Georgia State University
ScholarWorks @ Georgia State University

Mathematics Dissertations | Department of Mathematics and Statistics

12-5-2011

Statistical Inferences for the Youden Index

Haochuan Zhou

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

Recommended Citation
Zhou, Haochuan, "Statistical Inferences for the Youden Index" (2011). *Mathematics Dissertations.* Paper 5.

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 INFERENCES FOR THE YOUDEN INDEX

by

HAOCHUAN ZHOU

Under the Direction of Dr. Gengsheng Qin

ABSTRACT

In diagnostic test studies, one crucial task is to evaluate the diagnostic accuracy of a test. Currently, most studies focus on the Receiver Operating Characteristics Curve and the Area Under the Curve. On the other hand, the Youden index, widely applied in practice, is another comprehensive measurement for the performance of a diagnostic test. For a continuous-scale test classifying diseased and non-diseased groups, finding the Youden index of the test is equivalent to maximize the sum of sensitivity and specificity for all the possible values of the cut-point. This dissertation concentrates on statistical inferences for the Youden index. First, an auxiliary tool for the Youden index, called the diagnostic curve, is defined and used to evaluate the diagnostic test. Second, in the paired-design study to assess the diagnostic accuracy of two biomarkers, the difference in paired Youden indices frequently acts as an evaluation standard. We propose an exact confidence interval for the difference in paired Youden indices based on generalized pivotal quantities. A maximum likelihood estimate-based interval and a bootstrap-based interval are also included in the study. Third, for certain diseases, an intermediate level exists between diseased and non-diseased status. With such concern, we define the Youden index for three ordinal groups, propose the empirical estimate of the Youden index, study the asymptotic properties of the empirical Youden index estimate, and construct parametric and nonparametric confidence intervals for the Youden index. Finally, since covariates often affect the accuracy of a diagnostic test, therefore, we propose estimates for

the Youden index with a covariate adjustment under heteroscedastic regression models for the test results. Asymptotic properties of the covariate-adjusted Youden index estimators are investigated under normal error and non-normal error assumptions.

INDEX WORDS: Diagnostic test, Sensitivity, Specificity, ROC, AUC, Youden index, Optimal cut-off point, Paired design, Generalized pivotal quantity, Exact confidence interval, Covariates adjustment, Three ordinal group, Diagnostic curve, Empirical estimate

STATISTICAL INFERENCES FOR THE YOUDEN INDEX

by

HAOCHUAN ZHOU

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

2011

Copyright by
Haochuan Zhou
2011

STATISTICAL INFERENCES FOR THE YOUDEN INDEX

by

HAOCHUAN ZHOU

Committee Chair: Dr. Gengsheng Qin

Committee: Dr. Xu Zhang

Dr. Ruiyan Luo

Dr. Yi Zhao

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2011

DEDICATION

To my parents, adviser, and girl friend.

ACKNOWLEDGEMENTS

In my Ph.D. studying, myriad of genial assistances have been obtained from substantive people who are my mentors, instructors, classmates, and friends. Endless gratitude to give voice to by names; however, I would love to record my truly respects to the following persons.

The paramountly inexpressible gratitude to my adviser, Dr. Gengsheng Qin, for all his enlightening, indoctrinating, supporting, comprehending, and patient assistances. Dr. Qin scarified a lot of his efforts to assist me into the academic disquisitions step by step, without his companying, there is no chance I could wipe off all the brambles on the way to this researches. Rather than studying guidance, Dr. Qin also affectionately offered me valuable suggestions for my future career based on his insightful understanding of my characteristics and personalities. My appreciations to Dr. Qin are ceaseless.

With this opportunity, I would like to present my gratitude to my dissertation committee members, Dr. Xu Zhang, Dr. Ruiyan Luo and Dr. Yi Zhao, for spending their precious time on reading this dissertation and suggesting critical comments to assist improving this dissertation.

Also, numerous thanks to all the other faculties and staffs in the Department of Mathematics and Statistics at Georgia State University for their teaching and helping. Without their generosity and support, I would not achieve the requirements for Ph.D. Degree. Specially, I would love to appreciate Dr. Yu-Sheng Hsu, Dr. Yixin Fang, Dr. Xu Zhang, and Dr. Guantao Chen.

Acknowledgements to my esteemed friends, Meng Zhao, Xin Huang, Yitao Wang, Ruyang Cui, Binhuan Wang, Haci M. Akcin, Qian Chen, Yingge Qu and others, for those critical, wisdom suggestions regarding to this dissertation, for those discussions about the truth of life and mostly for their continuing support and encouragements from all aspects. They are my brothers and sisters without consanguinity and my treasures for this life.

My special gratitude delivers to Michael Fost for helping me improving the writing of this dissertation.

The unspeakable thanks are to my parents, Longhua Zhou, Weigang Zhao, and my girl friend, Jing Yang. Their loves are my greatest forces. I would love to devote my rest life to return more than the love that they contributed to me.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . . **v**

LIST OF FIGURES . . . . . **ix**

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

LIST OF ABBREVIATIONS . . . . . **xii**

Chapter 1 INTRODUCTION **1**

1.1 Diagnostic Tests and Their Evaluations 1

1.2 Estimation for a Youden Index 4

1.3 Comparison Between Two Diagnostic Tests 6

1.4 Evaluation of the Diagnostic Test Classifying Three Ordinal Groups 6

1.5 Covatiates Adjustment for ROC Related Indices 8

Chapter 2 DIAGNOSTIC CURVE AND INFERENCES FOR A YOUDEN INDEX **9**

2.1 Introduction 9

2.2 Diagnostic Curve 10

2.2.1 Definition 10

2.2.2 Point-wise Confidence Intervals for the Diagnostic Curve 11

2.2.3 Simulation Studies 13

2.3 Existing Methods for Inferences of a Youden Index 14

2.3.1 Point Estimation 14

2.3.2 Confidence Intervals for the Youden index 17

2.3.3 Practical Implications of Existing Methods 19

Chapter 3 INFERENCES FOR THE DIFFERENCE OF TWO YOUDEN INDICES 20

3.1 Introduction 20

3.2 Bootstrap and Maximum Likelihood Based Intervals 20

3.3 Exact Confidence Interval 24

3.4 Simulation Studies 27

3.5 A Real Application 29

Chapter 4 INFERENCES FOR YOUDEN INDEX OF THREE ORDINAL GROUPS 31

4.1 Introduction 31

4.2 New Definition of The Youden Index for Three Ordinal Groups 32

4.3 Estimations of Youden index for Three Ordinal Groups 34

4.3.1 Parametric Estimation 34

4.3.2 Nonparametric Estimation 35

4.4 Asymptotic Properties for the Parametric and Nonparametric Estimators 36

4.5 Confidence Intervals for the Youden index of Three Ordinal Groups 37

4.6 Simulation Studies 41

Chapter 5 NONPARAMETRIC COVARIATES ADJUSTMENT FOR YOUDEN INDEX 44

5.1 Introduction 44

5.2 Model and Methods 44

5.2.1 Heteroscedastic Regression Models for the Test Results 44

5.2.2 Covariate-adjusted Youden Index under Normal Error Assumption 45

5.2.3 Covariate-adjusted Youden Index without Normal Error Assumption . . 47

5.3 Asymptotic Properties of the Covariate-adjusted Estimators for the Youden Index 48

5.3.1 Asymptotic Properties of $\widehat{J}_N(z)$ 49

5.3.2 Asymptotic Properties of $\widehat{J}_E(z)$ 51

5.4 Confidence Intervals for the Youden Index and Simulation Studies 52
5.4.1 Confidence Intervals for the Covariate-adjusted Youden Index 52
5.4.2 Simulation Studies 54
5.5 A Real Application 56

Chapter 6 CONCLUSIONS AND FUTURE WORK 60

REFERENCES 62

APPENDICES 67

Appendix A SIMULATION RESULTS FOR CHAPTER TWO 67

Appendix B SIMULATION RESULTS FOR CHAPTER THREE 76

Appendix C SIMULATION RESULTS FOR CHAPTER FOUR 83

Appendix D PROOF FOR CHAPTER FOUR 87

Appendix E PROOF FOR CHAPTER FIVE 90

Appendix F SIMULATION RESULTS FOR CHAPTER FIVE 95

LIST OF FIGURES

Figure2.1 Diagnostic Curve Example 10

Figure5.1 The scatter plot of OGTT test vs. Age, left for cases, right for controls. Solid lines are local polynomial estimates for the mean functions. 57

Figure5.2 Nonparametric estimates for $J(Age)$: $\widehat{J}_N$ (solid), $\widehat{J}_E$ (dashed), and $\widehat{J}_{AC}$ (dot-dash). Point-wise confidence bands for $J(Age)$: BP band (dashed), ACNA band (dotdash). 59

FigureA.1 Normal-Normal, Maximum of DC is 0.60. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability. 68

FigureA.2 Normal-Normal, Maximum of DC is 0.80. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability. 69

FigureA.3 Normal-Normal, Maximum of DC is 0.90. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability. 70

FigureA.4 Normal-Normal, Maximum of DC is 0.95. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability. 71

FigureA.5 Normal-T, Maximum of DC is 0.60. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability. 72

FigureA.6 Normal-T, Maximum of DC is 0.80. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability. 73

FigureA.7 Normal-T, Maximum of DC is 0.90. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability. 74

FigureA.8 Normal-T, Maximum of DC is 0.95. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability. 75

FigureF.1 The MSE's of the estimators when ϵ_1 and ϵ_2 follow the standard normal distribution: solid line for $\widehat{J}_N$, dashed line for $\widehat{J}_E$, and dotdash line for $\widehat{J}_{AC}$. 95

FigureF.2 The MSE's of the estimators when ϵ_1 and ϵ_2 follow t-distribution with degree of freedom 4: solid line for $\widehat{J}_N$, dashed line for $\widehat{J}_E$, and dotdash line for $\widehat{J}_{AC}$. 96

FigureF.3 ϵ_1 and ϵ_2 follow the standard normal distribution. The point-wise confidence bands for $J(z)$: NA band (dashed), BP band (dotted), and ACNA band (dotdash). Solid line is the curve for the true values of $J(z)$. 97

FigureF.4 ϵ_1 and ϵ_2 follow t-distribution with degree of freedom 4. The point-wise confidence bands for $J(z)$: NA band (dashed), BP band (dotted), and ACNA band (dotdash). Solid line is the curve for the true values of $J(z)$. 98

LIST OF TABLES

Table4.1 Parameter Setting for Chapter 4 42

TableB.1 The coverage probabilities and average interval lengths of 95% Confidence Intervals for the difference in paired YIs: Scenario 1. 77

TableB.2 Table (B.1) Continue. 78

TableB.3 The coverage probabilities and average interval lengths of 95% Confidence Intervals for the difference in paired YIs: Scenario 2. 79

TableB.4 Table (B.3) Continue. 80

TableB.5 The coverage probabilities and average interval lengths of 95% Confidence Intervals for the difference in paired YIs: Scenario 3. 81

TableB.6 Table (B.5) Continue. 82

TableB.7 The coverage probabilities and average interval lengths of 95% ECI for the difference in paired YIs under mixture normal setting, for all scenarios. . 82

TableC.1 The coverage probabilities and average interval lengths of 95% Confidence Intervals for J_3. 84

TableC.2 The coverage probabilities and average interval lengths of 95% Confidence Intervals for c_{o1}. 85

TableC.3 The coverage probabilities and average interval lengths of 95% Confidence Intervals for c_{o2}. 86

LIST OF ABBREVIATIONS

- FN - False Negative
- FP - False Positive
- OGTT - Oral Glucose Tolerance Test
- ROC - Receiver Operating Characteristic
- YI - Youden index
- TPR - True Positive Rate
- FNR - False Negative Rate
- TNR - True Negative Rate
- FPR - False Positive Rate
- AUC - Area Under ROC Curve
- YIs - Youden indices
- ECI - Exact Confidence Interval
- GPQs - Generalized Pivotal Quantities
- AC - Agresti and Coull
- ACNA - AC normal approximation
- BPac - Bootstrap Percentile based on AC adjustment
- TCCR - Total Correctly Classified Rate
- AD - Alzheimer's Disease
- MCI - Mild Cognitive Impairment

- VUS - Volume under the ROC Surface
- MLE - Maximum Likelihood Estimate
- TTR - True Transitional Rate
- c.d.f. - Cumulative Distribution Function
- p.d.f. - Probability Distribution Function
- DC - Diagnostic Curve
- MSE - Mean Square Error
- BCa - Bias Corrected and accelerated
- BP - Bootstrap Percentile
- HBML - Hybrid Bootstrap and Maximum Likelihood
- LIL - Law of The Iterated Logarithm

Chapter 1

INTRODUCTION

1.1 Diagnostic Tests and Their Evaluations

Medical diagnosis is a commendable technology of modern medicine services. The progresses of medical science and technology provide new avenues for earlier and more accurate diagnosis of diseases. Screening tests offer the opportunity to identify hazardous or abnormal individuals. Accordingly, physicians could assign preventative therapies to cure the disease, to reduce the loss in social wealth and individual wealth, and to improve personal life qualities. Before new testing procedures become "golden" standards, diagnostic and/or screening procedures have to undergo rigorous evaluations. A reliable diagnostic method must recognize disease status of patients as early as possible to admit treatments without inadvertently categorizing healthy individuals. The discriminating capability of a test to correctly diagnose health and diseased individuals would be referred to as test performance, or test accuracy.

