared with proposed methods, estimates from estimating equations involve the selection of initial values, which will influence the convergence of the algorithm. Also, inference based on estimating equation estimates highly depends on accurate estimations of variance-covariance matrices that are often provided in complicated forms. Bootstrap would aid this yet at a price of computational burden.

6.3.2 Misspecified Models

So far, both disease models and verification models are assumed to be correctly specified in previous simulation studies. However, misspecification of underlying models may happen. SPE method is proposed based on robustness consideration, which could work in such situation. In the followings, we will evaluate the robustness of proposed bias-corrected jackknife EL confidence regions, especially for SPE-based confidence regions.

To introduce misspecification, we apply the setting used in [24]. V is generated from a

Bernoulli random variable with $P(V = 1) = 1$ for patients with $T > t^{(0.8)}$, i.e. all verified, and $P(V = 1) = 0.2$ for others, where $t^{(0.8)}$ is the 80-th quantile of the distribution of T. But we still model V by a logistic regression. Then a misspecification of the verification model happens. In previous settings, the disease is present if $Z_1 + Z_2 > h$, and T and A are generated from linear combinations of Z_1 and Z_2. Based on the discussion in [24], with $\nu_1 = 1$ and $\tau = 0$ for T and $\nu_2 = 0$ and $\tau_2 = 1$ for A, a probit model of D only linear in T is misspecified. As mentioned above, it is expected that bias-corrected jackknife EL confidence regions based on $l^J(\vartheta_0)$ with $Y_{i,\mathrm{SPE}(\vartheta)}$'s are still doubly robust.

When disease models are misspecified, confidence regions based on FI, MSI and SPE methods are evaluated, because the IPW method does not depend on models of disease. Coverage probabilities (which are from 70% to 80%, not reported here) for FI and MSI based confidence regions are much smaller than nominal levels, and results for SPE-based methods are displayed in Table 6.2. Table 6.2 shows that the proposed SPE-based jackknife EL confidence regions work well with moderate sample size cases ($n \geq 300$) when a misspecification of disease models is present.

When verification models are misspecified, only confidence regions based on IPW and SPE methods are evaluated, because FI and MSI methods do not depend on models of verification. Similarly, coverage probabilities for the IPW method are not good (not reported here), and only results from SPE based method are presented. Table 6.3 indicates that SPE-based jackknife EL confidence regions work well with moderate sample size cases ($n \geq 300$) for most cases when a misspecification of verification models is present. In the first case where disease prevalence is 0.3, $\nu_1 = \kappa_1 = 1$, and the cut-off level is selected to generate much higher sensitivity than specificity, resulting joint confidence regions which undercover true values. However, from simulation results, it is clear that their performance is improving as the sample size increases from 300 to 600. One reasonable explanation for such phenomena is still due to the asymptotic independence of jackknife pseudo samples. In order to get better results, we recommend the proposed methods in Chapter 5 based on weighted chi-square distributions, which are shown to perform well in all cases for moderate sample sizes

$n \geq 300$.

These observations confirm that the SPE-based jackknife EL method is "doubly robust". As mentioned in Chapter 5, SPE-based NA regions could perform well in some settings, but results are not stable.

6.4 Study of Neonatal Hearing Screening Data

We apply proposed various jackknife confidence regions in the presence of verification bias to the data set from a neonatal hearing screening data set. This data set is analyzed by Alonzo and Pepe [24]. In their paper, they applied several bias-corrected estimators of true and false positive rates to construct ROC curves and estimate areas under the ROC curve of screening tests for neonatal hearing loss.

Undetected hearing loss in infants is of great concern in practice, because it would result in serious problems with speech, social and emotional development. Thus, earlier diagnosis of such loss will lower the risk of infants. The identification of neonatal hearing impairment (INHI) study aims at assessing the accuracy of two passive electronic devices, the distorting product otoacoustic emissions (DPOAE) and the transient evoked otoacoustic emissions (TEOAE) tests. These two tests could be administered soon after birth [73], compared with the gold standard test for determining neonatal hearing loss, visual reinforcement audiometry (VRA), which cannot be administered until infants are 8 to 12 months old. Therefore, the evaluation of such two screening tests is necessary.

The subset of the INHI data used in this section was generated by Alonzo and Pepe [24], following a two-phase design. In the first phase, DPOAE and TEOAE test results are available for all infants. In the second phase, all infants with DPOAE test results greater than the 80-th quantile of the distribution of DPOAE test results at least on one ear are sent to be verified, and remaining infants are verified with probability 0.4. The subset includes TEOAE and DPOAE test results on 5101 ears, corresponding to 2763 infants. Also, verification statuses V_i's and VRA results, D_i's, for 1571 verified infants are also available. We follow the same argument in [24] to let T ≡DPOAE and A ≡TEOAE. Also logistic

regression for $P(D = 1|T, A)$ is utilized to obtain $\widehat{\rho}_i$'s. We use 2763 observations, one ear of each infant, in our real case study because the decision of verifying ears just depend on the ear with larger screening test values. Estimates of π_i could be written into a general estimating equation resulting in empirical estimates. For infants with DPOAE test results greater than the 80-th quantile of DPOAE results, $\widehat{\pi}_i = 1$; for infants with DPOAE test results below the threshold, $\widehat{\pi}_i = 0.394$. With a large data set at hand, we just simply remove the ith value of $\{\widehat{\pi}_i\}_{i=1}^n$ to obtain $T_{n,-i,\text{IPW}}(\vartheta)$ and $T_{n,-i,\text{SPE}}(\vartheta)$, because empirical estimates are stable for a large data set.

In order to select a reasonable cut-off point that generates both higher sensitivity and higher specificity of the screening test DPOAE, motivated by [68], we follow the procedure in Chapter 5. Firstly, proper joint confidence regions of the pair (sensitivity, cut-off level) at a fixed specificity value and the pair (specificity, cut-off level) at a fixed sensitivity value are constructed to investigate the relationship between the sensitivity/specificity and the cut-off level. In order to fix the specificity and the sensitivity, we check ROC curves provided in [24]. From those curves, the screening test DPOAE does not have a good performance, thus we fix both specificity and sensitivity at 0.6. With jackknife technique, joint empirical likelihood confidence regions could be directly constructed by applying Theorem 9. Contour plots are employed to show confidence regions. $l^J(\theta, 0.6, \tau)$ and $l^J(0.6, \eta, \tau)$ are evaluated at fine grids of (θ, τ) and (η, τ), and contours are connected according to the $(1 - \alpha)$th quantile of the chi-squared distribution with 2 degrees of freedom. Because the true model of verification is available, jackknife empirical likelihood confidence regions based on the IPW method are used as reference. Additionally, results from MSI-based jackknife empirical likelihood with the logistic regression assumption of $P(D = 1|T, A)$ are also presented here to make a comparison.

The left panel in Figure 6.1 shows contour curves of $l^J_{\text{IPW}}(\theta, 0.6, \tau)$, offering confidence regions for the pair (sensitivity, cut-off level) at nominal confidence levels 90%, 95% and 99%. The plot indicates that, only a narrow range around -4 of the cut-off level is compatible with the target specificity level of 0.6. The right panel in Figure 6.1 shows contour curves of

$l^J_{\text{IPW}}(0.6, \eta, \tau)$, offering confidence regions for the pair (specificity, cut-off level) at nominal levels 90%, 95% and 99%. This plot shows that, a variety of pairs of values for (η, τ) are compatible with the target sensitivity level of 0.6 in a narrow strip manner. Similarly we have Figure 6.2 for $l^J_{\text{MSI}}(\theta, 0.6, \tau)$ and $l^J_{\text{MSI}}(0.6, \eta, \tau)$. Although, both figures offer similar shapes of confidence regions at nominal levels 90% and 95%, MSI-based confidence regions of the pair of specificity and cut-off level have narrower ranges for the cut-off level. Therefore, MSI-based joint confidence regions with the logistics regression assumption for D seems to be optimistic by comparing the first two figures.

Joint confidence regions offer us a good chance to select reasonable cut-off levels compatible with both good specificity and sensitivity, 0.6 in this study, as we plot two regions in the same graph. Figure 6.3 shows both 95% IPW-based jackknife joint confidence region for the pair (sensitivity, cut-off level) at the fixed specificity level of 0.6 and the 95% confidence region for the pair (specificity, cut-off level) at the fixed sensitivity level of 0.6. Figure 6.4 shows similar regions from MSI-based jackknife joint confidence regions. The overlapping parts in two figures indicate a narrow interval around -4 of the cut-off level. Therefore, we select -4 as a reasonable value of the cut-off level.

Given the cut-off level, joint confidence regions of the specificity and the sensitivity could be obtained. Figure 6.5 and 6.6 show contour curves of $l^J_{\text{IPW}}(\theta, \eta, -4)$ and $l^J_{\text{MSI}}(\theta, \eta, -4)$, respectively, indicating 95% joint jackknife confidence regions for the pair (specificity, sensitivity) with the cut-off level fixed at -4. With the cut-off level fixed at -4, in Figure 6.5, the specificity ranges from 0.58 to 0.67, and the sensitivity has a larger variation, from 0.35 to 0.75; in Figure 6.6, the MSI-based region is optimistic, providing a smaller region, i.e., the specificity ranges from 0.60 to 0.65, and the sensitivity ranges from 0.40 to 0.73. Generally speaking, the performance of DPOAE, treated as a screening test, could not offer both high sensitivity and high specificity. Thus the gold standard VRA is required to verify neonatal hearing impairment.

Table 6.1 Correct models: Coverage probabilities of various jackknife empirical likelihood confidence regions with nominal confidence levels 90% and 95% in the presence of verification bias. Pre. means disease prevalence.

							$n = 200$		$n = 300$		$n = 400$		$n = 500$	
Type	Pre.	τ_0	ν_1	κ_1	η_0	θ_0	90%	95%	90%	95%	90%	95%	90%	95%
IPW	0.3	0.2	1	1	0.783	0.924	0.930	0.969	0.896	0.951	0.881	0.937	0.878	0.935
		0.4	1	1	0.855	0.864	0.897	0.948	0.888	0.940	0.885	0.940	0.894	0.942
		0.15	1	0	0.690	0.715	0.896	0.944	0.897	0.952	0.899	0.948	0.897	0.950
		-0.2	1	0	0.520	0.850	0.896	0.943	0.889	0.938	0.898	0.940	0.905	0.947
		0.15	0	1	0.690	0.715	0.897	0.945	0.901	0.948	0.906	0.953	0.907	0.949
		-0.2	0	1	0.520	0.850	0.897	0.946	0.887	0.941	0.900	0.944	0.890	0.944
	0.5	0	1	1	0.852	0.852	0.888	0.942	0.893	0.945	0.897	0.948	0.894	0.949
		-0.2	1	1	0.771	0.913	0.890	0.934	0.892	0.943	0.888	0.939	0.893	0.946
		0	1	0	0.696	0.696	0.901	0.945	0.900	0.946	0.904	0.949	0.911	0.956
		-0.4	1	0	0.495	0.851	0.858	0.910	0.871	0.924	0.888	0.934	0.905	0.947
		0	0	1	0.696	0.696	0.903	0.942	0.905	0.953	0.911	0.956	0.904	0.954
		-0.4	0	1	0.495	0.851	0.866	0.913	0.889	0.927	0.895	0.938	0.894	0.939
FI	0.3	0.2	1	1	0.783	0.924	0.868	0.919	0.888	0.940	0.887	0.937	0.895	0.941
		0.4	1	1	0.855	0.864	0.892	0.942	0.898	0.948	0.897	0.945	0.906	0.957
		0.15	1	0	0.690	0.715	0.901	0.953	0.889	0.949	0.902	0.953	0.908	0.960
		-0.2	1	0	0.520	0.850	0.895	0.946	0.888	0.939	0.891	0.945	0.902	0.949
		0.15	0	1	0.690	0.715	0.904	0.955	0.902	0.953	0.900	0.953	0.913	0.954
		-0.2	0	1	0.520	0.850	0.893	0.946	0.899	0.950	0.900	0.946	0.900	0.947
	0.5	0	1	1	0.852	0.852	0.877	0.924	0.902	0.947	0.901	0.949	0.910	0.955
		-0.2	1	1	0.771	0.913	0.874	0.927	0.892	0.937	0.884	0.940	0.910	0.953
		0	1	0	0.696	0.696	0.911	0.954	0.907	0.957	0.901	0.956	0.914	0.960
		-0.4	1	0	0.495	0.851	0.888	0.944	0.906	0.950	0.905	0.954	0.907	0.952
		0	0	1	0.696	0.696	0.912	0.953	0.902	0.953	0.905	0.952	0.906	0.957
		-0.4	0	1	0.495	0.851	0.897	0.944	0.908	0.952	0.899	0.949	0.907	0.958
MSI	0.3	0.2	1	1	0.783	0.924	0.880	0.920	0.888	0.939	0.891	0.943	0.898	0.947
		0.4	1	1	0.855	0.864	0.904	0.952	0.906	0.953	0.898	0.950	0.911	0.957
		0.15	1	0	0.690	0.715	0.903	0.954	0.896	0.951	0.902	0.954	0.905	0.956
		-0.2	1	0	0.520	0.850	0.896	0.944	0.895	0.942	0.897	0.948	0.902	0.947
		0.15	0	1	0.690	0.715	0.910	0.957	0.903	0.954	0.900	0.956	0.907	0.955
		-0.2	0	1	0.520	0.850	0.898	0.945	0.906	0.953	0.898	0.949	0.893	0.948
	0.5	0	1	1	0.852	0.852	0.889	0.935	0.906	0.951	0.911	0.955	0.914	0.954
		-0.2	1	1	0.771	0.913	0.875	0.922	0.900	0.943	0.893	0.947	0.917	0.958
		0	1	0	0.696	0.696	0.910	0.955	0.911	0.958	0.898	0.953	0.913	0.958
		-0.4	1	0	0.495	0.851	0.892	0.943	0.904	0.950	0.901	0.951	0.906	0.953
		0	0	1	0.696	0.696	0.912	0.957	0.906	0.950	0.906	0.953	0.908	0.956
		-0.4	0	1	0.495	0.851	0.898	0.946	0.908	0.952	0.900	0.951	0.906	0.956
SPE	0.3	0.2	1	1	0.783	0.924	0.854	0.899	0.879	0.928	0.876	0.930	0.890	0.940
		0.4	1	1	0.855	0.864	0.896	0.946	0.903	0.949	0.894	0.942	0.906	0.952
		0.15	1	0	0.690	0.715	0.896	0.943	0.890	0.950	0.899	0.949	0.905	0.953
		-0.2	1	0	0.520	0.850	0.878	0.932	0.866	0.866	0.892	0.944	0.897	0.945
		0.15	0	1	0.690	0.715	0.899	0.950	0.904	0.951	0.901	0.953	0.906	0.955
		-0.2	0	1	0.520	0.850	0.882	0.928	0.892	0.942	0.888	0.939	0.887	0.940
	0.5	0	1	1	0.852	0.852	0.885	0.935	0.895	0.945	0.900	0.951	0.903	0.949
		-0.2	1	1	0.771	0.913	0.851	0.896	0.883	0.930	0.878	0.932	0.896	0.945
		0	1	0	0.696	0.696	0.897	0.948	0.896	0.949	0.885	0.938	0.901	0.953
		-0.4	1	0	0.495	0.851	0.857	0.915	0.883	0.936	0.884	0.936	0.894	0.942
		0	0	1	0.696	0.696	0.898	0.949	0.894	0.945	0.900	0.950	0.904	0.951
		-0.4	0	1	0.495	0.851	0.868	0.925	0.883	0.934	0.880	0.935	0.891	0.946

Table 6.2 Misspecified disease models: Coverage probabilities of SPE-based joint empirical likelihood confidence regions with nominal confidence levels 90% and 95% in the presence of verification bias. Pre. means disease prevalence.

				$n = 300$		$n = 400$		$n = 500$	
Pre.	τ_0	η_0	θ_0	90%	95%	90%	95%	90%	95%
0.3	0.0	0.620	0.779	0.910	0.958	0.908	0.960	0.914	0.956
	-0.2	0.520	0.850	0.907	0.955	0.904	0.954	0.907	0.952
0.5	-0.2	0.599	0.781	0.915	0.957	0.905	0.953	0.909	0.957
	-0.4	0.495	0.851	0.913	0.956	0.909	0.956	0.909	0.961

Table 6.3 Misspecified verification models: Coverage probabilities of SPE-based joint empirical likelihood confidence regions with nominal confidence levels 90% and 95% in the presence of verification bias. Pre. means disease prevalence.

						$n = 300$		$n = 400$		$n = 500$		$n = 600$	
Pre.	τ_0	ν_1	κ_1	η_0	θ_0	90%	95%	90%	95%	90%	95%	90%	95%
0.3	0.2	1	1	0.783	0.924	0.768	0.817	0.805	0.850	0.846	0.889	0.859	0.916
	0.4	1	1	0.855	0.864	0.859	0.905	0.876	0.928	0.893	0.940	0.891	0.946
	0.15	1	0	0.690	0.715	0.885	0.937	0.884	0.941	0.903	0.953	0.897	0.947
	-0.2	1	0	0.520	0.850	0.864	0.917	0.877	0.931	0.888	0.937	0.885	0.939
	0.15	0	1	0.690	0.715	0.892	0.943	0.893	0.946	0.903	0.946	0.900	0.950
	-0.2	0	1	0.520	0.850	0.874	0.920	0.874	0.928	0.881	0.938	0.897	0.947
0.5	0	1	1	0.852	0.852	0.878	0.929	0.907	0.952	0.908	0.958	0.911	0.958
	-0.2	1	1	0.771	0.913	0.853	0.898	0.886	0.924	0.911	0.952	0.897	0.949
	0	1	0	0.696	0.696	0.890	0.945	0.888	0.949	0.911	0.954	0.907	0.953
	-0.4	1	0	0.495	0.851	0.902	0.943	0.902	0.956	0.906	0.954	0.906	0.952
	0	0	1	0.696	0.696	0.909	0.949	0.898	0.947	0.904	0.949	0.895	0.945
	-0.4	0	1	0.495	0.851	0.901	0.951	0.894	0.943	0.907	0.952	0.907	0.952

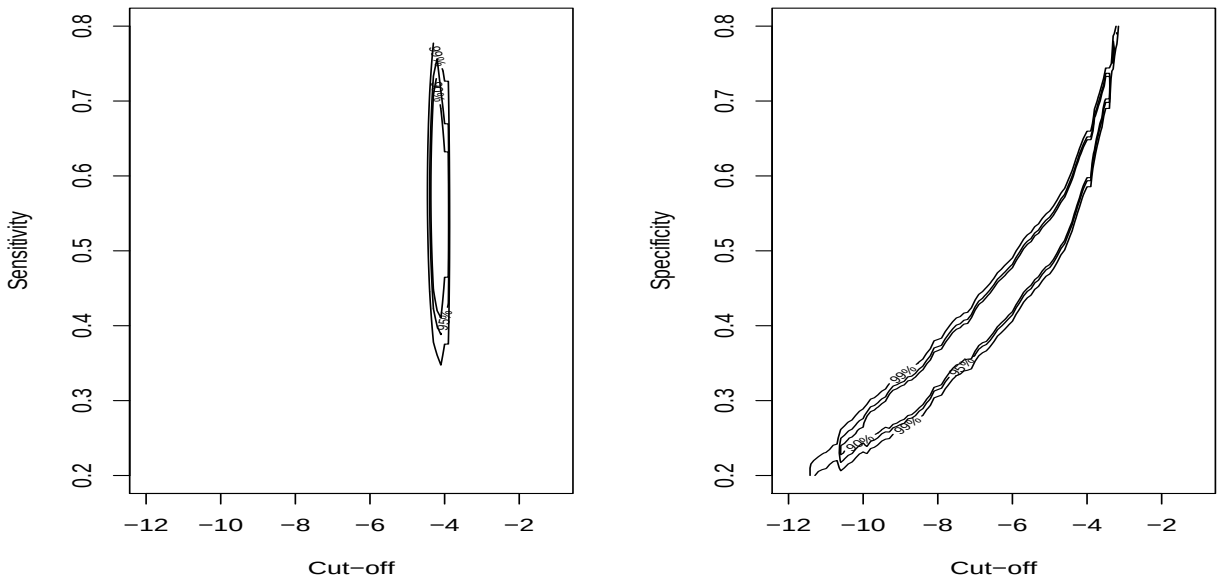

Figure 6.1 Left panel: contour curves of $l^J_{\mathrm{IPW}}(\theta, 0.6, \tau)$, offering confidence regions for the pair (cut-off level, sensitivity) at the fixed specificity level of 0.6. Right panel: contour curves of $l^J_{\mathrm{IPW}}(0.6, \eta, \tau)$, offering confidence regions for the pair (cut-off level, specificity) at the fixed sensitivity level of 0.6. Contours in both panels correspond to nominal confidence levels 90%, 95% and 99%.