For diagnostic tests with binary outcomes, each person is identified as either non-diseased or diseased. A perfectly accurate diagnostic process would identify all truly disordered individuals as diseased and all healthy individuals as non-diseased. In modern medical practice and research, it is paramount for physicians, or biological scientists to identify biomarkers or body symptoms to build a better diagnostic procedure that could determine the disease status of individuals much earlier, and more accurately. However, we are not yet close to approaching "perfection". Instead, diagnostic errors are common. There are two types of errors: classifying a diseased individual as non-diseased, refereed as the false negative (FN) error, and classifying a non-diseased individual as abnormal, known as the false positive (FP) error. Since inaccuracies exist, it is an essential role for statisticians to evaluate the accuracy of diagnostic tests.

Test outcomes could either be categorical, discrete, or continuous. In this dissertation, we focus on continuous test results since the majority of diagnostic procedure outcomes are continuous (Shapiro [52]). For instance, in diabetes diagnoses, the oral glucose tolerance test (OGTT) results

are the measure of glucose levels in human blood, a continuous outcome. In prostate cancer studies, after the confirmation of prostate cancer, the next concern for the patient is whether the cancer has spread to the neighboring lymph nodes. The detection of the spread of cancer is particularly crucial in patient management because a complete cure is often likely for patients in the early stages of prostate cancer. A new method with extremely small risk is used to detect the spread of the cancer by testing the "acid phosphatase level in blood serum", and this outcome is continuous (Aoki *et al.* [3]).

In recent decades, researchers have developed numerous statistical methodologies to evaluate the performance of diagnostic tests. Common measurements in literature for the accuracy of diagnostic tests are:

- Sensitivity and Specificity.
- Receiver Operating Characteristic (ROC) curve.
- The area under the ROC curve (AUC).
- Youden index (YI).

For convenience, without loss of generality, we assume that higher test values indicate higher probability of the disorder throughout this dissertation. The sensitivity, known as the true positive rate (TPR), is defined as the probability that the diseased subject test result is larger than a threshold or cut-off point c. The relative error rate or false negative rate (FNR) is defined as $FNR = 1 - TPR$. The specificity, known as the true negative rate (TNR), is defined as the probability that the non-diseased subject test result is smaller than c. The corresponding error rate is false positive rate (FPR), and $FPR = 1 - TNR$. Often, sensitivity and specificity are used to evaluate the test's capabilities of classifying diseased and non-diseased individuals at a given threshold. One valuable feature of sensitivity and specificity is that they are independent of the disease prevalence. For different diseases with different prevalence, sensitivities and specificities can be utilized. However, evaluating sensitivity and/or specificity at a given threshold value yields information only at that threshold, so only partial information of the performance power can be assessed.

In recent decades, the ROC curve has been given extensive attention in studies of assessments for diagnostic tests, practically and academically. The ROC curve is the plot of the sensitivity versus 1-specificity for all possible thresholds. It comprehensively describes the intrinsic diagnostic capability and provides a visual demonstration of the concession between TPR and FPR. The AUC is a summary measure for the ROC curve. It evaluates the overall discriminative power of a test. In general, the larger AUC (when ROC curve is closer to the upper-left corner) is, the higher the distinguishing power of the test is. Nonetheless, AUC only measures the overall accuracy of the test and explains nothing regarding individual parameters, such as sensitivity and specificity. Furthermore, AUC analysis does not provide relevant information concerning the optimal threshold or optimal cut-off point, another crucial issue in practice.

When evaluating diagnostic tests, it is essential to inform physicians how accurate those tests are. Equally important, information about the optimal threshold should also be supplied for the implementation of tests. According to Schisterman and Perkins [48], the optimal threshold for the positive test result of a disease would be the threshold leading to the maximum sum of TPR and TNR. The cut-off point determined by this maximizing procedure is equivalent to minimization of the misclassification likelihoods, the sum of FNR and FPR. Obtaining the cut-off point by such efforts certainly has the clinically desirable property of maximizing the total correct diagnosis rate and minimizing the overall misdiagnosis rate (Kim [27]).

Youden [66] introduced the following index:

$$J = \max_{c}\{sensitivity(c) + specificity(c) - 1\}, \tag{1.1}$$

$$= sensitivity(c_o) + specificity(c_o) - 1, \tag{1.2}$$

where c_o is the optimal cut-point of test results. This index is also an important numerical summary for the ROC curve. Youden mentioned that the index has several attractive features. For instances, the possible range of the Youden index value is between zero and one, inclusive. A value of zero indicates a totally useless test, and value of one indicates a perfect diagnostic test. The index is independent of the relative sizes of the diseased and non-diseased groups. It is also independent of the absolute sizes of the two groups. All the tests that share the same Youden index value make the same misclassification rates. In a ROC graph, the Youden indx is the maximum vertical distance

between the ROC curve and the diagonal chance line, and it is a comprehensive measurement of the optimal diagnostic capability.

Both AUC and the Youden index are valuable summaries of the performance of diagnostic procedures. However, the Youden index dose not receive equal attention to AUC in literature. Notwithstanding, in medical and biological sciences literature, numerous evaluations for new developed biomarkers and diagnostic procedures build up on the assessing of the Youden index. For instance, Demir *et al.* [10] applied it to identify the most reliable indices in differentiation between thalassemia trait and iron deficiency anemia. Schisterman and Perkins [48] calculated the Youden index to evaluate the performance of the Coronary Calcium Score, a biomarker for atherosclerosis. Also, Youden indices (YIs) could be found to compare the accuracy of different diagnostic procedures. For example, Hawass [19] used the Youden indices to compare the abilities of diagnostic tests. Castle *et al.* [6] applied Youden indices to contrast prototype hybrid capture 3 and hybrid capture 2 human papilloma-virus DNA assays for diagnosing high-grade cervical intra-epithelial neoplasia and cancer. Yerli *et al.* [65] used Youden indices to compare the two methods for diagnosing common parotid tumors: magnetic resonance imaging including diffusion-weighted imaging and fine-needle aspiration cytology.

Evidently, the Youden index has meritorious significance in practice. Therefore, the topics related to the Youden index need further and insightful studies. Evaluations of diagnostic tests via the Youden index could be based on either the point estimate or the confidence interval estimate. Practically, the confidence interval estimate, which offers a confidence range of the true parameter, is more valuable than a point estimate. Here, we mainly focus on the confidence interval estimation of the Youden index and its related topics.

1.2 Estimation for a Youden Index

The Youden index is a function of sensitivity and specificity that depends on the underlying distributions of the diseased and non-diseased populations. When the underlying distributions belong to a particular parametric family such as binormal distributions, Fluss *et al.* [14] provided parametric point estimate for the Youden index. Schisterman and Perkins [48] provided parametric confidence interval estimate for the index based on Delta method (Shao [51]) for the index and

offered nonparametric approaches. Most recently, Lai *et al.* [28] constructed the exact confidence interval (ECI) for the Youden index via generalized pivotal quantities (GPQs, see Weerahandi [60]) based on normal assumptions. Without parametric assumptions for the underlying distributions, Hsieh and Turnbull [22] studied the nonparametric point estimates for the index based on the empirical and the kernel estimates for the underlying distributions. They provided asymptotic properties of the two estimates; nonetheless, the asymptotic variance for the empirical estimate of the Youden index is still mysterious, thus confidence intervals for the Youden index can not be directly constructed. A few studies (e.g., Faraggi [13]) have considered constructing non-parametric confidence intervals for the Youden index and the corresponding cutoff point. Zhou and Qin [67] focused on construction of non-parametric confidence intervals for the Youden index and proposed two new non-parametric intervals for the index based on Agresti and Coull's [1] adjusted estimate (AC adjustment) for a binomial proportion. One interval, was constructed based on the asymptotic normality assumption of the adjusted estimate for the Youden index (ACNA interval), and the other interval was a bootstrap percentile interval based on AC adjustment estimate for the Youden index (BPac interval).

The parametric confidence intervals may be sensitive to departures from the distributional assumptions and can only provide a limited range of distributional forms. The nonparametric intervals may perform worse than parametric methods regarding to average coverage probability and average interval length, although they are robust.

Considering the concession between performance and robustness, in chapter 2, we introduce the diagnostic curve (DC), a supplementary device for the Youden index, as a summary of diagnostic accuracy. The value of $sensitivity(c) + specificity(c) - 1$ represents the total correctly classified rate (TCCR) at given threshold c. The diagnostic curve plots total correctly classified rate vs. the corresponding threshold. The methods introduced to analyze the diagnostic curve are robust and can be easily implemented. Some simulation studies are conducted to compare the proposed nonparametric methods. In chapter 2, we also summarize the existing inferential methods for the Youden index as the preliminaries for chapters 3, 4, and 5.

1.3 Comparison Between Two Diagnostic Tests

Rather than merely independently evaluate different biomarkers, comparison of the accuracies of different biomarkers is one of the major targets in diagnostic experiments. Clinical studies frequently apply matched-pair design to evaluate two diagnostic tests in the same subjects, based on the goal of reducing variation between subjects. Commonly used measurements for the contrast of diagnostic tests are specificity, sensitivity, effectiveness, and AUC. For instances, Wieand *et al.* [61] proposed to compare paired test outcomes based on sensitivity/specificity at a fixed level of specificity/sensitivity. Qin *et al.* [47] constructed new confidence intervals for the difference between two sensitivities at fixed specificities. Hsueh *et al.* [23] and Liu *et al.* [34] proposed methods to evaluate the difference in effectiveness. Tang *et al.* [55] proposed methodologies to assess the ratio of proportions of correct diagnosis. Li *et al.* [30] applied the difference in paired AUCs to investigate the contrast of two diagnostic tests.

Meanwhile, difference between Youden indices for two diagnostic procedures can also be applied to compare diagnostic tests' accuracy. However, to the author's knowledge, this topic has not been sufficiently explored. In chapter 3, we shall focus on constructing confidence intervals for the difference between paired Youden indices. According to the previous research, we construct two parametric intervals, which are the exact confidence interval and the normal approximation based interval. For comparison, we also introduce a nonparametric bootstrap interval. Abundant simulation studies are conducted to evaluate the performances of the proposed methods. Finally, we apply the proposed methods to analyze a real data.

1.4 Evaluation of the Diagnostic Test Classifying Three Ordinal Groups

With highly developed medical technologies, for some diseases, a transitional stage, could be detected and defined. For instance, the "famous" Alzheimer's Disease (AD), commonly existed in senior human population, is a serious brain function disordered malum and increases broad social burden. After years of pathological and clinical researches (in Braak and Braak [4], and Mckeel *et al.* [35]), an intermediate stage, the mild cognitive impairment (MCI) stage, can be detected. For any disease, regardless of its irreversible or reversible attribute, it is necessary to recognize the intermediate stage if exists, to gain the optimal timing window for medical interventions. Through

such timely treatments, we can reduce the loss of social wealth and the spending on the disease, and improve the life qualities of the patients. Therefore, diagnostic tests, which can identify the three stages (non-diseased/intermediate/diseased), is valuable, and the evaluating methodologies for such tests are necessary.

To classify subjects into three groups, two thresholds, saying c_1 and c_2 ($c_2 \geq c_1$), are necessary. Under this situation, redefine the sensitivity (TPR) to be the probability of the test results is larger than c_2 for disease subjects, and redefine specificity (TNR) to be the probability of the test results is smaller than c_1 for non-disease subjects. For the intermediate subjects, we could define the true transitional rate (TTR) as the probability that the test results is between c_1 and c_2.

Mossman [38], first defined a polyhedral ROC surface for discrete outcomes and introduced the volume under the ROC surface (VUS) as a measurement of the diagnostic accuracy in the three groups setting. More recently, Xiong et al. [62] proposed the ROC surface for continuous test results, and suggested using the full VUS under the surface to be the global summary for the accuracy of the test. Under parametric assumptions, they developed the maximum likelihood estimate (MLE) for the VUS.

Few literatures discussed the extension of the Youden index to three diagnostic groups case. Nakas *et al.* [39] extended the Youden index for three ordinal groups and defined it to be the maximum of $TPR + TNR + TTR$. They proposed parametric and nonparametric point estimates for Youden index and mainly focused on evaluating the precision and accuracy of the point estimates. But they did not introduce confidence interval for the Youden index. In chapter 4, we shall focus on making inferences for the Youden index under three ordinal groups. We alternatively define the Youden index to be the maximum of $\frac{1}{2}(TNR+TPR+TTR-1)$. This definition is a nature extension of the original Youden index. We fulfill this topic via proposing parametric and non-parametric point estimates, exploring the nonparametric asymptotic properties, and building the confidence intervals. Some simulation studies are conducted to evaluate the performances of proposed interval estimates.

1.5 Covatiates Adjustment for ROC Related Indices

In plenty situations, covariates information is available and could affect the accuracy of the tests. For example, a medical screening may be a potent detecting device of a disease, but patient characteristics, such as age, gender, or race, often impact test results. It may be that the definition of testing positive (or negative) should depend on covariates, or it may be that the accuracy of the test is less than optimal in certain settings (see Pepe [44]). Covariate-adjustment for the summary measures of the ROC curve has thus become indispensable in many diagnostic applications. Tosteson and Begg [59] and Toledano and Gatsonis [58] used a latent variable ordinal regression to model the distribution of the test results in the diseased and non-diseased populations. Thompson and Zucchini [57] and Obuchowski [40] calculated the ROC curve and AUC for a number of distinct combinations of covariates and then applied a general regression model. Pepe [41], [43], Dodd and Pepe [11] proposed a general regression framework and semi-parametric methods to model the dependence of the ROC curve and AUC on the covariates. Zhou *et al.* [68] and Pepe [44] gave a delightful introduction to why and how to adjust for covariate effects for ROC curves and a comprehensive review of the existing methods in estimating a covariate-specific ROC curve.