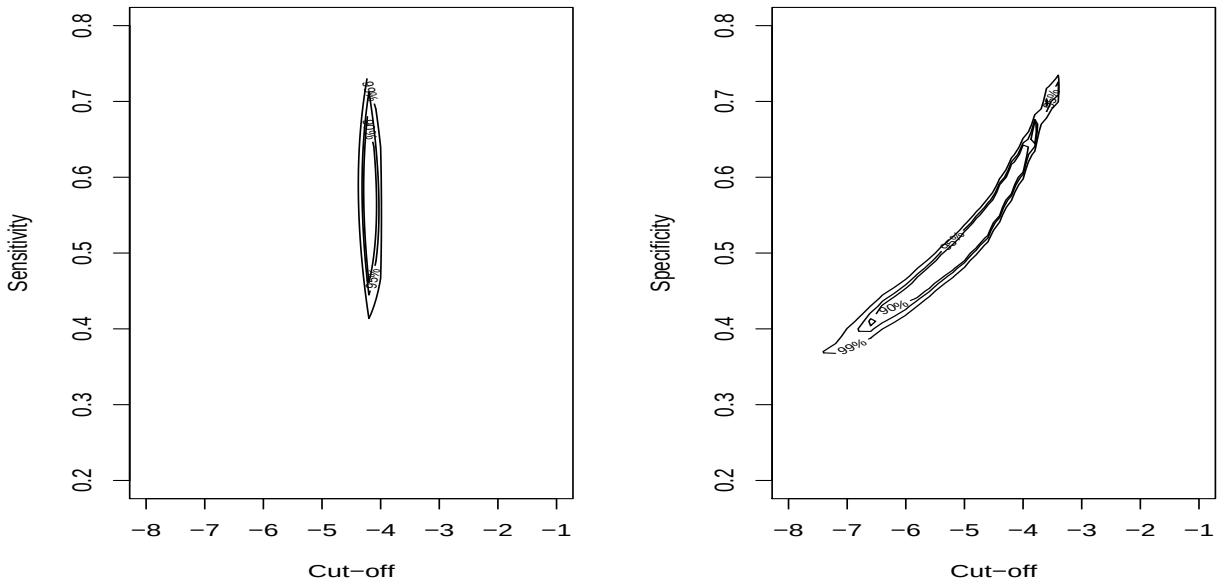

Figure 6.2 Left panel: contour curves of $l^J_{\mathrm{MSI}}(\theta, 0.6, \tau)$, offering confidence regions for the pair (cut-off level, sensitivity) at the fixed specificity level of 0.6. Right panel: contour curves of $l^J_{\mathrm{MSI}}(0.6, \eta, \tau)$, offering confidence regions for the pair (cut-off level, specificity) at the fixed sensitivity level of 0.6. Contours in both panels correspond to nominal confidence levels 90%, 95% and 99%.

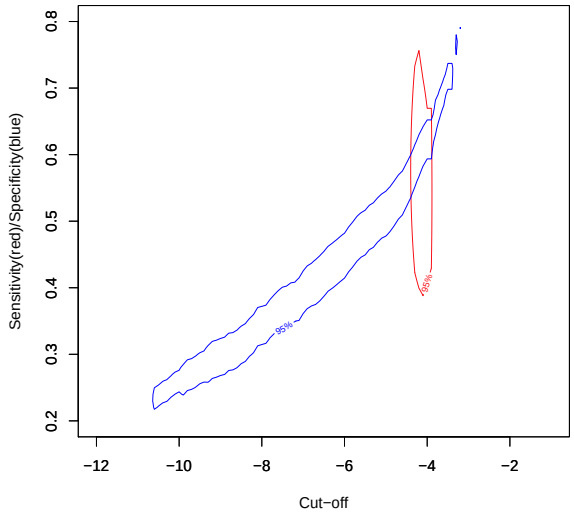

Figure 6.3 IPW-based joint confidence regions: in red, the 95% confidence region for the pair (cut-off level, sensitivity) at the fixed 0.6 level of specificity; In blue, the 95% confidence region for the pair (cut-off level, specificity,) at the fixed 0.6 level of sensitivity.

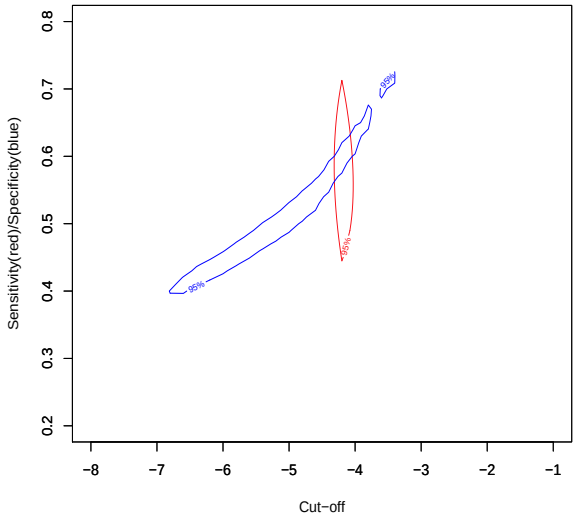

Figure 6.4 MSI-based joint confidence regions: in red, the 95% confidence region for the pair (cut-off level, sensitivity) at the fixed 0.6 level of specificity; In blue, the 95% confidence region for the pair (cut-off level, specificity,) at the fixed 0.6 level of sensitivity.

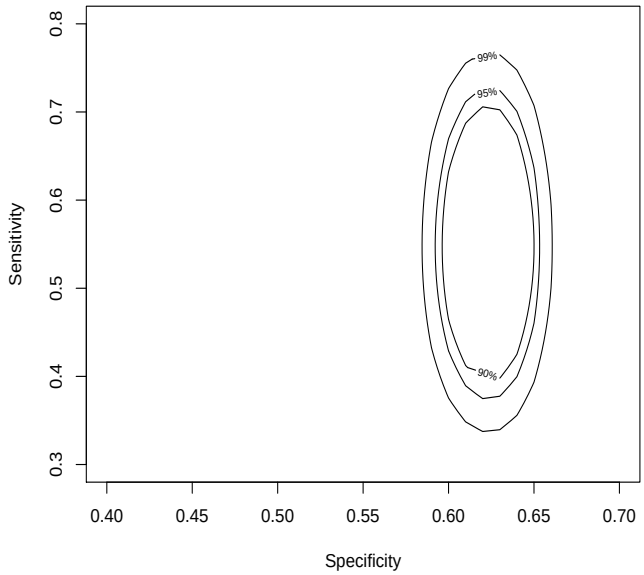

Figure 6.5 Contour curves of $l^J_{\mathrm{IPW}}(\theta, \eta, -4)$, offering the confidence regions for the pair (specificity, sensitivity) when the cut-off level τ is fixed at -4, at nominal coverage levels 90%, 95% and 99%.

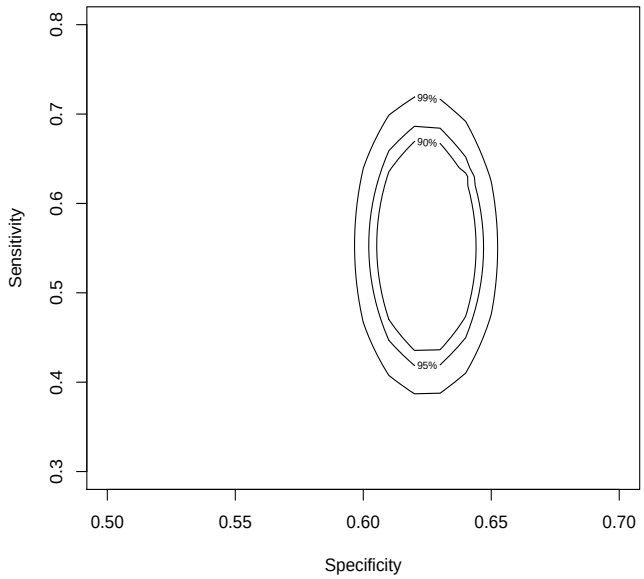

Figure 6.6 Contour curves of $l^J_{\mathrm{MSI}}(\theta, \eta, -4)$, offering the confidence regions for the pair (specificity, sensitivity) when the cut-off level τ is fixed at -4, at nominal coverage levels 90%, 95% and 99%.

CHAPTER 7

DISCUSSION AND FUTURE WORK

This dissertation focuses on the inference of ROC curves with missing data under both MCAR and MAR assumptions, which are usual in practice. Also various bias-corrected empirical likelihood confidence intervals for the sensitivity of ROC curves, the AUC and joint confidence regions for the sensitivity and the specificity with missing data are proposed. Simulation studies are conducted to evaluate the finite sample performance of all proposed methods. Additionally, all new methods have been applied to some real data sets in medical diagnostics to show their practical meanings.

In Chapter 2, We have established the EL-based theory and proposed two EL-based intervals for the sensitivity of a continuous-scale diagnostic test with missing data under the MCAR assumption. In Chapter 3, an imputation-based empirical likelihood method is proposed to construct confidence interval for the AUC with MCAR data. Furthermore, joint confidence regions of the pair (cut-off level, sensitivity) at a fixed value of specificity, the pair (cut-off level, specificity) at a fixed value of sensitivity or the pair (specificity, sensitivity) at a fixed cut-off value are constructed by applying the empirical likelihood method, under the MCAR assumption in Chapter 4 and the MAR assumption in Chapter 5. Chapter 6 applies the jackknife technique to the framework combining the empirical likelihood method and generalized estimating equations with nuisance parameters, proposed in Chapter 5, to simplify the inference procedure.

As we mentioned, MAR and MNAR assumptions are more general than the MCAR assumption, because missing values may depend on observed or even unobserved variables. These assumptions are more flexible to accommodate real cases. Then we could do the inference of the sensitivity given a specificity and the AUC under the more general MAR or MNAR assumption, rather than the MCAR assumption. These problems are challenging be-

cause the missing mechanism depends on different characteristics of patients. Some research have been done on these topics. Rotnitzky *et al.* [26] proposed a doubly robust estimator of the AUC under both MAR and MNAR assumptions. Later, He *et al.* [27] provided a direct estimate of the AUC in the presence of verification bias, and Fluss *et al.* [28] investigated the properties of the doubly robust method for estimating the ROC curve under verification bias. Long *et al.* [30] developed robust statistical methods for estimating the ROC AUC, and the proposed methods used information from auxiliary variables that are potentially predictive of the missingness of the biomarkers or the missing biomarker values.

So far, all these methods are based on normal approximation methods, which involve complicate variance-covariance matrices and require large sample size to obtain satisfactory results. In the future, we could apply the empirical likelihood method and the jackknife empirical likelihood method to construct robust confidence intervals of the sensitivity and the AUC under the MAR and the MNAR assumptions. Additionally, joint confidence regions of the sensitivity and the specificity as well as the cut-off level could be extended to MNAR cases.

REFERENCES

[1] Bamber, D. C. The area above the ordinal dominance graph and the area below the receiver operating characteristic curve graph. *Journal of Mathematical Psychology*, **12**, 387–415, 1975.

[2] Wolfe, D. A., and Hogg, R. V. On constructing statistics and reporting data. *American Statistician*, **25**, 27–30, 1971.

[3] Mann, H. B., and Whitney, D. R. On a test whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, **18**, 50–60, 1947.

[4] Adimari, G., and Chiogna, M. Simple nonparametric confidence regions for the evaluation of continuous-scale diagnostic tests. *The International Journal of Biostatistics* **6(1)**, Article 24, 2010.

[5] Metz, C. E. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, **8**, 283–298, 1978.

[6] Swets, J. A., and Pickett, R. M. *Evaluation of diagnostic systems: Methods from signal detection theory*, New York: Academic Press, 1982.

[7] Pepe, M. S. A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika*, **84**, 595–608, 1997.

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

[9] Linnet, K. Comparison of quantitative diagnostic tests: type I error, power, and sample size. *Statistics in Medicine*, **6**, 147–158, 1987.

[10] Hsieh, F., and Turnbull, B. Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *The Annals of Statistics*, **24**, 25–40, 1996.

[11] Zou, K. H., Hall, W., and Shapiro, D. E. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine*, **16**, 2143–2156, 1997.

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

[13] Pepe, M. S. *The statistical evaluation of medical tests for classification and prediction.* Oxford University Press, 2003.

[14] Zhou, X. H., Obuchowski, N. A. and McClish, D.M. *Statistical methods in diagnostic medicine (2nd edition).* Wiley & Son, 2011.

[15] Rubin, D. B. Inference and missing data. *Biometrika*, **63**, 581–590, 1976.

[16] Little, R. J. A., and Rubin, D. B. *Statistical analysis with missing data (2nd edition)*. New York: John Wiley & Sons, 2002.

[17] Begg, C. B., and Greenes, R. A. Assessment of diagnostic tests when disease is subject to selection bias. *Biometrics*, **39**, 207–216, 1983.

[18] Ransohoff, D. F., and Feinstein, A. R. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *The New England Journal of Medicine*, **299**, 926–930, 1978.

[19] Zhou, X. H. Maximum likelihood estimators of sensitivity and specificity corrected for verification bias. *Communications in Statistics - Theory and Methods*, **22**, 3177–3198, 1993.

[20] Zhou, X. H. A nonparametric maximum likelihood estimator for the receiver operating characteristic curve area in the presence of verification bias. *Biometrics*, **52**, 299–305, 1996.

[21] Hunink, M. G. M., Richardson, D. K., Doubilet, P. M. and Begg, C. B. Testing for fetal pulmonary maturity: ROC analysis involving covariates, verification bias, and combination testing. *Medical Decision Making*, **10**, 201–211, 1990.

[22] Rodenberg, C. A., and Zhou, X. H. ROC curve estimation when covariates affect the verification process. *Biometrics*, **56**, 1256–1262, 2000.

[23] Geert, J. M. G., Van der, H., Rogier, A., Donders, T., Theo, S., and Karel, G. M. M. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research:a clinical example. *Journal of Clinical Epidemiology*, **59**, 1102–1109, 2006.

[24] Alonzo, T. A., and Pepe, M. S. Assessing accuracy of a continuous screening test in the presence of verification bias. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 173–190, 2005.

[25] Carroll, R. J., Ruppert, D. and Stefanski, L. A. *Measurement Error in Nonlinear Models*. London: Chapman & Hall, 1995.

[26] Rotnitzky, A., Faraggi, D., and Schisterman, E. Doubly robust estimation of the area under the receiver-operating characteristic curve in the presence of verification bias. *Journal of the American Statistical Association*, **101**, 1276–1288, 2006.

[27] He, H., Lyness, J. M., and McDermott, M. P. Direct estimation of the area under the receiver operating characteristic curve in the presence of verification bias. Statistics in Medicine. *Statistics in Medicine*, **28**, 361–376, 2009.

[28] Fluss, R., Reiser, B., Faraggi, D. and Rotnitzky, A. Estimation of the ROC curve under verification bias. *Biometrical Journal*, **51**, 475–490, 2009.

[29] Liu, D., and Zhou, X. H. Semiparametric estimation of the covariate-specific ROC curve in presence of ignorable verification bias. *Biometrics*, **67**, 906–916, 2011.

[30] Long, Q., Zhang, X., and Johnson, B. A. Robust estimation of area under ROC curve using auxiliary variables in the presence of missing biomarker values. *Biometrics*, **67**, 559–567, 2011.

[31] Owen, A. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249, 1988.

[32] Owen, A. Empirical likelihood ratio confidence regions. *Annals of Statistics*, **18**, 90–120, 1990.

[33] Owen, A. *Empirical Likelihood*. New York: Chapman & Hall/CRC, 2001.

[34] Claeskens, G., Jing, B. Y., Peng, L., and Zhou, W. Empirical likelihood confidence regions for comparison distributions and ROC curves. *The Canadian Journal of Statistics*, **31**, 173–190, 2003.

[35] Gong, Y., Peng, L., and Qi, Y. C. Smoothed jackknife empirical likelihood method for ROC curve. *Journal of Multivariate Analysis*, **101**, 1520–1531, 2010.

[36] Qin, G. S., Davis, A. E., and Jing, B. Y. Empirical likelihood-based confidence intervals for the sensitivity of a continuous-scale diagnostic test at a fixed level of specificity. *Statistical Methods in Medical Research*, **20**, 217–231, 2011.

[37] Chen, S. X., and Van Keilegom, I. A review on empirical likelihood methods for regressions (with discussions). *Test*, **3**, 415–447, 2009.

[38] Zhou, X. H., and Qin, G. S. Improved confidence intervals for the sensitivity of a continuous-scale test at a fixed level of specificity of a continuous-scale diagnostic test. *Statistics in Medicine*, **24**, 465–477, 2005.

[39] Horváth, L., Horváth, Z., and Zhou, W. Confidence bands for ROC curves. *ournal of Statistical Planning and Inference*, **138**, 1894–1904, 2008.

[40] Qin, G. S., and Zhou, X. H. Empirical likelihood inference for the area under the ROC curve. *Biometrics*, **62**, 612–622, 2006.

[41] Jing, B. Y., Yuan, J., and Zhou, W. Jackknife empirical likelihood. *Journal of the American Statistical Association*, **104**, 1224–1232, 2009.

[42] Gong, Y., Peng, L., and Qi, Y. C. Smoothed jackknife empirical likelihood method for ROC curve. *Journal of Multivariate Analysis*, **101**, 1520–1531, 2010.

[43] Qin, J., and Lawless, J. Empirical likelihood and general estimating equations. *Annuals of Statistics*, **22**, 300–325, 1994.

[44] Hjort, N. L., McKeague, I. W. and Keilegom, I. V. Extending the scope of empirical likelihood. *Ann. Statistics*, **37**, 1079–1111, 2009.

[45] Li, M., Peng, L., and Qi, Y. Reduce computation in profile empirical likelihood medhod. *Canadian Journal of Statistics*, **39**, 370–384, 2011.

[46] Peng, L. Approximate jackknife empirical likelihood method for estimating equations. *Canadian Journal of Statistics*, **40**, 110–123, 2012.

[47] Wang, Q., and Rao, J. N. K. Empirical likelihood-based inference under imputation for missing response data. *The Annals of Statistics*, **30**, 896–924, 2002A.

[48] Wang, Q., and Rao, J. N. K. Empirical likelihood-based inference in linear models with missing response data. *Scandinavian Journal of Statistics*, **29**, 563–576, 2002B.

[49] Liang, H., and Zhou, Y. Semiparametric Inference for ROC Curves with Censoring. *Scandinavian Journal of Statistics*, **35**, 212–227, 2008.

[50] Qin, Y., and Qian, Y. Empirical likelihood confidence intervals for the differences of quantiles with missing data. *Acta Mathematicae Applicatae Sinica (English Series)*, **25**, 105–116, 2009.

[51] Wang, D., and Chen, S. X. Empirical likelihood for estimating equations with missing values. *The Annals of Statistics*, **37**, 490–517, 2009.

[52] Qin, J., Zhang, B., and Leung, D. H. Y. Empirical likelihood in missing data problems. *Journal of the American Statistical Association*, **104**, 1492–1503, 2009.

[53] Chen, J., Rao, J. N. K.,and Sitter, R. R. Efficient random imputation for missing data in complex surveys. *Statistica Sinica*, **10**, 1153–1169, 2000.

[54] Rao, J. N. K., and Shao, J. Jackknife variance estimation with survey data under hot deck imputation. *Biometrika*, **79**, 811–822, 1992.