While extensive studies focused on covariates adjustment for the ROC curve and the AUC can be found in literature, not much effort focused on the Youden index. Faraggi [13] and Schisterman *et al.* [50] used regression models under normal assumption to adjust the AUC and Youden index for covariates. Yao *et al.* [64] generalized the approaches of Faraggi [13] and Schisterman *et al.* [50] to establish a covariate-adjusted Mann-Whitney estimator for the AUC. They also mentioned that the methods can be extended to other measures related to ROC curves, but they did not provide details. In chapter 5, under the heteroscedastic regression models for the test results, we study the asymptotic properties for the Youden index at given covariates value with and without normal error assumptions. Comprehensive simulation studies are conducted to assess the proposed methodologies. Lastly, we utilize the proposed methods on a real data.

Chapter 2

DIAGNOSTIC CURVE AND INFERENCES FOR A YOUDEN INDEX

2.1 Introduction

For a diagnostic test only identifying diseased and non-diseased groups, let X and Y be the test results of a continuous-scale diagnostic test for health and diseased subjects, respectively. Without loss of generality, assume that X and Y are independent. Denote $X_1, X_2, \ldots, X_{n_-}$ be a random sample from non-diseased people, and $Y_1, Y_2, \ldots, Y_{n_+}$ be a random sample from diseased patients. Let F_- and F_+ be the cumulative distribution functions (c.d.f.), and f_- and f_+ be the corresponding probability density functions (p.d.f.) for the non-diseased and diseased populations, respectively. The Youden index can be expressed as:

$$J = \max_c \{F_-(c) - F_+(c)\} \tag{2.1}$$

$$= F_-(c_o) - F_+(c_o), \tag{2.2}$$

where c_o stands for the optimal threshold and its uniqueness depends on F_- and F_+ .

As mentioned in section 1.2, the point estimate and interval estimate for J can be determined via parametric or nonparametric approaches. Parametric methods performs well; however, they struggle with the validation of parametric distribution assumptions. Nonparametric methods are robust. However, without the information of the underlying distributions, the finite sample performance may be worse than parametric methods. One of the main difficulties of the nonparametric approaches is that the asymptotic variance of the nonparametric estimator for J is still unknown due to the lack of a closed form of the optimal threshold. Creating a balance between performance and robustness, and avoiding the estimation of the optimal threshold, we introduce the concept of the diagnostic curve to evaluate diagnostic tests.

2.2 Diagnostic Curve

2.2.1 Definition

The Youden index represents the maximum total correctly classified rate, which is achieved at the optimal threshold, of a diagnostic test. Ignoring the maximization procedure in the definition (1.1), we could define the diagnostic curve (DC) as follows:

$$J(c) = sensitivity(c) + specificity(c) - 1 \tag{2.3}$$

$$= F_{-}(c) - F_{+}(c). \tag{2.4}$$

$J(c)$ measures the diagnostic accuracy of the test at any given cut-off value c. The diagnostic curve displays the relationship between the accuracy of a test and its cut-off point. The figure below is a plot for diagnostic curve. This diagnostic curve is generated from normal distributions for the non-diseased and diseased population with equal standard deviations 1. The means of the non-diseased and diseased are selected to be 0 and 2.563104, respectively, to achieve the maximum of the curve is 0.8. From this graph, we can view the overall trend of the variation in the accuracy depends on

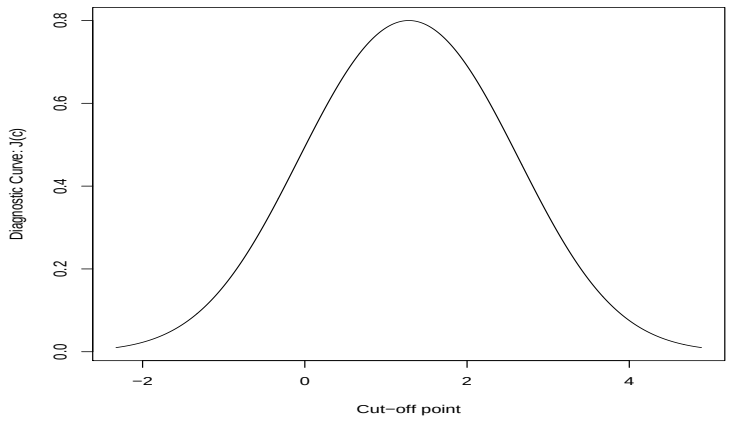

Figure 2.1. Diagnostic Curve Example

the cut-off values. The maximum of this curve represents the maximum accuracy that the test can

achieve, and it is equivalent to the Youden index. Thus, the diagnostic curve plays an analogous role as the Youden index.

Notice that, at a given c, (2.4) is the difference between two binomial proportions. Re-express (2.4) for convenience as:

$$J(c) \quad = \quad p_X(c) - p_Y(c), \tag{2.5}$$

where $p_X(c)$ denotes the true negative rate (TNR), and $p_Y(c)$ denotes the false negative rate (FNR), let $q_X(c) = 1 - p_X(c)$ and $q_Y(c) = 1 - p_Y(c)$ stand for the false positive rate (FPR) and true positive rate (TPR) at given c, respectively.

2.2.2 Point-wise Confidence Intervals for the Diagnostic Curve

Without assuming any parametric form for the underlying distributions of the test results, we offer an ultimately nonparametric evaluation for the test. For a given cut-off c, we have the empirical estimate of $J(c)$

$$\widehat{J}_E(c) = \widehat{p}_X(c) - \widehat{p}_Y(c), \tag{2.6}$$

where $\widehat{p}_X(c) = \frac{1}{n_-}\sum_{i=1}^{n_-} I(X_i \leq c)$ and $\widehat{p}_Y(c) = \frac{1}{n_+}\sum_{j=1}^{n_+} I(Y_j \leq c)$, and $I(\cdot)$ is the indicator function. Since $J(c)$ is the difference between two proportions, we can apply some existing methods to construct the point-wise confidence intervals for $J(c)$.

Zhou *et al.* [70] developed two new confidence intervals for the difference between two binomial proportions, via adjusting the skewness based on an Edgeworth expansion directly and indirectly. Based on their simulation study, we decide to choose the confidence interval constructed in the indirect way, called the TT interval. To introduce this interval, we list the equations for the

following quantities at given threshold below.

$$\begin{aligned}
\delta(c) &= \left(\frac{n_- + n_+}{n_-}\right)^2 p_X(c)q_X(c)(1-2p_X(c)) - \left(\frac{n_- + n_+}{n_+}\right)^2 p_Y(c)q_Y(c)(1-2p_Y(c)) \quad (2.7)\\
\sigma(c) &= \left(\frac{n_- + n_+}{n_-}p_X(c)q_X(c) + \frac{n_- + n_+}{n_+}p_Y(c)q_Y(c)\right)^{1/2}, \qquad o(c) = \frac{\delta(c)}{6\sigma^2(c)},\\
s(c) &= \frac{(n_- + n_+)(1-2p_X(c))}{2n_-} - o(c).
\end{aligned}$$

Then the $(1-\alpha)100\%$ TT confidence interval for $J(c)$ at a given c, denoted by $(L_{TT}(c), U_{TT}(c))$, can be expressed as:

$$\begin{aligned}
L_{TT}(c) &= \widehat{J}_E(c) - \left(\frac{\widehat{p}_X(c)\widehat{q}_X(c)}{n_-} + \frac{\widehat{p}_Y(c)\widehat{q}_Y(c)}{n_+}\right)^{1/2} g^{-1}(z_{1-\alpha/2}), \quad (2.8)\\
U_{TT}(c) &= \widehat{J}_E(c) - \left(\frac{\widehat{p}_X(c)\widehat{q}_X(c)}{n_-} + \frac{\widehat{p}_Y(c)\widehat{q}_Y(c)}{n_+}\right)^{1/2} g^{-1}(z_{\alpha/2}),
\end{aligned}$$

where $g^{-1}(t) = (m+n)^{1/2}(\widehat{s}(c)\widehat{\sigma}(c))^{-1}\{(1+3(\widehat{s}(c)\widehat{\sigma}(c))((m+n)^{-1/2}t - (m+n)^{-1}\widehat{o}(c)\widehat{\sigma}(c)))^{1/3} - 1\}$. $z_{1-\alpha/2}$ and $z_{\alpha/2}$ are normal quantiles.

The second interval, denoted by $(L_{AC}(c), U_{AC}(c))$, is the AC interval proposed by Agresti and Caffo [2]. It is easy to implement and is one of best intervals for the difference of two binomial proportions (Zhou *et al.* [70]). Let $\widehat{p}$ represents the original sample proportion for a single binomial proportion p, namely, $\widehat{p}$ =(number of success)/(sample size). The original Agresti and Coull-adjusted (AC-adjusted) sample proportion for one sample case (Agresti and Coull [1]), $\tilde{p}^o$, can be obtained by adding two success and two failure to the original sample, namely $\tilde{p}^o$ =(number of success+2)/(sample size+4). According to this modification, $\tilde{p}^o$ shrink the sample proportion to 0.5, and of course when sample size increases, this shrinkage is much less influential. They also showed that $\tilde{p}^o$ is equivalent to the Bayesian estimate for the binomial proportion with a Beta prior distribution having both shape parameters 2. Furthermore, the adjusted sample proportion has a smaller mean square error (MSE) than the original sample proportion. The AC-adjusted confidence interval for one binomial proportion has the same form as the classical Wald interval; however, $\widehat{p}$ is replaced by $\tilde{p}^o$ in the formula and the updated size is used. Agresti and Coull [1] showed that the AC-adjusted interval for one binomial proportion p has outstanding performances. Later, Agresti and Caffo [2] pointed out that the strategy of adding 1 success and 1 failure to

each of the independent original samples works well when compare two binomial proportions from two independent samples. Based on this, we state the AC confidence interval at a given c for the diagnostic curve below:

$$\begin{aligned} L_{AC}(c) &= \tilde{J}_E(c) - z_{1-\alpha/2}\sqrt{\tilde{p}_X(c)\tilde{q}_X(c)/(n_- + 2) + \tilde{p}_Y(c)\tilde{q}_Y(c)/(n_+ + 2)}, \\ U_{AC}(c) &= \tilde{J}_E(c) + z_{1-\alpha/2}\sqrt{\tilde{p}_X(c)\tilde{q}_X(c)/(n_- + 2) + \tilde{p}_Y(c)\tilde{q}_Y(c)/(n_+ + 2)}, \end{aligned} \tag{2.9}$$

where $\tilde{p}_X(c) = (\sum_{i=1}^{n_-} I(X_i \leq c) + 1)/(n_- + 2)$ and $\tilde{q}_X(c) = 1 - \tilde{p}_X(c)$. $\tilde{p}_Y(c)$ and $\tilde{q}_Y(c)$ are similarly defined, and $\tilde{J}_E(c) = \tilde{p}_X(c) - \tilde{p}_Y(c)$.

Lastly, we apply the traditional Wald type confidence interval for comparison. At a fixed c, the Wald interval is $(L_{Wald}(c), U_{Wald}(c))$, where

$$\begin{aligned} L_{wald}(c) &= \widehat{J}_E(c) - z_{1-\alpha/2}\sqrt{\widehat{p}_X(c)\widehat{q}_X(c)/n_- + \widehat{p}_Y(c)\widehat{q}_Y(c)/n_+}, \\ U_{wald}(c) &= \widehat{J}_E(c) + z_{1-\alpha/2}\sqrt{\widehat{p}_X(c)\widehat{q}_X(c)/n_- + \widehat{p}_Y(c)\widehat{q}_Y(c)/n_+}. \end{aligned} \tag{2.10}$$

2.2.3 Simulation Studies

To evaluate the performances of the TT, AC and Wald intervals for $J(c)$, we simulate two scenarios. In scenario 1, the underlying distributions for the non-diseased and diseased populations are both normal. In scenario 2, the underlying distribution for the non-diseased population is normal, but the distribution for the diseased is the student-t distribution with degree of freedom 4. Without loss of generality, we specify the mean and standard deviation for the healthy population as 0 and 1, respectively. We allow the maximum values of diagnostic curve to be 0.6, 0.8, 0.9, and 0.95 for each scenario. Sample sizes are selected to be $(50, 50)$ in these simulation studies. Finally, a reasonable range for the possible cut-off points is determined to be from the 20-th percentile of F_- to the 80-th percentile of F_+ for both scenarios.

From the simulation results presented in Appendix A, we conclude that the classical Wald interval does not perform as well as the TT and the AC intervals. Some interesting observations from the simulation studies should be pointed out. When the maximum of the curve is 0.9, which means, in the neighborhood of the optimal threshold, the true negative rate is close to 1 and the false negative rate is close to 0, the TT confidence intervals will lose the power of capturing the true

parameter. Moreover, when the maximum of the curve is 0.95, which indicates, in the neighborhood of the optimal threshold, the true negative rate is extremely close to 1 and the false negative rate is also extremely close to 0, the TT interval would over capture the true parameter, and the average coverage probability exceed the nominal level. These facts of the performance for the TT confidence interval are congruent to the simulation results showed in Zhou *et al.* [70] (page 109, Figure 4) and indicate that the TT interval is sensitive to true value of the difference between two binomial proportions. The performance of the AC point-wise confidence interval is more stable. Looking at the average interval length, in the neighborhood of the the maximum point on the diagnostic curve, the AC confidence interval requires a longer length to capture the true difference between two binomial proportions than the TT confidence interval does.

In summary, we would recommend to apply either the TT point-wise confidence interval or the AC point-wise confidence interval for the diagnostic curve when the maximum of the diagnostic curve is not close to 1. Otherwise, the AC point-wise confidence interval is preferred.

2.3 Existing Methods for Inferences of a Youden Index

Now, we summarize existing methods for inferences of the Youden index as preliminaries for later chapters.