[55] Neyman, J. Contributions to the theory of sampling of human populations. *Journal of the American Statistical Association*, **33**, 101–116, 1938.

[56] Tenenbein, A. A double sampling scheme for estimating from binomial data with misclassifications. *Journal of the American Statistical Association*, **65**, 1350–1361, 1970.

[57] Alonzo, T. A., Pepe, M. S. and Lumley, T. Estimating disease prevalence in two-phase studies. *Biostatistics*, **4**, 313–326, 2003.

[58] Gao, S., Hui, S. L., Hall, K. S. and Hendrie, H. C. Estimating disease prevalence from two-phase surveys with non-response at the second phase. *Statistics in Medicine*, **19**, 2101–2114, 2000.

[59] Gastwirth, J. L. The first-median test: a two-sided version of the the control median test. *Journal of the American Statistical Association*, **63**, 692–706, 1968.

[60] Chakraborti, S., and Mukerjee, R. A Confidence Interval for a Measure Associated With the Comparison of a Treatment With a Control. *South African Statistical Journal*, **23**, 219–230, 1989.

[61] Platt, R. W., Hanley, J. A., and Yang, H. Bootstrap confidence intervals for the sensitivity of a quantitative diagnostic test. *Statistics in Medicine*, **19**, 313–322, 2000.

[62] Duong, T. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software*, **21**,7, URL http://www.jstatsoft.org/v21/i07, 2007.

[63] Shao, J., and Sitter, R. R. Bootstrap for imputed survey data. *Journal of the American Statistical Association*, **91**, 1278–1288, 1996.

[64] Willett, C., Daly, W., and Warshaw, A. CA 19-9 is an index of response to neoadjunctive chemoradiation therapy in pancreatic cancer. *American Journal of Surgery*, **172**, 350–352, 1996.

[65] Wieand, S., Gail, M. H., James, B. R., and James, K. R. A family of non-parametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*, **76**, 585–592, 1989.

[66] Pepe, M. S., and Cai, T. The analysis of placement values for evaluating discriminatory measures. *Biometrics*, **60**, 528–535, 2004.

[67] Robins, J. M., Rotnitzky, A., Zhao, L. P. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, **89**, 846–866, 1994.

[68] Adimari, G., and Guolo, A. A note on the asymptotic behavior of empirical likelihood statistics. *Statistical Methods and Applications*, **19**, 463–476, 2010.

[69] Schaefer, R. L. Bias correction in maximum likelihood logistic regression. *Statistics in Medicine*, **2**, 71–78, 1983.

[70] Spitzer, R. L., Gibbon, M. and Williams, J. B. W. *Structured Clinical Interview for Axis I DSM-IV Disorders.* Biometrics Research Department, New York State Psychiatric Institute, 1994.

[71] Linn, B. S., Linn, M. W. and Gurel, L. Cumulative illness rating scale. *Journal of the American Geriatrics Society*, **16**, 622–626, 1968.

[72] Tukey, J. W. Bias and confidence in not-quite large samples. *Annals of Statistics*, **29**, 614, 1958.

[73] Norton, S. J., Gorga, M.P., Widen, J.E., Folsom, R.C., Sininger, Y., Cone-Wesson, B., Vohr, B.R., and Fletcher, K.A. Identification of neonatal hearing impairment: a multicenter investigation. *Ear Hearng*, **21**, 348–356, 2000.

[74] Chen, J., and Rao, J. N. K. Asymptotic normality under two-phase sampling designs. *Statistica Sinica*, **17**, 1047–1064, 2007.

[75] Sen, P. K. A note on asymptotically distribution-free confidence bounds for $P(X < Y)$ based on two independent samples. *Sankhyã, Series A*, **29**, 95–102, 1967.

[76] Sen, P. K. On some convergence properties of U-statistics. *Calcutta Statistical Association Bulletin*, **10**, 1–18, 1960.

[77] Qin, G. S., and Jing, B. Y. Empirical likelihood for censored linear regression. *Scandinavian Journal of Statistics*, **28**, 661–673, 2001.

Appendix A

PROOFS OF CHAPTER 2

In order to prove Proposition 1, Proposition 2 and Theorem 1, a few lemmas are necessary.

Lemma 1 *(Chen and Rao [74]) Let U_n, V_n be two sequences of random variables and $\mathscr{B}_n$ be a σ-algebra. Assume that: (i) There exists $\sigma_{1n} > 0$ such that $\sigma_{1n}^{-1} V_n \xrightarrow{d} \mathcal{N}(0,1)$ as $n \to \infty$, where V_n is $\mathscr{B}_n$ measurable; (ii) $E(U_n|\mathscr{B}_n) = 0$ and $VAR(U_n|\mathscr{B}_n) = \sigma_{2n}^2$ such that*

$$\sup_t |P(\sigma_{2n}^{-1} U_n \leq t|\mathscr{B}_n) - \Phi(t)| = o_p(1),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution; (iii) $\gamma_n^2 = \sigma_{1n}^2/\sigma_{2n}^2 = \gamma^2 + o_p(1)$. Then,

$$\frac{U_n + V_n}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} \xrightarrow{d} \mathcal{N}(0,1) \qquad as\ n \to \infty.$$

Proof of Proposition 1. Let $\bar{I}_{1r} = \frac{1}{r_X} \sum_{i \in S_{r_X}} I(X_i \leq x)$ and $\mathscr{B}_m = \sigma(X_i, \delta_{X_i}, i = 1, \ldots, m)$. Then, we have

$$E(I(X_i^* \leq x)|\mathscr{B}_m) = \bar{I}_{1r},\ \mathrm{VAR}(I(X_i^* \leq x)|\mathscr{B}_m) = \frac{1}{r_X} \sum_{i \in S_{r_X}} \{I(X_i \leq x) - \bar{I}_{1r}\}^2,$$

and the following decomposition:

$$\begin{aligned} \widetilde{F}(x) &= \frac{1}{m} \left[\sum_{i \in S_{r_X}} I(X_i \leq x) + m_X \bar{I}_{1r} + \sum_{i \in S_{m_X}} (I(X_i^* \leq x) - \bar{I}_{1r}) \right] \\ &= \frac{1}{r_X} \sum_{i \in S_{r_X}} I(X_i \leq x) + \frac{m_X}{m} \frac{1}{m_X} \sum_{i \in S_{m_X}} (I(X_i^* \leq x) - \bar{I}_{1r}) \equiv I + II. \end{aligned}$$

Based on the fact that the empirical distribution is uniformly consistent, the first part (I) is uniformly consistent with $F(x)$, and the second part (II) uniformly tends to 0. Therefore, $\widetilde{F}(x)$ is a uniformly consistent estimate for $F(x)$. Additionally,

$$\begin{aligned}\sqrt{m}(\widetilde{F}(x)-F(x)) &= \frac{1}{\sqrt{m}}\left[\sum_{i\in S_{r_X}} I(X_i\le x)+m_X\bar{I}_{1r}-mF(x)+\sum_{i\in S_{m_X}}(I(X_i^*\le x)-\bar{I}_{1r})\right]\\ &= \frac{\sqrt{m}}{\sqrt{r_X}}\frac{1}{\sqrt{r_X}}\sum_{i\in S_{r_X}}(I(X_i\le x)-F(x))\\ &\quad+\frac{\sqrt{m_X}}{\sqrt{m}}\frac{1}{\sqrt{m_X}}\sum_{i\in S_{m_X}}(I(X_i^*\le x)-\bar{I}_{1r})\equiv V_m+U_m.\end{aligned}$$

It is clear that V_m is $\mathscr{B}_m$ measurable. Combining the MCAR assumption, the Central Limit Theorem and the Slutsky's Theorem, it follows that

$$V_m \xrightarrow{d} \mathcal{N}(0,\pi_1^{-1}F(x)(1-F(x))).$$

From the Berry-Essen's Central Limit Theorem for independent variables, we have

$$\sup_x |P(\sigma_{2m}^{-1}U_m\le x|\mathscr{B}_m)-\Phi(x)|=o_p(1),$$

where $\sigma_{2m}^2=\frac{m_X}{m}\mathrm{VAR}(I(X_i^*\le x)|\mathscr{B}_m)$.

Also we know $\mathrm{VAR}(I(X_i^*\le x)|\mathscr{B}_m)\xrightarrow{P}F(x)(1-F(x))$. Combined with $\frac{m_X}{m}\xrightarrow{P}1-\pi_1$, we have $\sigma_{2m}^2\xrightarrow{P}(1-\pi_1)F(x)(1-F(x))$.

Then, as $m\to\infty$, by Lemma 1 and the Slutsky's Theorem,

$$\frac{1}{\sqrt{m}}\sum_{i=1}^m I(\tilde{X}_i\le x)-\sqrt{m}F(x)\xrightarrow{d}\mathcal{N}(0,\sigma_X^2)$$

where $\sigma_X^2=(1-\pi_1+\pi_1^{-1})F(x)(1-F(x))$.

Similarly, we could prove the same results for $\widetilde{G}(y)$.

Proof of Corollary 1. The proof is the same as it of Proposition 1 by disregarding

the random hot deck imputation part.

Proof of Proposition 2. This proposition is an extension of the result in the paper of Gastwirth [59]. Thus, some adjustments are needed.