2.3.1 Point Estimation

The point estimate for the Youden index would either be based on parametric or nonparametric assumptions.

The major parametric point estimate for J and c_o is from Schisterman and Perkins [48], in which they applied normality assumptions for X and Y:

$$X_i \sim N(\mu_x, \sigma_x^2), i = 1, 2, \ldots, n_-; \quad Y_j \sim N(\mu_y, \sigma_y^2), j = 1, 2, \ldots, n_+.$$

With this assumption, we can obtain the closed form of the optimal cut-off point:

$$c_o = \frac{\mu_x(b^2 - 1) - a + b\sqrt{a^2 + (b^2 - 1)\sigma_x ln(b^2)}}{(b^2 - 1)}, \tag{2.11}$$

where $a = \mu_y - \mu_x$ and $b = \sigma_y/\sigma_x$. When the two populations have equal standard deviations, the optimal cut-point will be the average of the two population means, namely, $c_o = \frac{\mu_x+\mu_y}{2}$. The point estimate for the Youden index under normal assumption is:

$$J = \Phi\left(\frac{\mu_y - c_o}{\sigma_y}\right) + \Phi\left(\frac{c_o - \mu_x}{\sigma_x}\right) - 1, \tag{2.12}$$

where $\Phi(\cdot)$ is the cumulative normal distribution function. To estimate the J and c_o, simply plug the maximum likelihood estimates (MLEs), $\widehat{\mu}_x$, $\widehat{\mu}_y$, $\widehat{\sigma}_x$, and $\widehat{\sigma}_y$, of μ_x, μ_y, σ_x, and σ_y into (2.11) and (2.12) to obtain $\widehat{c}_{oML}$ and $\widehat{J}_{ML}$, the MLE for c_o and J. $\widehat{c}_{oML}$ and $\widehat{J}_{ML}$ can be expressed as:

$$\widehat{c}_{oML} = \frac{\widehat{\mu}_x(\widehat{b}^2 - 1) - \widehat{a} + \widehat{b}\sqrt{\widehat{a}^2 + (\widehat{b}^2 - 1)\widehat{\sigma}_x ln(\widehat{b}^2)}}{(\widehat{b}^2 - 1)}, \tag{2.13}$$

where the maximum likelihood estimator $\widehat{\mu}_x = \frac{1}{n_-}\sum_{i=1}^{n_-} X_i$, $\widehat{\sigma}_x = \sqrt{\frac{1}{n_-}\sum_{i=1}^{n_-}(X_i - \widehat{\mu}_x)^2}$, and $\widehat{\mu}_y$ and $\widehat{\sigma}_y$ are similarly defined. Furthermore, $a = \widehat{\mu}_y - \widehat{\mu}_x$ and $b = \widehat{\sigma}_y/\widehat{\sigma}_x$, when the two populations have equal standard deviations, $c_o = \frac{\widehat{\mu}_x+\widehat{\mu}_y}{2}$. Then,

$$\widehat{J}_{ML} = \Phi\left(\frac{\widehat{\mu}_y - \widehat{c}_{oML}}{\widehat{\sigma}_y}\right) + \Phi\left(\frac{\widehat{c}_{oML} - \widehat{\mu}_x}{\widehat{\sigma}_x}\right) - 1. \tag{2.14}$$

When the underlying distributions are other parametric distributions, the parametric MLEs for c_o and J can still be derived. For instance, Schisterman and Perkins [48] assumed that the distributions for the test results are both Gamma distributions. Although there is no closed form for c_o and J, the MLEs still can be found via numerical algorithms, such as Newton-Raphson algorithm.

On the other hand, two main nonparametric point estimates (Hsieh and Turnbull [22]) for J have been developed. One method is based on the empirical distributions $\widehat{F}_-(c)$ and $\widehat{F}_+$ for F_- and F_+, where

$$\widehat{F}_-(c) = \frac{1}{n_-}\sum_{i=1}^{n_-} I(X_i \le c), \text{ and } \widehat{F}_+(c) = \frac{1}{n_+}\sum_{j=1}^{n_+} I(Y_j \le c). \tag{2.15}$$

Another method is based on the kernel smoothed estimates for the underlying distributions of the test results. Let $\tilde{f}_-$ and $\tilde{f}_+$ represent the kernel smoothed density estimates for f_- and f_+,

respectively. $\tilde{f}_-$ and $\tilde{f}_+$ are defined as follows:

$$\tilde{f}_-(c) = \frac{1}{n_-}\sum_{i=1}^{n_-}\frac{1}{h_{n_-}}k(\frac{c-X_i}{h_{n_-}}), \tag{2.16}$$

$$\tilde{f}_+(c) = \frac{1}{n_+}\sum_{j=1}^{n_+}\frac{1}{h_{n_+}}k(\frac{c-Y_j}{h_{n_+}}), \tag{2.17}$$

where $k(\cdot)$ is a well defined kernel function, and h_{n_-} and h_{n_+} are the bandwidths selected for $\tilde{f}_-$ and $\tilde{f}_+$, respectively. Chose some appropriate constant r, we can determine $h_{n_-} = rn_-^{-1/(2v-1)}$ for some $v > 2$, and similar for h_{n_+}. Let $\tilde{F}_-(c)$ and $\tilde{F}_+$ be the kernel smoothed c.d.f. estimates for F_- and F_+, respectively. $\tilde{F}_-(c)$ and $\tilde{F}_+$ are given as:

$$\tilde{F}_-(c) = \frac{1}{n_-}\sum_{i=1}^{n_-}K(\frac{c-X_i}{h_{n_-}}), \quad \tilde{F}_+(c) = \frac{1}{n_+}\sum_{j=1}^{n_+}K(\frac{c-Y_j}{h_{n_+}}). \tag{2.18}$$

$K(\cdot)$ represents the kernel distribution and is defined as $K(t) = \int_{-\infty}^{t} k(x)dx$.

We have the empirical estimate for J

$$\begin{aligned} \widehat{J}_E &= \max_c\{\widehat{F}_-(c) - \widehat{F}_+(c)\} && (2.19)\\ &= \widehat{F}_-(\widehat{c}_{oE}) - \widehat{F}_+(\widehat{c}_{oE}), && (2.20) \end{aligned}$$

where $\widehat{c}_{oE}$ is the empirical estimate for c_o. Hsieh and Turnbull [22] defined $\widehat{c}_{oE}$ as:

$$\widehat{c}_{oE} = median\{t_0|\widehat{F}_-(t_0) - \widehat{F}_+(t_0) = \max_t\Big(\widehat{F}_-(t) - \widehat{F}_+(t)\Big)\}. \tag{2.21}$$

The kernel smoothed estimate for J can be expressed as:

$$\begin{aligned} \widehat{J}_K &= \max_c\{\tilde{F}_-(c) - \tilde{F}_+(c)\} && (2.22)\\ &= \tilde{F}_-(\widehat{c}_{oK}) - \tilde{F}_+(\widehat{c}_{oK}), && (2.23) \end{aligned}$$

where $\widehat{c}_{oK}$ represents the kernel smoothed estimate for the optimal threshold. Under some regulations, Hsieh and Turnbull [22] defined $\widehat{c}_{oK}$ as:

$$\widehat{c}_{oK} \quad = \quad median\{t_0 | t_0 \in (i_l, i_u), \text{and} \tilde{f}_-(t_0) = \tilde{f}_+(t_0)\}, \tag{2.24}$$

where the interval (i_l, i_u) was assumed to be a sub-interval in the intersection of the supports of F_- and F_+.

Hsieh and Turnbull [22] studied the nonparametric estimates for J and c_o. Under some mild conditions, they proved the strong consistence of $\widehat{J}_E$, $\widehat{J}_K$, $\widehat{c}_{oE}$, and $\widehat{c}_{oK}$. Further, if $n_-/n_+ \to \lambda^2 > 0$, they proved that $\sqrt{n_-}(\widehat{J}_E - J)$ converges in distribution to a linear combination of two independent Brownian bridges, which is $\lambda^{-1}B_1(F_-(c_o)) - B_2(F_+(c_o))$, where $B(\cdot)$ stands for Brownian bridges on interval $[0, 1]$. They also stated that $\widehat{J}_K$ has a smaller mean square error (MSE) than dose $\widehat{J}_E$.

2.3.2 Confidence Intervals for the Youden index

Confidence intervals for the Youden index are of particular interest in this dissertation. In the literature, a few studies focused on constructing confidence intervals for the Youden index, utilizing both parametric and nonparametric methods.

With normality assumptions for the distribution of X and Y, Lai *et al.* [28] constructed the exact confidence interval (ECI) via generalized pivotal quantities (GPQs) for J and c_o. Detailed introductions to GPQ can be found in Weerahandi [60] and Hanning *et al.* [18]. We cite the definition of GPQ from Li *et al.* [30] as follows:

"Let S be a random variable with distribution $F_s(\theta; \eta)$, where θ is a parameter of interest and η is a vector of nuisance parameters. Let s be an observed value of S and $Q = Q(S; s; \theta; \eta)$ be a function of $(S; s; \theta; \eta)$. The random quantity Q is said to be a GPQ if it satisfies the following two conditions:

(a) The distribution of Q does not depend on any unknown parameters.

(b) The observed value of Q, say $q = Q(s; s; \theta; \eta)$, is free of the nuisance parameter vector η."

Under similar parametric assumptions, Schisterman and Perkins [48] provided Delta confidence intervals for J and c_o. The $(1-\alpha)100\%$ Delta confidence interval for J is $\widehat{J}_{ML} \pm z_{1-\alpha/2}\sqrt{\widehat{Var}(\widehat{J}_{ML})}$, where $\widehat{Var}(\widehat{J}_{ML})$ is the estimate for $Var(\widehat{J}_{ML})$ via plugging in the MLEs of populations means and

standard deviations. $Var(\widehat{J}_{ML})$ can be derived by the Delta method and can be expressed as follow:

$$\begin{aligned} Var(\widehat{J}_{ML}) \approx & (\frac{\partial J}{\partial \mu_x})^2 Var(\widehat{\mu}_x) + (\frac{\partial J}{\partial \sigma_x})^2 Var(\widehat{\sigma}_x) + \\ & (\frac{\partial J}{\partial \mu_y})^2 Var(\widehat{\mu}_y) + (\frac{\partial J}{\partial \sigma_x})^2 Var(\widehat{\sigma}_y), \end{aligned} \tag{2.25}$$

where ∂ represents the partial differential operating. It is well known that $Var(\widehat{\mu}_x) = \frac{\sigma_x^2}{n_-}$ and $Var(\widehat{\mu}_y) = \frac{\sigma_y^2}{n_+}$. Schisterman and Perkins [48] stated the approximate estimate for $Var(\widehat{\sigma}_x)$ and $Var(\widehat{\sigma}_y)$ to be $\frac{\sigma_x^2}{2(n_- -1)}$ and $\frac{\sigma_y^2}{2(n_+ -1)}$, respectively. Similarly, the Delta interval can be constructed for c_o. It should be pointed out that the derivation of the approximate variance via Delta method is complex. Alternatively, the parametric bootstrap procedure can be used to estimate the variance of the proposed MLE, and the resulting bootstrap based confidence interval is much easier to construct than the Delta interval.

In a nonparametric setting, Schisterman and Perkins [48] introduced the Bias Corrected and accelerated (BCa) and the Bootstrap Percentile (BP) intervals for J and c_o. They have specifically explained the procedure for constructing BP interval. The construction of the BCa interval are referred to Carpenter and Bithell [5] or Yang *et al.* [63].

Inspired by Agresti and Coull's [1] adjusted estimate for a binomial proportion, Zhou and Qin [67] proposed using the AC-adjusted confidence interval for J. Although the Youden index can be interpreted as the difference between two binomial proportions; these two binomial proportions in the Youden index are dependent since the estimation of the optimal threshold is involved. Unlike the application of Agresti and Caffo [2] for the point-wise confidence interval to the diagnostic curve, we use the original AC-adjusted binomial proportions for the Youden index. Define the AC-adjusted empirical estimate for J to be

$$\widehat{J}_{AC} = \max_c \{\frac{1}{n_- + 4}(\sum_{i=1}^{n_-} I(X_i \le c) + 2) - \frac{1}{n_+ + 4}(\sum_{j=1}^{n_+} I(Y_j \le c) + 2)\}. \tag{2.26}$$

By Theorem 3.1 in Hsieh and Turnbull [22], $\widehat{J}_{AC}$ is asymptotically normal at given c. However, the asymptotic variance of $\widehat{J}_{AC}$ is still unknown. Therefore, a bootstrap is needed to estimate the variance of $\widehat{J}_{AC}$. Re-sampling the original observations B times, we can obtain B bootstrap replications $\{\widehat{J}_{AC}^{*b} : b = 1, 2, \ldots, B\}$. Then, the Agresti and Coull normal approximation (ACNA)

interval is defined as follows

$$\left(\bar{J}^*_{AC} - z_{1-\alpha/2}\sqrt{Var(\widehat{J}^*_{AC})},\ \bar{J}^*_{AC} + z_{1-\alpha/2}\sqrt{Var(\widehat{J}^*_{AC})}\right),$$

where $\bar{J}^*_{AC} = \frac{1}{B}\sum_{b=1}^{B}\widehat{J}^{*b}_{AC}$, and $Var(\widehat{J}^*_{AC}) = \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{J}^{*b}_{AC} - \bar{J}^*_{AC})^2$.

2.3.3 Practical Implications of Existing Methods

Based on the simulation studies in the papers mentioned, we make the following conclusions concerning choice of methodologies.

In practice, for the point estimate of J or c_o, if the parametric assumptions are believed to be valid, we would recommend efficient and accurate parametric approach. Otherwise, we prefer the empirical estimate rather than kernel smoothed estimate because the selection of the kernel function and optimal bandwidth is sometimes questionable.