Let e_f be the standardized "excess random variable" defined by

$$e_f = \sum_{i=1}^{m} \frac{r_i - F^{-1}(p)}{\sqrt{mp(1-p)}},$$

where

$$r_i = \begin{cases} 1, & \text{if } \tilde{X}_i \leq F^{-1}(p) \\ 0, & \text{o.w.} \end{cases}$$

and, let

$$e_g^* = \sum_{j=1}^{n} \frac{s_j - G(F^{-1}(p))}{\sqrt{nG(F^{-1}(p))(1-G(F^{-1}(p)))}},$$

where

$$s_j = \begin{cases} 1, & \text{if } \tilde{Y}_j \leq G(F^{-1}(p)) \\ 0, & \text{o.w.} \end{cases}$$

Based on similar calculation in the paper of Gastwirth [59], we have

$$\begin{aligned} u^* &= \widetilde{G}(\widetilde{F}^{-1}(p)) - G(F^{-1}(p)) \\ &= [G(F^{-1}(p))(1-G(F^{-1}(p)))n^{-1}]^{1/2} e_g^* + \frac{g(F^{-1}(p))}{f(F^{-1}(p))} p(1-p) m^{-1/2} e_f + o_p(m^{-1/2}). \end{aligned}$$

It is clear that e_g^* and e_f are independent. By Proposition 1, it follows that

$$e_g^* \xrightarrow{d} \mathcal{N}(0, 1 - \pi_2 + \pi_2^{-1}), \quad e_f \xrightarrow{d} \mathcal{N}(0, 1 - \pi_1 + \pi_1^{-1}).$$

Therefore, combined with $R(p) = 1 - G(F^{-1}(p))$, we have

$$\sqrt{n}\sigma_1^{-1}(p)u^* = \sqrt{n}\sigma_1^{-1}(p)(\widetilde{G}(\widetilde{F}^{-1}(p)) - G(F^{-1}(p)) \xrightarrow{d} \mathcal{N}(0,1),$$

Lemma 2 *Under the same conditions as in Theorem 1, the followings hold:*

(i). $\frac{1}{n}\sum_{j=1}^n \widetilde{W}_j^2(p) \xrightarrow{P} \sigma^2(p)$;

(ii). $\frac{1}{n^{1/2}\sigma_1(p)}\sum_{j=1}^n \widetilde{W}_j(p) \xrightarrow{d} \mathcal{N}(0,1)$.

Proof of Lemma 2.

Lemma 2(i) follows from the uniform consistency of $\widetilde{F}$ and the following:

$$\begin{aligned}
\left|\frac{1}{n}\sum_{j=1}^n \widetilde{W}_j^2(p) - \frac{1}{n}\sum_{j=1}^n W_j^2(p)\right| &\leq \frac{2}{n}\sum_{j=1}^n |I(\widetilde{U}_j \leq 1-p) - I(U_j \leq 1-p)| \xrightarrow{P} 0, \\
\frac{1}{n}\sum_{j=1}^n W_j^2(p) &= \frac{1}{n}\sum_{j=1}^n [I(U_j \leq 1-p) - R(p)]^2 \\
&= \frac{1}{n}\sum_{j=1}^n \left[I(U_j \leq 1-p) + R^2(p) - 2R(p)I(U_j \leq 1-p)\right] \\
&\xrightarrow{P} R(p)(1-R(p)) = \sigma^2(p),
\end{aligned}$$

because we can similarly show $\frac{1}{n}\sum_{j=1}^n I(U_j \leq 1-p) \xrightarrow{P} R(p)$ as in Proposition 1.

Lemma 2(ii) follows from Proposition 2 and the following identity

$$\frac{1}{n^{1/2}\sigma_1(p)}\sum_{j=1}^n \widetilde{W}_j(p) = \sqrt{n}\sigma_1^{-1}(p)(1 - \widetilde{G}(\widetilde{F}^{-1}(p)) - R(p)) = -\sqrt{n}\sigma_1^{-1}(p)(\widetilde{G}(\widetilde{F}^{-1}(p)) - G(F^{-1}(p))).$$

Proof of Theorem 1.

Based on Lemma 2 and the same procedure of the proof of Theorem 3.1 in Qin *et al.* [36], it is straight forward to obtain the result. Key steps are listed as follows.

Similarly, we could prove that $|\tilde{\lambda}| = O_p(n^{-1/2})$. Based on Taylor expansion, we have

$$\begin{aligned}
\tilde{l}(R(p)) &= 2\sum_{j=1}^{n} \log\left(1 + \tilde{\lambda}\widetilde{W}_j(p)\right) \\
&= 2\sum_{j=1}^{n} \left[\tilde{\lambda}\widetilde{W}_j(p) - \frac{1}{2}\left(\tilde{\lambda}\widetilde{W}_j(p)\right)^2\right] + r_n,
\end{aligned}$$

with $|r_n| = O_p(n^{-1/2})$.

By applying similar arguments in Qin *et al.* [36], it follows that

$$\begin{aligned}
\tilde{\lambda} &= \left(\sum_{j=1}^{n} \widetilde{W}_j(p)^2\right)^{-1} \sum_{j=1}^{n} \widetilde{W}_j(p) + O_p(n^{-1}), \\
\sum_{j=1}^{n} \tilde{\lambda}\widetilde{W}_j(p) &= \sum_{j=1}^{n} (\tilde{\lambda}\widetilde{W}_j(p))^2 + O_p(n^{-1/2}).
\end{aligned}$$

Combining all previous results and Lemma 2, it follows that

$$\begin{aligned}
c(p)\tilde{l}(R(p)) &= c(p)\sum_{j=1}^{n} \tilde{\lambda}\widetilde{W}_j(p) + o_p(1) \\
&= \frac{\sigma^2(p)}{\frac{1}{n}\sum_{j=1}^{n} \widetilde{W}_j^2(p)} \left[\frac{1}{n^{1/2}\sigma_1(p)} \sum_{j=1}^{n} \widetilde{W}_j(p)\right]^2 + o_p(1) \\
&\xrightarrow{d} \chi_1^2.
\end{aligned}$$

Appendix B

PROOFS OF CHAPTER 3

In order to prove Theorem 2, a few lemmas are necessary.

Lemma 3 *Under the same conditions as in Theorem 2, the followings hold:*

(i). $\frac{1}{n}\sum_{j=1}^{n}\widetilde{W}_j^2(p) \xrightarrow{P} \sigma_0^2$*, where* $\sigma_0^2 = E[F^2(Y)] - \delta_0^2$*;*

(ii). $(\frac{mn}{m+n})^{1/2}\frac{\widetilde{\delta}-\delta_0}{S} \xrightarrow{d} \mathcal{N}(0,1)$*, where* $\widetilde{\delta}$ *is defined by (3.1).*

Proof of Lemma 3.

(i) From the uniform consistency of $\widetilde{F}$ in Lemma 1, it follows that

$$\frac{1}{n}\sum_{j=1}^{n}\widetilde{F}^2(\widetilde{Y}_j) - \frac{1}{n}\sum_{j=1}^{n}F^2(\widetilde{Y}_j) \xrightarrow{P} 0.$$

By using the similar technique employed in the proof of Lemma 1, we get that

$$\begin{aligned} & \frac{1}{n}\sum_{j=1}^{n}F^2(\widetilde{Y}_j) \\ = \; & \frac{1}{r_Y}\sum_{j\in S_{r_Y}}F^2(Y_j) + \frac{m_Y}{n}\frac{1}{m_Y}\sum_{j\in S_{m_Y}}(F^2(Y_j^*) - \bar{F}_{1r}) \\ \xrightarrow{P} \; & E[F^2(Y)], \end{aligned}$$

where $\bar{F}_{1r} = \frac{1}{r_Y}\sum_{j\in S_{r_Y}}F^2(Y_j)$. Therefore,

$$\frac{1}{n}\sum_{j=1}^{n}\widetilde{F}^2(\widetilde{Y}_j) \xrightarrow{P} E[F^2(Y)].$$

Similarly, we can prove that $\frac{1}{n}\sum_{j=1}^{n}\widetilde{F}(\widetilde{Y}_j) \xrightarrow{P} E[F(Y)] = \delta_0$. Combining the above results,

from Lemma 3(i), it follows that:

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n}\widetilde{W}_j^2(\delta_0) &= \frac{1}{n}\sum_{j=1}^{n}\left(\widetilde{F}(\widetilde{Y}_j)-\delta_0\right)^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}\widetilde{F}^2(\widetilde{Y}_j)-\frac{2\delta_0}{n}\sum_{j=1}^{n}\widetilde{F}(\widetilde{Y}_j)+\delta_0^2 \\
&\xrightarrow{P} E[F^2(Y)]-\delta_0^2=\sigma_0^2.
\end{aligned}$$

(ii) If the data set is complete, Sen [75] has proved similar result. Based on imputed data, some necessary modifications are needed. Let

$$\begin{aligned}
&\alpha_\delta=\int_0^1 F^2(y)\mathrm{d}G(y), \quad \beta_\delta=\int_0^1[1-G(x)]^2\mathrm{d}F(x), \\
&n_0=\frac{mn}{m+n}, \\
&\mathscr{B}_n=\sigma(\tilde{Y}_j, j=1,\ldots,n), \qquad \mathscr{A}_m=\sigma(\tilde{X}_i, i=1,\ldots,m).
\end{aligned}$$

Then, the variance of $\sqrt{n_0}\widetilde{\delta}$ can be calculated as follows:

$$\mathrm{VAR}(\sqrt{n_0}\widetilde{\delta})=\mathrm{VAR}\left(E(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n)\right)+E\left(\mathrm{VAR}(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n)\right). \tag{B.1}$$

For the first term of the right-hand side in (B.1), from

$$\begin{aligned}
E(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n) &= \frac{\sqrt{n_0}}{mn}\sum_{j=1}^{n}\sum_{i=1}^{m}E[I(\tilde{X}_i\le\tilde{Y}_j)|\mathscr{B}_n] \\
&= \frac{\sqrt{n_0}}{n}\sum_{j=1}^{n}[\widetilde{F}(\tilde{Y}_j)],
\end{aligned}$$

it follows that

$$\begin{aligned}
\mathrm{VAR}\left(E(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n)\right) &= \mathrm{VAR}\left(\frac{\sqrt{n_0}}{n}\sum_{j=1}^{n}[\widetilde{F}(\tilde{Y}_j)]\right) \\
&\to \frac{1}{1+\tau}(1-\pi_2+\pi_2^{-1})(\alpha_\delta-\delta_0^2),
\end{aligned}$$

where the last step follows from Lemma 1.

As for the second term of the right-hand side in (B.1), from

$$
\begin{aligned}
&\mathrm{VAR}\left(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n\right)\\
&= \frac{n_0}{m^2n^2}\mathrm{VAR}\left(\sum_{j=1}^{n}\sum_{i=1}^{m} I(\tilde{X}_i \le \tilde{Y}_j)|\mathscr{B}_n\right)\\
&= \frac{n_0 m}{m^2n^2}\left[(1-\pi_1+\pi_1^{-1})\mathrm{VAR}\left(\sum_{j=1}^{n} I(X \le \tilde{Y}_j|\mathscr{B}_n)\right) + o_P(1)\right]\\
&= \frac{n_0}{mn^2}\left[(1-\pi_1+\pi_1^{-1})\left(\sum_{j=1}^{n} F(\tilde{Y}_j) + 2\sum_{j\le k} E(I(X \le \tilde{Y}_j)I(X \le \tilde{Y}_k)|\mathscr{B}_n) - \big(\sum_{j=1}^{n} F(\tilde{Y}_j)\big)^2\right) + o_P(1)\right],
\end{aligned}
$$

it follows that

$$
\begin{aligned}
&E\left(\mathrm{VAR}(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n)\right)\\
=\ & \frac{n_0}{mn^2}\Bigg[(1-\pi_1+\pi_1^{-1})\Bigg(\sum_{j=1}^{n} EF(\tilde{Y}_j)\\
&+2\sum_{j\le k} E\big(E(I(X \le \tilde{Y}_j)I(X \le \tilde{Y}_k)|\mathscr{B}_n)\big) - E\big(\sum_{j=1}^{n} F(\tilde{Y}_j)\big)^2\Bigg) + o(1)\Bigg]\\
=\ & \frac{n_0}{mn^2}\Bigg[(1-\pi_1+\pi_1^{-1})\Bigg(n\delta_0 + 2\sum_{j\le k} E\big(E(I(X \le \tilde{Y}_j)I(X \le \tilde{Y}_k)|\mathscr{B}_n)\big)\\
&- \big(\mathrm{VAR}\big(\sum_{j=1}^{n} F(\tilde{Y}_j)\big) + E^2\big(\sum_{j=1}^{n} F(\tilde{Y}_j)\big)\big)\Bigg) + o(1)\Bigg]\\
=\ & \frac{n_0}{mn^2}\Bigg[(1-\pi_1+\pi_1^{-1})\Bigg(n\delta_0 + 2\sum_{j\le k} E\big(E(I(X \le \tilde{Y}_j)I(X \le \tilde{Y}_k)|\mathscr{B}_n)\big)\\
&- \big((1-\pi_2+\pi_2^{-1})n(\alpha_\delta - \delta_0^2) + (n\delta_0)^2 + o(n)\big)\Bigg) + o(1)\Bigg].
\end{aligned}
$$

From

$$
\begin{aligned}
& \sum_{j\le k} E\left(E(I(X\le \tilde{Y}_j)I(X\le \tilde{Y}_k)|\mathscr{B}_n)\right) \\
= \;& E\left[E\left(\sum_{j\le k} I(X\le \tilde{Y}_j)I(X\le \tilde{Y}_k)|\mathscr{B}_n\right)\right] \\
= \;& EE\left[\left(\sum_{\substack{j\le k\\ j,k\in S_{r_Y}}} + \sum_{\substack{j\le k\\ j\in S_{r_Y},k\in S_{m_Y}}} + \sum_{\substack{j\le k\\ j\in S_{m_Y},k\in S_{r_Y}}} + \sum_{\substack{j\le k\\ j,k\in S_{m_Y}}}\right) I(X\le \tilde{Y}_j)I(X\le \tilde{Y}_k)\middle|\mathscr{B}_n\right] \\
= \;& EE\left[\left(\sum_{\substack{j\le k\\ j,k\in S_{r_Y}}} + \sum_{\substack{j\le k\\ j\in S_{r_Y},k\in S_{m_Y}}} + \sum_{\substack{j\le k\\ j\in S_{m_Y},k\in S_{r_Y}}} + \sum_{\substack{j\le k\\ j,k\in S_{m_Y}}}\right) I(X\le \tilde{Y}_j)I(X\le \tilde{Y}_k)\middle|X)\right] \\
= \;& EE\Bigg[\sum_{\substack{j\le k\\ j,k\in S_{r_Y}}} \big(1-G(X)\big)^2 \\
& + \sum_{\substack{j\le k\\ j\in S_{r_Y},k\in S_{m_Y}}} \left(\frac{1}{r_Y}(1-G(X)) + \frac{r_Y-1}{r_Y}(1-G(X))^2\right) \\
& + \sum_{\substack{j\le k\\ j\in S_{m_Y},k\in S_{r_Y}}} \left(\frac{1}{r_Y}(1-G(X)) + \frac{r_Y-1}{r_Y}(1-G(X))^2\right) \\
& + \sum_{\substack{j\le k\\ j,k\in S_{m_Y}}} \frac{1}{r_Y^2}\left(r_Y(1-G(X)) + r_Y(r_Y-1)(1-G(X))^2\right)\Big|X,\sigma(\delta_{Y_j}, j=1,\ldots,n)\Bigg] \\
= \;& EE\Bigg[\frac{n(n-1)}{2}(1-G(X))^2 \\
& + \sum_{\substack{j\le k\\ j \text{ or } k\in S_{m_Y}}} \frac{1}{r_Y}\left((1-G(X)) - (1-G(X))^2\right)\Big|X,\sigma(\delta_{Y_j}, j=1,\ldots,n)\Bigg] \\
= \;& EE\Bigg[\frac{n(n-1)}{2}(1-G(X))^2 \\
& + \frac{n(n-1)-r_Y(r_Y-1)}{2r_Y}\left((1-G(X)) - (1-G(X))^2\right)\Big|X,\sigma(\delta_{Y_j}, j=1,\ldots,n)\Bigg],
\end{aligned}
$$

it follows that

$$
\begin{aligned}
& 2\sum_{j\le k} E\left(E(I(X\le \tilde{Y}_j)I(X\le \tilde{Y}_k)|\mathscr{B}_n)\right) \\
= \;& EE\Big[2(1-G(X))^2 + (\frac{n(1-\pi_2^2)}{\pi_2} + O_P(n))\Big((1-G(X)) - (1-G(X))^2\Big)\Big|X\Big] \\
= \;& n(n-1)\beta_\delta + EE\Big[(\frac{n(1-\pi_2^2)}{\pi_2} + O_P(n))\Big((1-G(X)) - (1-G(X))^2\Big)\Big|X\Big].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& E\left(\mathrm{VAR}(\sqrt{n_0}\widetilde{\delta}|\mathscr{B}_n)\right) \to \frac{\tau}{1+\tau}(1-\pi_1+\pi_1^{-1})(\beta_\delta - \delta_0^2), \\
& \mathrm{VAR}(\sqrt{n_0}\widetilde{\delta}) \to \frac{1}{1+\tau}(1-\pi_2+\pi_2^{-1})(\alpha_\delta - \delta_0^2) \\
& \qquad\qquad + \frac{\tau}{1+\tau}(1-\pi_1+\pi_1^{-1})(\beta_\delta - \delta_0^2).
\end{aligned}
$$

In order to prove Lemma 3(ii), we need to show that S and $\mathrm{VAR}(\sqrt{n_0}\widetilde{\delta})$ converge to the same limit. Let

$$
\begin{aligned}
V_{10}(\tilde{X}_i) &= \frac{1}{n}\sum_{j=1}^{n} I(\tilde{X}_i \le \tilde{Y}_j), \quad i = 1,\ldots,m; \\
V_{01}(\tilde{Y}_j) &= \frac{1}{m}\sum_{i=1}^{m} I(\tilde{X}_i \le \tilde{Y}_j), \quad j = 1,\ldots,n.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
S_{10}^2 &= \frac{1}{m-1}\sum_{i=1}^{m}\left[V_{10}(\tilde{X}_i) - \widetilde{\delta}\right]^2 \\
&= \frac{1}{m-1}\sum_{i=1}^{m}\left[V_{10}^2(\tilde{X}_i) - 2V_{10}(\tilde{X}_i)\widetilde{\delta} + \widetilde{\delta}^2\right].
\end{aligned}
$$

By Lemma 1, we have

$$
\widetilde{\delta} \xrightarrow{P} \delta_0, \text{ and } V_{10}(\tilde{X}_i) \xrightarrow{P^*} 1 - G(\tilde{X}_i)
$$

where P^* is the probability measure on $\mathscr{A}_m$. Thus,

$$S_{10}^2 \xrightarrow{P} \beta_\delta - \delta_0^2.$$

Similarly, we have

$$S_{01}^2 \xrightarrow{P} \alpha_\delta - \delta_0^2.$$

Therefore,

$$\begin{aligned} S^2 &= \frac{m(1-\pi_2+\pi_2^{-1})S_{01}^2 + n(1-\pi_1+\pi_1^{-1})S_{10}^2}{m+n} \\ &\xrightarrow{P} \frac{1}{1+\tau}(1-\pi_2+\pi_2^{-1})(\alpha_\delta - \delta_0^2) \\ &\quad + \frac{\tau}{1+\tau}(1-\pi_1+\pi_1^{-1})(\beta_\delta - \delta_0^2). \end{aligned}$$

Based on Sen [75], S_{01}^2 and S_{10}^2 have the alternative algebraic expressions in Theorem 2. Finally, from Lemma 1, the Slutsky's theorem and the similar procedures of structural convergence of U-statistics in Sen ([75], [76]), it follows that

$$\sqrt{n_0}\frac{\widetilde{\delta} - \delta_0}{S} = \left(\frac{mn}{m+n}\right)^{1/2} \frac{\widetilde{\delta} - \delta_0}{S} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof of Theorem 2.

Based on Lemma 3 and the same procedure of the proof of Theorem 1 in Qin and Zhou [40], it is straight forward to obtain the result.

Proof of Theorem 3.

Based on Theorem 2 and the same procedure of the proof of Theorem 2 in Qin and Zhou [40], it is straight forward to obtain the result.

Appendix C

PROOFS OF CHAPTER 4

Proof of Theorem 4: With similar arguments with Adimari and Chiogna [4], $\tilde{l}(\theta_0, \eta_0, \tau_0)$ is finite with probability tending to 1, as $\min\{m, n\} \to +\infty$. By using Taylor expansion and some algebra, it follows that

$$\tilde{l}(\theta_0, \eta_0, \tau_0) = m\frac{[\widetilde{F}_X(\tau_0) - \eta_0]^2}{\eta_0(1-\eta_0)} + n\frac{[\widetilde{F}_Y(\tau_0) - 1 + \theta_0]^2}{\theta_0(1-\theta_0)} + o_p(1).$$

Bases on the results in Proposition 1,

$$\sqrt{m}[\widetilde{F}_X(\tau_0) - \eta_0] \xrightarrow{d} \mathcal{N}(0, \sigma_X^2)$$

where $\sigma_X^2 = (1 - \pi_1 + \pi_1^{-1})\eta_0(1-\eta_0)$, and

$$\sqrt{n}[\widetilde{F}_Y(\tau_0) - 1 + \theta_0] \xrightarrow{d} \mathcal{N}(0, \sigma_Y^2)$$

where $\sigma_Y^2 = (1 - \pi_2 + \pi_2^{-1})\theta_0(1-\theta_0)$.