Considering the confidence interval estimation, armed with normal assumptions, the exact confidence interval outperforms all other methods, especially when the sample size is extremely small. Nonetheless, as long as the parametric MLEs of J or c_o can be found, the Delta interval can be applied. If the underlying distributions are unknown, we highly favor the nonparametric ACNA interval.

Chapter 3

INFERENCES FOR THE DIFFERENCE OF TWO YOUDEN INDICES

3.1 Introduction

In this chapter we focus on constructing the confidence intervals for the difference between two Youden indices in paired study design. Under the normality assumption for the distributions of continuous test results for diseased and non-diseased populations, based on the asymptotic property of the maximum likelihood estimate (MLE), confidence interval for the differences of paired Youden indices can be constructed via the asymptotic normality property of the MLE, and the corresponding variance estimation can be implemented by Delta method. However, this MLE-based confidence interval suffers from obtaining the complicated expression for the asymptotic variance via Delta method, due to the correlations between the paired tests on the same individual. Recently, the exact confidence regions for the parameter vector containing nuisance parameters based on generalized pivotal quantities (GPQs, see Weerahandi [60]) presented exceptionally small sample performances in terms of coverage probability. Substantial applications of GPQs can be found, for instances, in the bioequivalence research (McNally *et al.* [36]), the inter-laboratory trials research (Iyer *et al.* [25]), the construction of tolerance intervals (Liao and Iyer [31] and Liao *et al.* [32]), and the multivariate analysis of variance (Gamage *et al.* [16]). In diagnostic studies, Li *et al.* [30] established the ECI for the difference in paired AUCs via GPQs. Lai *et al.* [28] proposed GPQ based exact confidence interval (ECI) for one Youden index. Motivated by these research works, the primary goal in this chapter is to develop a GPQ-based ECI for the difference in paired Youden indices. Further, we propose an MLE and bootstrap hybrid based confidence interval and a nonparametric BP interval as a comparison.

3.2 Bootstrap and Maximum Likelihood Based Intervals

Matched-pair design is frequently applied to evaluate two diagnostic tests in the same subjects for the purpose of reducing variation between subjects. Suppose we have two continuous-scale

diagnostic procedures Test 1 and Test 2. Let $\mathbf{X} = (X_1, X_2)$ and $\mathbf{Y} = (Y_1, Y_2)$ be two numerical vectors representing the test results for the non-diseased and diseased populations, respectively. We also assume that $\mathbf{X}$ and $\mathbf{Y}$ are independent. Let F_{-1}, F_{-2}, F_{+1} and F_{+2} are the respective marginal distributions of X_1, X_2, Y_1 and Y_2. Then, the Youden index of test Test-i is

$$\begin{aligned} J_i &= \max_c \{P(Y_i \geq c) + P(X_i \leq c) - 1\} \\ &= F_{-i}(c_{oi}) - F_{+i}(c_{oi}), \text{ for } i = 1, 2, \end{aligned} \tag{3.1}$$

where c_{oi} is the optimal cut-off point for the i-th test ($i = 1, 2$). The target parameter, $D = J_1 - J_2$, is the difference between two Youden indices from two paired diagnostic tests.

Let X_{ik} be the result of Test i from the non-diseased population for $k = 1 \cdots n_-$, and $Y_{ik'}$ be the result of Test i from the diseased population for $k' = 1 \cdots n_+$. Without specification of the underlying distributions for the non-diseased and diseased populations, the empirical estimate for J_i is

$$\begin{aligned} \widehat{J}_{iE} &= \frac{\sum_{k=1}^{n_-} I(X_{ik} \leq \widehat{c}_{oi}^E)}{n_-} + \frac{\sum_{k'=1}^{n_+} I(Y_{ik'} \geq \widehat{c}_{oi}^E)}{n_+} - 1 \\ &= \frac{\sum_{k=1}^{n_-} I(X_{ik} \leq \widehat{c}_{oi}^E)}{n_-} - \frac{\sum_{k'=1}^{n_+} I(Y_{ik'} < \widehat{c}_{oi}^E)}{n_+}, \ i = 1, 2, \end{aligned} \tag{3.2}$$

where $I(\cdot)$ is the indicator function, and $\widehat{c}_{oi}^E$ is the empirical estimate for the optimal cut-off point of Test i, which could follow the definition (2.21). Then, the empirical estimate for D is

$$\widehat{D}_E = \widehat{J}_{1E} - \widehat{J}_{2E}. \tag{3.3}$$

The exact distribution of $\widehat{D}_E$ cannot be obtained. Further, the asymptotic distribution of $\widehat{D}_E$ is still unknown in literature when F_- and F_+ are not specified. Therefore, some ad hoc methods, like bootstrap technology, are necessary to construct confidence intervals for D. For example, we can construct a bootstrap percentile interval for D by using the following bootstrap procedures:

1. Draw a bootstrap re-sample of size n_-, X_{ik}^*'s, with replacement from the non-diseased sample X_{ik}'s and a separate re-sample of size n_+, $Y_{ik'}^*$'s, with replacement from the diseased sample $Y_{ik'}$'s, for $i = 1, 2$.

2. Calculate the bootstrap version $\widehat{D}_E^*$ of $\widehat{D}_E$ by using the bootstrap samples.

3. Repeat the first two steps B (it is suggested that $B \geq 200$) times to obtain the set of bootstrap replications $\{\widehat{D}_E^{*b} : b = 1, 2, \ldots, B\}$.

Then, the $(1-\alpha)100\%$ level BP confidence interval for D is:

$$\left(\widehat{D}^*_{E_{\alpha/2}}, \widehat{D}^*_{E_{1-\alpha/2}}\right), \tag{3.4}$$

where $\widehat{D}^*_{E_{\alpha/2}}$ is the $\alpha/2 \times 100$-th percentile of $\{\widehat{D}_E^{*b} : b = 1, 2, \ldots, B\}$, and $\widehat{D}^*_{E_{1-\alpha/2}}$ is the $(1-\alpha/2) \times$ 100-th percentile of $\{\widehat{D}_E^{*b} : b = 1, 2, \ldots, B\}$.

The nonparametric BP interval is easy to implement but it could have serious under-coverage problem for the difference between two Youden indices. To obtain more accurate confidence interval for D, we further assume that

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathbf{N_2}(\mu_x, \mathbf{\Sigma}_x),$$

and

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathbf{N_2}(\mu_y, \mathbf{\Sigma}_y),$$

where $\mu_x = \begin{bmatrix} \mu_{x1} \\ \mu_{x2} \end{bmatrix}$, $\mathbf{\Sigma}_x = \begin{bmatrix} \sigma^2_{x1} & \sigma_{x12} \\ \sigma_{x12} & \sigma^2_{x2} \end{bmatrix}$, $\mu_y = \begin{bmatrix} \mu_{y1} \\ \mu_{y2} \end{bmatrix}$, $\mathbf{\Sigma}_x = \begin{bmatrix} \sigma^2_{y1} & \sigma_{y12} \\ \sigma_{y12} & \sigma^2_{y2} \end{bmatrix}$.

Under above normal distributional assumptions, the Youden indices for the two diagnostic tests can be explicitly expressed as

$$J_i = \Phi\left(\frac{\mu_{yi} - c_{oi}}{\sigma_{yi}}\right) + \Phi\left(\frac{c_{oi} - \mu_{xi}}{\sigma_{xi}}\right) - 1, \ i = 1, 2, \tag{3.5}$$

where $\Phi(\cdot)$ denotes the standard normal c.d.f., and c_{oi} is the optimal cut-off point for test i. If assuming that the mean test result in the diseased population is greater that in the non-diseased population, i.e., $\mu_{yi} > \mu_{xi}$ for $i = 1, 2$, then c_{oi} has the following closed form expression:

$$c_{oi} = \frac{\mu_{xi}(b_i^2 - 1) - a_i + b_i\sqrt{a_i^2 + (b_i^2 - 1)\sigma_{xi} ln(b_i^2)}}{(b_i^2 - 1)},$$

where $a_i = \mu_{yi} - \mu_{xi}$ and $b_i = \sigma_{yi}/\sigma_{xi}$, $i = 1, 2$. Particularly, if $\sigma_{yi} = \sigma_{xi}$, then

$$c_{oi} = \frac{\mu_{xi} + \mu_{yi}}{2}.$$

Therefore, the difference D between the two Youden indices is a function of parameters μ_{x1}, μ_{x2}, σ_{x1}, σ_{x2}, μ_{y1}, μ_{y2}, σ_{y1}, and σ_{y2}. To estimate D, firstly, we find the MLEs of all the parameters, namely, $\widehat{\mu}_{x1}$, $\widehat{\mu}_{x2}$, $\widehat{\sigma}_{x1}$, $\widehat{\sigma}_{x2}$, $\widehat{\mu}_{y1}$, $\widehat{\mu}_{y2}$, $\widehat{\sigma}_{y1}$, and $\widehat{\sigma}_{y2}$. Then, for the i-th $i = 1, 2$ test, we plug in these MLEs into (2.11) to obtain the MLEs of c_{o1} and c_{o2}, denoted by $\widehat{c}_{o1ML}$ and $\widehat{c}_{o2ML}$, respectively. According to (2.14), we can obtain the MLEs for J_1 and J_2 for Test 1 and Test 2, denoted by $\widehat{J}_{1ML}$ and $\widehat{J}_{2ML}$, respectively. Hence, the MLE of D is:

$$\widehat{D}_{ML} \quad = \quad \widehat{J}_{1ML} - \widehat{J}_{2ML}. \tag{3.6}$$

$\widehat{D}_{ML}$ inherits the asymptotic normality from parametric estimators in normal distributions. After finding the asymptotic variance of the estimate via Delta Method (Shao [51]), we could construct a confidence interval for D. However, we do not recommend the Delta method here because the estimation of the asymptotic variance $Var(\widehat{D}_{ML})$ of $\widehat{D}_{ML}$ involves in complicated calculation and estimation of too many unknown parameters. In stead, we propose to utilize the bootstrap procedures mentioned above to obtain the bootstrap replications of $\widehat{D}_{ML}$, denoted as $\{\widehat{D}^{*b}_{ML} : b = 1, 2, \ldots, B\}$. Then, we can calculate the bootstrap variance estimate $Var(\widehat{D}^{*}_{ML})$ of $Var(\widehat{D}_{ML})$. A $(1-\alpha)100\%$ hybrid bootstrap and maximum likelihood based confidence interval (hereafter called HBML confidence interval) for D is proposed as follows:

$$\left(\widehat{D}_{ML} - z_{1-\alpha/2}\sqrt{Var(\widehat{D}^{*}_{ML})}, \widehat{D}_{ML} + z_{1-\alpha/2}\sqrt{Var(\widehat{D}^{*}_{ML})}\right), \tag{3.7}$$

where $Var(\widehat{D}^{*}_{ML}) = \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{D}^{*b}_{ML} - \frac{1}{B}\sum_{b=1}^{B}\widehat{D}^{*b}_{ML})^2$.

3.3 Exact Confidence Interval

Since lots of sensational GPQ applications have been reported in literature, in this section, we focus on developing GPQ-based exact confidence interval for the difference between two Youden indices with paired data.

We can modify the results (Li *et al.* [30]) for the difference in paired AUCs to derive the GPQ for the difference in paired Youden indices. Under the normal distribution assumption for the test results, the parameter of interest, D, is a function of parameters μ_x, μ_y, σ_{x1}, σ_{x2}, σ_{y1}, and σ_{y2}. The functions of parameters of interest are

$$\left(\mu_x, \mu_y, \beta_x, \beta_y, \sigma^2_{x2|1}, \sigma^2_{y2|1}\right) \equiv \left(\mu_x, \mu_y, \frac{\sigma_{x12}}{\sigma^2_x}, \frac{\sigma_{y12}}{\sigma^2_y}, \sigma^2_{x2} - \frac{\sigma^2_{x12}}{\sigma^2_{x1}}, \sigma^2_{y2} - \frac{\sigma^2_{y12}}{\sigma^2_{y1}}\right). \tag{3.8}$$

Notice that $\left(\beta_x, \beta_y, \sigma^2_{x2|1}, \sigma^2_{y2|1}\right)$ are nuisance parameters for D.

Let $\mathbf{X_1}, \cdots, \mathbf{X_{n_-}}$ and $\mathbf{Y_1}, \cdots, \mathbf{Y_{n_+}}$ be random samples from $\mathbf{N_2}(\mu_\mathbf{x}, \mathbf{\Sigma_x})$ and $\mathbf{N_2}(\mu_\mathbf{y}, \mathbf{\Sigma_y})$, respectively. Then the sufficient estimators for $(\mu_\mathbf{x}, \mathbf{\Sigma_x})$ and $(\mu_\mathbf{y}, \mathbf{\Sigma_y})$ are

$$\left(\widehat{\mu}_\mathbf{x}, \widehat{\mathbf{\Sigma}}_\mathbf{x}\right) = \left(\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{bmatrix}, \frac{1}{n_- - 1}\begin{bmatrix} SSX_1 & SSX_{12} \\ SSX_{12} & SSX_2 \end{bmatrix}\right),$$

and

$$\left(\widehat{\mu}_\mathbf{y}, \widehat{\mathbf{\Sigma}}_\mathbf{y}\right) = \left(\begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}, \frac{1}{n_+ - 1}\begin{bmatrix} SSY_1 & SSY_{12} \\ SSY_{12} & SSY_2 \end{bmatrix}\right),$$

where, $\bar{X}_i = \frac{1}{n_-}\sum_k X_{ik}$, $SSX_i = \sum_k \left(X_{ik} - \bar{X}_i\right)^2$, $SSX_{12} = \sum_k \left(X_{1k} - \bar{X}_1\right)\left(X_{2k} - \bar{X}_2\right)$, and $\bar{Y}_i$'s, SSY_i's and SSY_{12} are similarly defined based on the sample $\mathbf{Y_i}$'s.

Using these sufficient estimators, the parameters listed in (3.8) can be estimated by

$$\begin{aligned} &\left(M_x, M_y, B_x, B_y, SSX_{2|1}, SSY_{2|1}\right) = \\ &\left(\widehat{\mu}_x, \widehat{\mu}_y, \frac{SSX_{12}}{SSX_1}, \frac{SSY_{12}}{SSY_1}, SSX_2 - \frac{SSX_{12}^2}{SSX_1}, SSY_2 - \frac{SSY_{12}^2}{SSY_1}\right). \end{aligned} \tag{3.9}$$

Let U_{X1}, U_{Y1}, $U_{X2|1}$, $U_{Y2|1}$, Z_{BX}, Z_{BY}, Z_{MX}, and Z_{MY} be the quantities defined as follows:

$$U_{X1} = \frac{SSX_1}{\sigma^2_{x1}} \sim \chi^2_{n_- -1}, \qquad U_{Y1} = \frac{SSY_1}{\sigma^2_{y1}} \sim \chi^2_{n_+ -1} \tag{3.10}$$

$$U_{X2|1} = \frac{SSX_{2|1}}{\sigma^2_{x2|1}} \sim \chi^2_{n_- -2}, \qquad U_{Y2|1} = \frac{SSY_{2|1}}{\sigma^2_{y2|1}} \sim \chi^2_{n_+ -2} \tag{3.11}$$

$$Z_{BX} = (B_x - \beta_x)\sqrt{\frac{SSX_1}{\sigma^2_{x2|1}}} \sim N(0,1), \qquad Z_{BY} = (B_y - \beta_y)\sqrt{\frac{SSY_1}{\sigma^2_{y2|1}}} \sim N(0,1) \tag{3.12}$$

$$Z_{MX} = \left(\frac{\mathbf{\Sigma_x}}{n_-}\right)^{-1/2} (M_x - \mu_x) \sim N_2(0, \mathbf{I}) \tag{3.13}$$

$$Z_{MY} = \left(\frac{\mathbf{\Sigma_y}}{n_+}\right)^{-1/2} (M_y - \mu_y) \sim N_2(0, \mathbf{I}). \tag{3.14}$$

It is easy to verify that these quantities satisfy conditions (a) and (b) in the definition of GPQ (in section 2.3.2). Hence they are the pivotal quantities corresponding to estimators in (3.9).

Let $m_x = (m_{x1}, m_{x2})^T$, $m_y = (m_{y1}, m_{y2})^T$, b_x, b_y, $ssx_{2|1}$, ssx_1, $ssy_{2|1}$, and ssy_1 be the observed values of M_x, M_y, B_x, B_y, $SSX_{2|1}$, SSX_1, $SSY_{2|1}$, and SSY_1, respectively. By the mutual independence among $(M_x, \widehat{\mathbf{\Sigma}}_x)$ and $(M_y, \widehat{\mathbf{\Sigma}}_y)$ as well as the mutual independence among $(B_x, SSX_{2|1}, SSX_1)$ and $(B_y, SSY_{2|1}, SSY_1)$, we can find the following GPQs for $(\sigma^2_{x1}, \sigma^2_{x2|1}, \beta_x)$ and $(\sigma^2_{y1}, \sigma^2_{y2|1}, \beta_y)$:

$$\left(Q_{\sigma^2_{x1}}, Q_{\sigma^2_{x2|1}}, Q_{\beta_x}\right) = \left(\frac{ssx_1}{U_{X1}}, \frac{ssx_{2|1}}{U_{X2|1}}, b_x - Z_{BX}\sqrt{\frac{1}{U_{X2|1}}\frac{ssx_{2|1}}{ssx_1}}\right), \tag{3.15}$$

$$\left(Q_{\sigma^2_{y1}}, Q_{\sigma^2_{y2|1}}, Q_{\beta_y}\right) = \left(\frac{ssy_1}{U_{Y1}}, \frac{ssy_{2|1}}{U_{Y2|1}}, b_y - Z_{BY}\sqrt{\frac{1}{U_{Y2|1}}\frac{ssy_{2|1}}{ssy_1}}\right). \tag{3.16}$$

Similarly, the GPQs for $(\sigma^2_{x2}, \sigma_{x12})$ and $(\sigma^2_{y2}, \sigma_{y12})$ can be obtained as follows:

$$\left(Q_{\sigma^2_{x2}}, Q_{\sigma_{x12}}\right) = \left(Q^2_{\beta_x} Q_{\sigma^2_{x1}} + Q_{\sigma^2_{x2|1}}, Q_{\beta_x} R_{\sigma^2_{x1}}\right), \tag{3.17}$$

$$\left(Q_{\sigma^2_{y2}}, Q_{\sigma_{y12}}\right) = \left(Q^2_{\beta_y} Q_{\sigma^2_{y1}} + Q_{\sigma^2_{y2|1}}, Q_{\beta_y} R_{\sigma^2_{y1}}\right). \tag{3.18}$$

Therefore, $Q_{\Sigma_X} = \begin{bmatrix} Q_{\sigma^2_{x1}} & Q_{\sigma_{x12}} \\ Q_{\sigma_{x12}} & Q_{\sigma^2_{x2}} \end{bmatrix}$ and $Q_{\Sigma_Y} = \begin{bmatrix} Q_{\sigma^2_{y1}} & Q_{\sigma_{y12}} \\ Q_{\sigma_{y12}} & Q_{\sigma^2_{y2}} \end{bmatrix}$ are the GPQs for Σ_X and Σ_Y, respectively. The GPQs for μ_x and μ_y are given by

$$Q_{\mu_x} = \begin{bmatrix} Q_{\mu_{x1}} \\ Q_{\mu_{x2}} \end{bmatrix} = m_x - \left(\frac{Q_{\Sigma_x}}{n_-}\right)^{1/2} Z_{MX}, \tag{3.19}$$

$$Q_{\mu_y} = \begin{bmatrix} Q_{\mu_{y1}} \\ Q_{\mu_{y2}} \end{bmatrix} = m_y - \left(\frac{Q_{\Sigma_y}}{n_+}\right)^{1/2} Z_{MY}. \tag{3.20}$$

From (3.5), we can see that the Youden indices J_i's are functions of μ_{xi}, μ_{yi}, σ_{xi}, σ_{yi} and c_{oi} $(i = 1, 2)$. From equations (3.15)-(3.20), $Q_{\sigma_{xi}} = \sqrt{Q_{\sigma^2_{xi}}}$ and $Q_{\sigma_{yi}} = \sqrt{Q_{\sigma^2_{yi}}}$, we get the GPQs for c_{oi}'s:

$$Q_{c_{oi}} = \frac{Q_{\mu_{xi}}(Q^2_{b_i} - 1) - Q_{a_i} + Q_{b_i}\sqrt{Q^2_{a_i} + (Q^2_{b_i} - 1)Q_{\sigma_{xi}} ln(Q^2_{b_i})}}{(R^2_{b_i} - 1)} \tag{3.21}$$

where $Q_{a_i} = Q_{\mu_{yi}} - Q_{\mu_{xi}}$, and $Q_{b_i} = Q_{\sigma_{yi}}/Q_{\sigma_{xi}}$, for $i = 1, 2$. If $\sigma_{yi} = \sigma_{xi}$, then

$$Q_{c_{oi}} = \frac{Q_{\mu_{xi}} + Q_{\mu_{yi}}}{2}. \tag{3.22}$$

Therefore,

$$Q_D = Q_{J_1} - Q_{J_2}, \tag{3.23}$$

is the GPQ for the difference in paired Youden indices, where

$$Q_{J_i} = \Phi\left(\frac{Q_{\mu_{yi}} - Q_{c_{oi}}}{Q_{\sigma_{yi}}}\right) + \Phi\left(\frac{Q_{c_{oi}} - Q_{\mu_{xi}}}{Q_{\sigma_{xi}}}\right) - 1, \qquad i = 1, 2. \tag{3.24}$$

For the given non-diseased test results $(\mathbf{x_1}, \cdots, \mathbf{x_{n_-}})$ from $\mathbf{N_2}(\mu_\mathbf{x}, \mathbf{\Sigma_x})$ and the diseased test results $(\mathbf{y_1}, \cdots, \mathbf{y_{n_+}})$ from $\mathbf{N_2}(\mu_\mathbf{y}, \mathbf{\Sigma_y})$, to construct the GPQ-based confidence interval, we propose the following Monte-Carlo algorithm:

1. Compute the sample mean and sample covariance matrix using the diseased and the non-diseased test results respectively.

2. Use (3.10)-(3.14) to generate U_{X1}, U_{Y1}, $U_{X2|1}$ and $U_{Y2|1}$ from the corresponding chi-squared distributions and generate Z_{BX}, Z_{BY}, Z_{MX}, and Z_{MY} from the corresponding normal distributions.

3. Calculate $Q_{\mu_{xi}}$, $Q_{\mu_{yi}}$, $Q_{\sigma_{xi}}$, $Q_{\sigma_{yi}}$ and $Q_{c_{oi}}$ by using (3.15)-(3.22).

4. Calculate Q_D by using (3.23) and (3.24).

5. Repeat step 2 to step 4 H times (here we recommend H=10000) to obtain H repetitions of Q_D: $\{Q_{D_1}, Q_{D_2}, \cdots, Q_{D_H}\}$.

Consequently, the $100(1-\alpha)\%$ GPQ-based exact confidence interval for the difference between the paired Youden indices is defined as follows:

$$\left(Q_{D_{\alpha/2}}, Q_{D_{1-\alpha/2}}\right), \tag{3.25}$$

where $Q_{D_{\alpha/2}}$ is the $\alpha/2 \times 100$-th percentile of Q_{D_h}'s and $Q_{D_{1-\alpha/2}}$ is the $(1-\alpha/2) \times 100$-th percentile of Q_{D_h}'s.

3.4 Simulation Studies

In this section, we implement two simulation studies to evaluate the finite sample performance of the proposed confidence intervals, namely, the exact confidence interval based on GPQs, the HBML interval and the nonparametric BP interval.

In the first simulation study, the underlying non-diseased distribution and diseased distribution are $\mathbf{N_2}(\mu_\mathbf{x}, \mathbf{\Sigma_x})$ and $\mathbf{N_2}(\mu_\mathbf{y}, \mathbf{\Sigma_y})$ respectively. Without loss of generality, we choose $(\mu_{x1}, \mu_{x2}) = (0, 0)$, and $(\sigma_{x1}, \sigma_{x2}) = (\sigma_{y1}, \sigma_{y2}) = (1, 1)$. The sample sizes are chosen to be $n_-, n_+ = 10, 20, 50, 100$ with $n_- \leq n_+$. Three scenarios have been considered. In scenario 1, (μ_{y1}, μ_{y2}) is fixed at $(3.289708, 2.563104)$ such that $D = J_1 - J_2 = 0.9 - 0.8 = 0.1$; both T_1 and T_2 have high Youden index values, but the difference between the two Youden indices is small. In scenario 2, (μ_{y1}, μ_{y2}) is fixed at $(3.289708, 1.6832426)$ such that $D = J_1 - J_2 = 0.9 - 0.6 = 0.3$;

in this case, T_1 has high Youden index value, but T_2 has lower Youden index value than T_1, and the difference between the two Youden indices is big. In scenario 3, (μ_{y1}, μ_{y2}) is chosen to be $(1.6832426, 1.3489795)$ such that $D = J_1 - J_2 = 0.6 - 0.5 = 0.1$; both T_1 and T_2 have low Youden index values, but the difference between the two Youden indices is small. Finally, we set σ_{x12} and σ_{y12} to be 0.1 (low), 0.5 (medium) and 0.9 (high) which represent different levels of correlation between two diagnostic tests. Under each parameter setting, we simulate 5000 random samples from the underlying normal distributions. In the bootstrap procedure, we draw $B = 500$ bootstrap re-samples. The coverage probabilities and average interval lengths of the ECI, HBML and BP intervals are calculated based on the simulated samples. The results of the simulation study are reported in Table B.1 to Table B.6.

From Table B.1 to Table B.6, we make the following observations:

- Among all the proposed methods, the ECI interval outperforms the HBML and BP intervals in terms of coverage probability and average interval length in most simulation settings. However, when sample size is small $(n_-, n_+) \leq (20, 20)$, the ECI interval sacrifices the interval length to achieve the coverage probability's closure to nominal level. Expect small sample size situations, namely $(n_-, n_+) \leq (50, 50)$, the HBML method provides acceptable coverage accuracy.

- Comparing the HBML and BP methods, we observe that the HBML interval is more stable than the BP interval in three scenarios considered here. The performance pattern of the HBML method is as expected. As the sample size increases, its performance enhanced constantly. However, the performance of the BP interval varies a lot, and it performs worse when sample sizes are unequal.

- An interesting observation is the good performance of the BP method in scenario 3 in which the BP interval performs far better than in scenarios 1 and 2. In scenarios 1 and 2, $(J_1, J_2) = (0.9, 0.8)$, $(0.9, 0.6)$, respectively. In both cases, $J_1 = 0.9$, which requires the corresponding sensitivity and specificity values being at least 0.9 (a proportion close to the boundary 1). In fact, (J_i) is the difference between two proportions. Its nonparametric estimate $(\widehat{J}_{iE})$ may be a poor estimate for the difference between two proportions when one of the proportions is close to 1, which may explain the poor performance of the BP interval in scenarios 1 and 2.