Then the result follows.

Appendix D

PROOFS OF CHAPTER 5

Proof of Theorem 5: By Taylor Expansion, we have that

$$\widehat{l}(\psi_0) = 2\sum_{i=1}^{n}\log(1+\widehat{t}'U(W_i,\psi_0,\widehat{\beta})) = 2\left(\sum_{i=1}^{n}\widehat{t}'U(W_i,\psi_0,\widehat{\beta}) - \frac{\sum_{i=1}^{n}(\widehat{t}'U(W_i,\psi_0,\widehat{\beta}))^2}{2} + o_p(1)\right).$$

Based on the standard methods used in empirical likelihood literature ([31],[32],[33]), we can get

$$\sum_{i=1}^{n}\widehat{t}'U(W_i,\psi_0,\widehat{\beta}) = \sum_{i=1}^{n}(\widehat{t}'U(W_i,\psi_0,\widehat{\beta}))^2 + O_p(n^{-1/2}).$$

Therefore,

$$\widehat{l}(\psi_0) = \sum_{i=1}^{n}\widehat{t}'U(W_i,\psi_0,\widehat{\beta}) + o_p(1).$$

Observe that

$$\begin{aligned}
0 &= Q_{1n}(\widehat{t},\psi_0,\widehat{\beta}) \\
&= Q_{1n}(0,\psi_0,\boldsymbol{\beta}_0) + \frac{\partial Q_{1n}(0,\psi_0,\beta_0)}{\partial t'}(\widehat{t}-0) + \frac{\partial Q_{1n}(0,\psi_0,\beta_0)}{\partial \beta'}(\widehat{\beta}-\beta_0) + o_p(\delta_n) \\
0 &= Q_{2n}(\widehat{\beta}) = Q_{2n}(\beta_0) + \frac{\partial Q_{2n}(\beta_0)}{\partial \beta'}(\widehat{\beta}-\beta_0) + o_p(\delta_n),
\end{aligned}$$

where $\delta_n = \|\widehat{t}\| + \|\widehat{\beta}-\beta_0\|$.

In matrix notation, we have

$$\begin{pmatrix} \widehat{t} \\ \widehat{\beta}-\beta_0 \end{pmatrix} = S_n^{-1}\begin{pmatrix} -Q_{1n}(0,\psi_0,\beta_0) + o_p(\delta_n) \\ -Q_{2n}(\beta_0) + o_p(\delta_n) \end{pmatrix},$$

where

$$
\begin{aligned}
S_n &= \begin{pmatrix} \frac{\partial Q_{1n}(0,\psi_0,\beta_0)}{\partial t'} & \frac{\partial Q_{1n}(0,\psi_0,\beta_0)}{\partial \beta'} \\ 0 & \frac{\partial Q_{2n}(\beta_0)}{\partial \beta'} \end{pmatrix} \\
&\xrightarrow{P} S = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix} = \begin{pmatrix} E\left[\frac{\partial Q_1(W_i,0,\psi_0,\beta_0)}{\partial t'}\right] & E\left[\frac{\partial Q_1(W_i,0,\psi_0,\beta_0)}{\partial \beta'}\right] \\ 0 & E\left[\frac{\partial V(W_i,\beta_0)}{\partial \beta'}\right] \end{pmatrix}.
\end{aligned}
$$

From this and $Q_{1n}(0,\psi_0,\beta_0) = (1/n)\sum_{i=1}^n U(W_i,\psi_0,\beta_0) = O_p(n^{-1/2})$, it follows that $\delta_n = O_p(n^{-1/2})$.

By applying the block-wise inverse technique,

$$
S^{-1} = \begin{pmatrix} S_{11}^{-1} & -S_{11}^{-1}S_{12}S_{22}^{-1} \\ 0 & S_{22}^{-1} \end{pmatrix},
$$

then

$$
\widehat{t} = -S_{11}^{-1}Q_{1n}(0,\psi_0,\beta_0) + S_{11}^{-1}S_{12}S_{22}^{-1}Q_{2n}(\beta_0) + O_p(n^{-1/2}).
$$

By regular methods of estimating equations, it follows that

$$
\sqrt{n}(-S_{22})(\widehat{\beta} - \beta_0) = \sqrt{n}Q_{2n}(\beta_0) + o_p(1).
$$

Combined with that Taylor expansion of $Q_{1n}(0,\psi_0,\widehat{\beta})$ at β_0, we have

$$
Q_{1n}(0,\psi_0,\widehat{\beta}) = Q_{1n}(0,\psi_0,\beta_0) + S_{12}(\widehat{\beta} - \beta_0) + o_p(n^{-1/2}).
$$

Therefore,

$$
\begin{aligned}
& \widehat{l}(\psi_0) \\
= \; & n\left(Q_{1n}(0,\psi_0,\beta_0) + S_{12}(\widehat{\beta}-\beta_0)\right)' \left(-S_{11}^{-1}Q_{1n}(0,\psi_0,\beta_0) + S_{11}^{-1}S_{12}S_{22}^{-1}Q_{2n}(\beta_0)\right) + o_p(1) \\
= \; & nQ_{1n}'(0,\psi_0,\beta_0)(-S_{11})^{-1}Q_{1n}(0,\psi_0,\boldsymbol{\beta}_0) + nQ_{1n}'(0,\psi_0,\beta_0)(-S_{11})^{-1}S_{12}(-S_{22})^{-1}Q_{2n}(\beta_0) \\
& +n(-S_{22}(\widehat{\beta}-\beta_0))'(-S_{22})^{-1}S_{12}'[(-S_{11})^{-1}Q_{1n}(0,\psi_0,\beta_0) + (-S_{11})^{-1}S_{12}(-S_{22})^{-1}Q_{2n}(\beta_0)] + o_p(1) \\
= \; & n\left(Q_{1n}'(0,\psi_0,\beta_0), Q_{2n}'(\beta_0)\right) \begin{pmatrix} (-S_{11})^{-1} & (-S_{11})^{-1}S_{12}(-S_{22})^{-1} \\ (-S_{22})^{-1}S_{12}'(-S_{11})^{-1} & (-S_{22})^{-1}S_{12}'(-S_{11})^{-1}S_{12}(-S_{22})^{-1} \end{pmatrix} \\
& \times \begin{pmatrix} Q_{1n}(0,\psi_0,\beta_0) \\ Q_{2n}(\beta_0) \end{pmatrix} + o_p(1)
\end{aligned}
$$

Let

$$
S^* = \mathrm{Cov}\left[(U'(W_i,\psi_0,\beta_0), V'(W_i,\beta_0))'\right].
$$

Then,

$$
\sqrt{n}(S^*)^{-1/2} \begin{pmatrix} Q_{1n}(0,\psi_0,\beta_0) \\ Q_{2n}(\beta_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, I_{(p+q)\times(p+q)}).
$$

Finally, we have that

$$
\begin{aligned}
\widehat{l}(\psi_0) &= \sqrt{n}\left(Q'_{1n}(0,\psi_0,\beta_0), Q'_{2n}(\beta_0)\right)(S^*)^{-1/2} \\
&\quad \times (S^*)^{1/2}\begin{pmatrix} (-S_{11})^{-1} & (-S_{11})^{-1}S_{12}(-S_{22})^{-1} \\ (-S_{22})^{-1}S'_{12}(-S_{11})^{-1} & (-S_{22})^{-1}S'_{12}(-S_{11})^{-1}S_{12}(-S_{22})^{-1} \end{pmatrix} \\
&\quad \times (S^*)^{1/2}\sqrt{n}(S^*)^{-1/2}\begin{pmatrix} Q_{1n}(0,\psi_0,\beta_0) \\ Q_{2n}(\beta_0) \end{pmatrix} \\
&= \left[\sqrt{n}(S^*)^{-1/2}\left(Q'_{1n}(0,\psi_0,\beta_0), Q'_{2n}(\beta_0)\right)'\right]'\Lambda\left[\sqrt{n}(S^*)^{-1/2}\begin{pmatrix} Q_{1n}(0,\psi_0,\beta_0) \\ Q_{2n}(\beta_0) \end{pmatrix}\right] + o_p(1) \\
&\xrightarrow{d} Y'\Lambda Y
\end{aligned}
$$

where $Y \sim \mathcal{N}(0, I_{(p+q)\times(p+q)})$.

Note that

$$
\Lambda = (S^*)^{\frac{1}{2}}\begin{pmatrix} I \\ (-S_{22})^{-1}S'_{12} \end{pmatrix}(-S_{11})^{-1}\begin{pmatrix} I & S_{12}(-S_{22})^{-1} \end{pmatrix}(S^*)^{\frac{1}{2}},
$$

thus Λ has rank p.

By applying the same method in the Lemma 3 of Qin and Jing [77],

$$
\widehat{l}(\psi_0) \xrightarrow{d} r_1\chi^2_{1,1} + \cdots + r_p\chi^2_{1,p}
$$

where $\chi^2_{1,j}, j = 1,\ldots,p$ are independent chi-squared random variables with one degree of freedom, and the weights $r_1,\ldots,r_p$ are none-zero eigenvalues of Λ.

Proofs of Theorem 6-8:

It is straightforward to obtain these theorems by applying Theorem 5. Only a few modifications are needed.

In Theorem 5, note that

$$S^* = \begin{pmatrix} -S_{11} & -S_{12} \\ -S'_{12} & -S_{22} \end{pmatrix}.$$

Thus the Λ^* in Remark 2 is equal to

$$\Lambda^* = I - (-S_{11})^{-1} S_{12} (-S_{22})^{-1} S'_{12}.$$

Therefore, Λ shares the same eigenvalues with Λ^*.