However, in scenario 3, the true Youden indices are 0.6 and 0.5. Both values are not close to the boundary, and the BP method doesn't suffer from such boundary effect and thus performs significantly better.

In the second simulation study, we explore the robustness of the proposed GPQ-based ECI interval. For this purpose, we generate diseased and non-diseased samples from mixed normal distributions, i.e., $\mathbf{X} \sim p_x\mathbf{N_2}(\mu_\mathbf{x}, \mathbf{\Sigma_x}) + (1-p_x)\mathbf{N_2}(\mu_\mathbf{mix}, \mathbf{\Sigma_{mix}})$ and $\mathbf{Y} \sim p_y\mathbf{N_2}(\mu_\mathbf{y}, \mathbf{\Sigma_y}) + (1-p_y)\mathbf{N_2}(\mu_\mathbf{mix}, \mathbf{\Sigma_{mix}})$, where $(\mu_\mathbf{x}, \mathbf{\Sigma_x})$ and $(\mu_\mathbf{y}, \mathbf{\Sigma_y})$ are the same as those used in the first simulation study, $p_x = p_y = 0.9$, $\mathbf{\Sigma_{mix}} = \begin{bmatrix} 2 & 1.6 \\ 1.6 & 2 \end{bmatrix}$, and $\mu_\mathbf{mix} = \mu_\mathbf{y}/\mathbf{2}$. We apply the $\mu_\mathbf{y}$ values in above three scenarios to explore the performance of ECI interval when the underlying distributions are misspecified. The simulation results are reported in Table B.7. From Table B.7, we can observe that the performance of the ECI interval with misspecified underlying distributions is not as good as when the underline distributions are exactly normal as expected. However, its performance is still acceptable.

3.5 A Real Application

The pancreatic cancer data discussed by Wieand et al. [61] is re-analyzed here. The dataset was from a case-control study, in which, 51 controls and 90 cases involved. In that study, for each subject, the outcomes of laboratory tests on biomarker CA-125 and biomarker CA-19-9 were measured separately for the purpose of identifying the disease status. Both of the test results are positive values. We are interesting in the comparison of the accuracy of the two biomarkers in detecting pancreatic cancer. The original test results from the two populations are not normally distributed. After taking log transformation to the CA19-9 test results and Box-Cox transformation with the power parameter $\varphi = -0.425$ to the CA-125 test results, the transformed data would follow the normal distribution. The point estimate for the difference between Youden indices of two biomarkers is $D = 0.4110$ by the parametric method. By applying the non-parametric method to the original data, the estimated D is 0.2745. In addition, the 95 percent confidence intervals for D are $(0.2369, 0.5850)$ and $(0.0745, 0.4451)$ by HBML and BP methods, respectively. The 95 percent ECI interval for D is $(0.2557, 0.5660)$, which is the shortest one among the three intervals. All these confidence intervals do not include 0. Therefore, we can conclude that the accuracy of CA-125 is

significantly higher than that of CA-19-9 at 95 percent significant level and recommend biomarker CA-125 for pancreatic cancer detection in this example.

Chapter 4

INFERENCES FOR YOUDEN INDEX OF THREE ORDINAL GROUPS

4.1 Introduction

As highlighted in section 1.4, it is necessary to extend the original Youden index to be a suitable measure of accuracy for diagnostic tests with three ordinal groups. Let W denote the continuous test result from the intermediate population. Denote $W_1, W_2, \ldots, W_{n_0}$ be a random sample from intermediate patients. Let F_0 and f_0 be the c.d.f. and p.d.f. for W, respectively. Also, assume X, W, and Y are independent without loss of generality.

Following the definitions of sensitivity, specificity, and the true transitional rate in section 1.4, for given thresholds c_1 and c_2, they can be expressed as:

$$Sen(c_2) = P(Y \geq c_2) = 1 - F_+(c_2), \tag{4.1}$$

$$Spe(c_1) = P(X \leq c_1) = F_-(c_1), \tag{4.2}$$

and

$$\begin{aligned} Sin(c_1, c_2) &= P(c_1 \leq W \leq c_2) \\ &= F_0(c_2) - F_0(c_1), \end{aligned} \tag{4.3}$$

respectively. The quantity $Sin(c_1, c_2)$ represents the probability that a randomly selected transitional patient's test result falls in the interval $[c_1, c_2]$.

Nakas *et al.* [39] extended the Youden index for three ordinal groups via defining the Youden index as a maximum value of summation of Sen, Spe, and Sin, namely,

$$\begin{aligned} J_{3Nakas}(c_{o1}, c_{o2}) &= \max_{c_1,c_2}\{Sen + Spe + Sin\} \\ &= \max_{c_1,c_2}\{F_-(c_1) + F_0(c_2) - F_0(c_1) - F_+(c_2) + 1\} \\ &= F_-(c_{o1}) + F_0(c_{o2}) - F_0(c_{o1}) - F_+(c_{o2}) + 1, \end{aligned} \tag{4.4}$$

where (c_{o1}, c_{o2}) are the pair of optimal cut-off points. They also generalized this definition to be a weighted summation of Sen, Spe, and Sin, and expressed as

$$J^*_{3Nakas}(c_{o1}, c_{o2}) = \max_{c_1,c_2}\{\omega_1 Sen + \omega_2 Spe + \omega_3 Sin\}. \tag{4.5}$$

where the three weights, ω_1, ω_2 and ω_3 could be understood as the prevalence of the three classes in the whole population. Under such circumstance, the J^*_{3Nakas} is equivalent to the ratio of maximum correctness criterion and maximum expected utility (He and Frey [20]). Such Youden index has possible values between 1 and 3. $J_{3Nakas} = 1$ indicates that the test results' distributions of three groups fully overlap, therefore, the test is a pointless test procedure. If the three distributions are perfectly discriminated to each other, then $J_{3Nakas} = 3$, which indicate the test is a perfect diagnostic procedure. Since the original Youden index's range is between 0 and 1, this definition is not a nature extension of the original Youden index.

4.2 New Definition of The Youden Index for Three Ordinal Groups

In this chapter, we alternatively define the Youden index for three groups as follows:

$$J_3(c_{o1}, c_{o2}) = \frac{1}{2}\max_{c_1,c_2}\{Sen + Spe + Sin - 1\} \tag{4.6}$$

$$= \frac{1}{2}\max_{c_1,c_2}\{F_-(c_1) - F_0(c_1) + F_0(c_2) - F_+(c_2)\} \tag{4.7}$$

Assume that,

(A4.1) there exist a pair of values c_{1o} and c_{2o} such that $f_-(c_{1o}) = f_0(c_{1o})$, $f_-(t) \leq f_0(t)$ if $t \leq c_{1o}$, and $f_-(t) \geq f_0(t)$ if $t \geq c_{1o}$; $f_0(c_{2o}) = f_+(c_{2o})$, $f_0(t) \leq f_+(t)$ if $t \leq c_{2o}$, and $f_0(t) \geq f_+(t)$ if $t \geq c_{2o}$.

This assumption indicates that stochastically, X is less than W, and W is less than Y, i.e., $F_+(t) \leq F_0(t) \leq F_-(t)$ for all t. It also indicates that there is a unique pair of optimal thresholds, which are c_{o1} and c_{o2}. Furthermore, based on this assumption, we have

$$J_3(c_{o1}, c_{o2}) = \frac{1}{2}\left[\max_{c_1}\{F_-(c_1) - F_0(c_1)\} + \max_{c_2}\{F_0(c_2) - F_+(c_2)\}\right] \tag{4.8}$$

$$= \frac{1}{2}\{[F_-(c_{o1}) - F_0(c_{o1})] + [F_0(c_{o2}) - F_+(c_{o2})]\}. \tag{4.9}$$

Notice that, the first half in (4.8) is the Youden index for measuring the capacity of the diagnostic procedure in discriminating the health group and the intermediate group; and the second half is the Youden index for measuring the capability of the diagnostic procedure in discriminating the intermediate group and the diseased group. Thus, the J_3 defined in this way could be interpreted as the average discriminative accuracy of the test in identifying X and W, and W and Y.

Denote $J_{-0} = \max_{c_1}\{F_-(c_1) - F_0(c_1)\}$, $J_{0+} = \max_{c_2}\{F_0(c_2) - F_+(c_2)\}$, and J_3 for $J_3(c_{o1}, c_{o2})$. Then, we have $J_3 = \frac{1}{2}(J_{-0} + J_{0+})$. J_3 inherits the desire features from the original Youden index and offers superb conveniences, computationally and theoretically:

- The possible range of its value is from zero to one inclusive. A value zero states a totally useless test and the value one indicates an ideal diagnostic test.
- The index is independent of the relative sizes of the non-diseased, intermediate, and diseased groups. It is also independent of the absolute sizes of the three groups.
- All the tests that have the same index value make the same total number of misclassifications per hundred patients.
- The procedure to calculate J_3 can be simplified to calculate the average of the two Youden indices for the two pairs of groups.

In this chapter, we would focus on deriving the asymptotic property of the empirical estimate for J_3. Chernoff [7] studied the asymptotic properties for the empirical estimate of mode. He obtained the limiting distribution of the estimate and proved that the convergence rate is $O_p(n^{-1/3})$. Later, Kim and Pollard [26] generalized the Chernoff's results to a higher dimension via a modified empirical process method and obtained the similar conclusion as in Chernoff [7]. Hsieh and Turnbull [22] applied those theoretical results to the nonparametric estimate for Youden index and developed the asymptotic distributions of the estimators for Youden index and its optimal cut-off point under two ordinal groups case. Here, we first extend Hsieh and Turnbull's [22] work to obtain the asymptotic distribution for the empirical estimate of J_3. Then, we propose several confidence intervals for J_3 and its optimal cut-off points.

4.3 Estimations of Youden index for Three Ordinal Groups

4.3.1 Parametric Estimation

In this section, we assume that

$$X \sim N(\mu_x, \sigma_x^2), \qquad W \sim N(\mu_w, \sigma_w^2), \qquad and \qquad Y \sim N(\mu_y, \sigma_y^2). \tag{4.10}$$

Under these assumptions, the optimal cut-off points have the following closed forms:

$$c_{o1} = \frac{\mu_x(b_1^2 - 1) - a_1 + b_1\sqrt{a_1^2 + (b_1^2 - 1)\sigma_ln(b_1^2)}}{(b_1^2 - 1)}, \tag{4.11}$$

$$c_{o2} = \frac{\mu_w(b_2^2 - 1) - a_2 + b_2\sqrt{a_2^2 + (b_2^2 - 1)\sigma_0 ln(b_2^2)}}{(b_2^2 - 1)}, \tag{4.12}$$

where $a_1 = \mu_w - \mu_x$, $b_1 = \sigma_w/\sigma_x$, $a_2 = \mu_y - \mu_w$, and $b_2 = \sigma_y/\sigma_w$. Similarly, in the presence of equal group standard deviations between X and W or/and between W and Y, the optimal cut-off points would be $c_{o1} = \frac{\mu_x+\mu_w}{2}$, $c_{02} = \frac{\mu_w+\mu_y}{2}$. Therefore,

$$J_3 = \frac{1}{2}\{\Phi\left(\frac{c_{o1} - \mu_x}{\sigma_x}\right) - \Phi\left(\frac{c_{o1} - \mu_w}{\sigma_w}\right) + \Phi\left(\frac{c_{o2} - \mu_w}{\sigma_w}\right) - \Phi\left(\frac{c_{o2} - \mu_y}{\sigma_y}\right)\}. \tag{4.13}$$

We obtain the MLEs for c_{o1}, c_{o2}, and J_3 via plugging in the MLEs, $\widehat{\mu}_x$, $\widehat{\mu}_w$, $\widehat{\mu}_y$, $\widehat{\sigma}_x$, $\widehat{\sigma}_w$, and $\widehat{\sigma}_y$, of μ_x, μ_w, μ_y, σ_x, σ_w, and σ_y into (4.11), (4.12), and (4.13). $\widehat{c}_{o1ML}$, $\widehat{c}_{02ML}$, and $\widehat{J}_{3ML}$ are given as

follows:

$$\widehat{c}_{o1ML} = \frac{\widehat{\mu}_x(\widehat{b}_1^2 - 1) - \widehat{a}_1 + \widehat{b}_1\sqrt{\widehat{a}_1^2 + (\widehat{b}_1^2 - 1)\widehat{\sigma}_- ln(\widehat{b}_1^2)}}{(\widehat{b}_1^2 - 1)},$$

$$\widehat{c}_{o2ML} = \frac{\widehat{\mu}_w(\widehat{b}_2^2 - 1) - \widehat{a}_2 + \widehat{b}_2\sqrt{\widehat{a}_2^2 + (\widehat{b}_2^2 - 1)\widehat{\sigma}_0 ln(\widehat{b}_2^2)}}{(\widehat{b}_2^2 - 1)},$$

$$\widehat{J}_{3ML} = \frac{1}{2}\{\Phi\left(\frac{\widehat{c}_{o1ML} - \widehat{\mu}_x}{\widehat{\sigma}_x}\right) - \Phi\left(\frac{\widehat{c}_{o1ML} - \widehat{\mu}_w}{\widehat{\sigma}_w}\right) + \Phi\left(\frac{\widehat{c}_{o2ML} - \widehat{\mu}_w}{\widehat{\sigma}_w}\right) - \Phi\left(\frac{\widehat{c}_{o2ML} - \widehat{\mu}_y}{\widehat{\sigma}_y}\right)\},$$

where $\widehat{a}_1 = \widehat{\mu}_w - \widehat{\mu}_x$, $\widehat{b}_1 = \widehat{\sigma}_w/\widehat{\sigma}_x$, $\widehat{a}_2 = \widehat{\mu}_y - \widehat{\mu}_w$, and $\widehat{b}_2 = \widehat{\sigma}_y/\widehat{\sigma}_w$. When $\widehat{b}_1 = 0$ and/or $\widehat{b}_2 = 0$, $\widehat{c}_{o1ML} = \frac{\widehat{\mu}_x + \widehat{\mu}_w}{2}$ and/or $\widehat{c}_{o2ML} = \frac{\widehat{\mu}_w + \widehat{\mu}_y}{2}$.

4.3.2 Nonparametric Estimation

When the underlying distributions for the test results are unknown, nonparametric estimates for Youden index and its associated optimal cut-off point are necessary.

A simple nonparametric estimate for J_3 can be obtained by substituting the c.d.f. F_-, F_0, and F_+ in (4.8) by their corresponding empirical distribution functions. The empirical estimate for J_3 is

$$\widehat{J}_{3E} = \frac{1}{2}\left[\max_{c_1}\{\widehat{F}_-(c_1) - \widehat{F}_0(c_1)\} + \max_{c_2}\{\widehat{F}_0(c_2) - \widehat{F}_+(c_2)\}\right], \tag{4.14}$$

where

$$\widehat{F}_-(c) = \frac{\sum_{i_-=1}^{n_-} I(X_{i_-} \leq c)}{n_-} \tag{4.15}$$

$$\widehat{F}_0(c) = \frac{\sum_{i_0=1}^{n_0} I(W_{i_0} \leq c)}{n_0}$$

$$\widehat{F}_+(c) = \frac{\sum_{i_+=1}^{n_+} I(Y_{i_+} \leq c)}{n_+}. \tag{4.16}$$

If let $\widehat{c}_{o1E}$ and $\widehat{c}_{o2E}$ be the empirical estimate for c_{o1} and c_{o2} which maximize (4.14), then (4.14) could be expressed as:

$$\widehat{J}_{3E} = \frac{1}{2}\left[\left(\widehat{F}_{-}(\widehat{c}_{o1E}) - \widehat{F}_0(\widehat{c}_{o1E})\right) + \left(\widehat{F}_0(\widehat{c}_{o2E}) - \widehat{F}_{+}(\widehat{c}_{o2E})\right)\right]. \tag{4.17}$$

Without specific assumption of the underlying distribution for the test results, there are no closed forms for $\widehat{c}_{o1E}$ and $\widehat{c}_{o2E}$. The estimates can only be found via numerical search. It is probable that there may be multiple pairs of $\widehat{c}_{o1E}$ and $\widehat{c}_{o2E}$ which could maximize (4.14). Here, we define the empirical estimates for the pair of optimal cut-off points as follows:

$$\begin{aligned} \widehat{c}_{o1E} &= median\{t_0 | \widehat{F}_{-}(t_0) - \widehat{F}_0(t_0) = \max_t \left(\widehat{F}_{-}(t) - \widehat{F}_0(t)\right)\}, \\ \widehat{c}_{o2E} &= median\{t_0 | \widehat{F}_0(t_0) - \widehat{F}_{+}(t_0) = \max_t \left(\widehat{F}_0(t) - \widehat{F}_{+}(t)\right)\}. \end{aligned} \tag{4.18}$$

4.4 Asymptotic Properties for the Parametric and Nonparametric Estimators

The asymptotic properties for the proposed parametric and nonparametric estimators are essential in inferential analysis. For instance, we might construct the confidence intervals for the Youden index and its optimal cut-off point based on their asymptotic distributions.

When the underlying distributions are normal, asymptotic properties for $\widehat{J}_{3ML}$ could be derived by the Delta method (see Schisterman and Perkins [48]). Then, we can apply the asymptotic normality for $\widehat{J}_{3ML}$ to construct confidence interval for J_3.

We focus on the asymptotic properties for the empirical estimator. Since no specific assumptions are made for F_g, $g = -, 0, +$, further assumptions are necessary to develop the main results in this section. Similar to the assumptions made in Hsieh and Turnbull [22] for the original Youden index, a slightly weaker version of A4.1 is assumed to restrict the uniqueness of the optimal cut-off points. That is:

(A4.2) For any $\delta_1 \geq 0$, there exists ϵ_1 (≥ 0), such that

$$\sup_{|x - c_{o1}| \geq \delta_1} [F_{-}(x) - F_0(x)] \leq F_{-}(c_{o1}) - F_0(c_{o1}) - \epsilon_1,$$

and for any $\delta_2 \geq 0$, there exists $\epsilon_2\,(\geq 0)$, such that

$$\sup_{|x-c_{o2}|\geq\delta_2} [F_0(x) - F_+(x)] \leq F_0(c_{o2}) - F_+(c_{o2}) - \epsilon_2.$$

Now let $\mathcal{C}^{(d)}(\mathcal{S})$ denote the class of functions with continuous k-th derivative in the support $\mathcal{S}$, $\mathcal{S} \subset \Re$. Further assume that

(A4.3) F_g, $g = -, 0, +$ are in $\mathcal{C}^{(2)}(l, u)$. $c_{o1} \in (l_1, u_1)$ and $c_{o2} \in (l_2, u_2)$ with $u_1 \leq l_2$, and the interval (l_1, u_2) is a subset of (l, u). Further, assume that the domains of F_- and F_0 overlap, and the intersection of their domains contains (l_1, u_1). Also assume that the domains of F_0 and F_+ overlap, and the intersection of their domains contains (l_2, u_2).

(A4.4) $|f_-^{'}(c_{o1}) - f_0^{'}(c_{o1})| = d_1 > 0$, and $|f_0^{'}(c_{o2}) - f_+^{'}(c_{o2})| = d_2 > 0$.

Based results in Kim and Pollard [26], and Hsieh and Turnbull [22], we can obtain the following theorem,

Theorem 4.1. *Let F_g, $g = -, 0, +$ satisfied (A4.2), (A4.3), and (A4.4). If $n_-/n_0 \to \lambda_1^2$ and $n_0/n_+ \to \lambda_2^2$ for some positive finite λ_1 and λ_2, then*

$\widehat{J}_{3E}$ converges to J_3 almost surely, and $\sqrt{4n_0}\left(\widehat{J}_{3E} - J_3\right)$ converges in distribution to $\left[\lambda_1^{-1}B_1(F_-(c_{o1})) - B_2(F_0(c_{o1})) + B_2(F_0(c_{o2})) - \lambda_2 B_3(F_+(c_{o2}))\right]$, where B_1, B_2, and B_3 are independent Brownian bridges on $[0, 1]$.

Its proof can be found in Appendix D.

4.5 Confidence Intervals for the Youden index of Three Ordinal Groups

In this section, we propose parametric and nonparametric confidence intervals for J_3.

Under assumption (4.10), we first propose the GPQs based exact confidence interval (ECI) for J_3. Notice that J_3 is a function of parameters μ_x, μ_w, μ_y, σ_x, σ_w, and σ_y. And the functions of parameters of interest are

$$\left(\mu_x, \mu_w, \mu_y, \sigma_x^2, \sigma_w^2, \sigma_y^2\right). \tag{4.19}$$

Then the sufficient estimators for (μ_x, σ_x^2), (μ_w, σ_w^2), and $\left(\mu_y, \sigma_y^2\right)$ are

$$\left(\widehat{\mu}_x, \widehat{\sigma}_{xs}^2\right) = \left(\bar{X}, \frac{1}{n_- - 1} SSX\right),$$

$$\left(\widehat{\mu}_w, \widehat{\sigma}_{ws}^2\right) = \left(\bar{W}, \frac{1}{n_0 - 1} SSW\right),$$

and

$$\left(\widehat{\mu}_y, \widehat{\sigma}_{ys}^2\right) = \left(\bar{Y}, \frac{1}{n_+ - 1} SSY\right),$$

respectively, where,$\bar{X} = \frac{1}{n_-}\sum_{i=1}^{n_-} X_i$, $SSX = \sum_{i=1}^{n_-}\left(X_i - \bar{X}\right)^2$, and $\bar{Y}$, SSY, $\bar{W}$, and SSW are similarly defined.

Let U_X, U_W, U_Y, Z_X, Z_W and Z_Y be the quantities defined as follows:

$$U_X = \frac{SSX}{\sigma_x^2} \sim \chi^2_{n_- - 1}, \tag{4.20}$$

$$U_W = \frac{SSW}{\sigma_w^2} \sim \chi^2_{n_0 - 1}, \tag{4.21}$$

$$U_Y = \frac{SSY}{\sigma_y^2} \sim \chi^2_{n_+ - 1}, \tag{4.22}$$

$$Z_X = \left(\frac{\sigma_x}{n_-}\right)^{-1/2} (\bar{X} - \mu_x) \sim N(0,1), \tag{4.23}$$

$$Z_W = \left(\frac{\sigma_w}{n_0}\right)^{-1/2} (\bar{W} - \mu_w) \sim N(0,1), \tag{4.24}$$

$$Z_Y = \left(\frac{\sigma_y}{n_+}\right)^{-1/2} (\bar{Y} - \mu_y) \sim N(0,1). \tag{4.25}$$

It is easy to verify that these quantities satisfy conditions (a) and (b) in the definition of GPQs. Denote $\widehat{\mu}_x$, $\widehat{\mu}_w$, $\widehat{\mu}_y$, ssx, ssw, and ssy be the observed values of $\bar{X}$, $\bar{W}$, $\bar{Y}$, SSX, SSW, and SSY, respectively. We can construct the GPQs for σ_x^2, σ_w^2, and σ_y^2 as follows:

$$Q_{\sigma_x^2} = \frac{ssx}{U_X}, \qquad Q_{\sigma_w^2} = \frac{ssw}{U_W}, \qquad \text{and} \qquad Q_{\sigma_y^2} = \frac{ssy}{U_Y}. \tag{4.26}$$

Then, the GPQs for σ_x, σ_w, σ_y are

$$Q_{\sigma_x} = \sqrt{Q_{\sigma_x^2}}, \qquad Q_{\sigma_w} = \sqrt{Q_{\sigma_w^2}}, \qquad \text{and} \qquad Q_{\sigma_y} = \sqrt{Q_{\sigma_y^2}}. \tag{4.27}$$

The GPQs for μ_x, μ_w, and μ_y are defended as follows:

$$Q_{\mu_x} = \widehat{\mu}_x - \frac{Q_{\sigma_x}}{\sqrt{n_-}} Z_X, \tag{4.28}$$

$$Q_{\mu_w} = \widehat{\mu}_w - \frac{Q_{\sigma_w}}{\sqrt{n_0}} Z_W, \tag{4.29}$$

$$Q_{\mu_y} = \widehat{\mu}_y - \frac{R_{\sigma_y}}{\sqrt{n_+}} Z_Y. \tag{4.30}$$

Based on these GPQs, the GPQs $Q_{c_{o1}}$ and $Q_{c_{o2}}$ for c_{o1} and c_{o2} are respectively given as:

$$Q_{c_{o1}} = \frac{Q_{\mu_x}(Q_{b_1}^2 - 1) - Q_{a_1} + Q_{b_1}\sqrt{Q_{a_1}^2 + (Q_{b_1}^2 - 1)Q_{\sigma_x} ln(Q_{b_1}^2)}}{(R_{b_1}^2 - 1)} \tag{4.31}$$

where $Q_{a_1} = Q_{\mu_w} - Q_{\mu_x}$, and $Q_{b_1} = Q_{\sigma_w}/Q_{\sigma_x}$. If $\sigma_w = \sigma_x$, then

$$Q_{c_{o1}} = \frac{Q_{\mu_x} + Q_{\mu_w}}{2}. \tag{4.32}$$

$$Q_{c_{o2}} = \frac{Q_{\mu_w}(Q_{b_2}^2 - 1) - Q_{a_2} + Q_{b_2}\sqrt{Q_{a_2}^2 + (Q_{b_2}^2 - 1)Q_{\sigma_w} ln(Q_{b_2}^2)}}{(R_{b_2}^2 - 1)} \tag{4.33}$$

where $Q_{a_2} = Q_{\mu_y} - Q_{\mu_w}$, and $Q_{b_2} = Q_{\sigma_y}/Q_{\sigma_w}$. If $\sigma_y = \sigma_w$, then

$$Q_{c_{o2}} = \frac{Q_{\mu_w} + Q_{\mu_y}}{2}. \tag{4.34}$$

Therefore, the GPQ Q_{J_3} for J_3 is given as:

$$Q_{J_3} = \frac{1}{2}\{\Phi\left(\frac{Q_{c_{o1}} - Q_{\mu_x}}{Q_{\sigma_x}}\right) - \Phi\left(\frac{Q_{c_{o1}} - Q_{\mu_w}}{Q_{\sigma_w}}\right) + \tag{4.35}$$

$$\Phi\left(\frac{Q_{c_{o2}} - Q_{\mu_w}}{Q_{\sigma_w}}\right) - \Phi\left(\frac{Q_{c_{o2}} - Q_{\mu_y}}{Q_{\sigma_y}}\right)\}. \tag{4.36}$$

Follow the similar Monte-Carlo algorithm in Chapter 3, simulate the quantity Q_{J3} H times to obtain $\{Q_{J3}^h : h = 1, 2, \ldots, H\}$. Then, the $100(1-\alpha)\%$ GPQ-based ECI for J_3 is defined as follows:

$$\left(Q_{J3_{\alpha/2}}, Q_{J3_{1-\alpha/2}}\right), \tag{4.37}$$

where $Q_{J3_{\alpha/2}}$ and $Q_{J3_{1-\alpha/2}}$ are the $\alpha/2 \times 100$-th and the $(1-\alpha/2) \times 100$-th percentiles of Q^h_{J3}'s. The exact confidence intervals can be similarly constructed for c_{o1} and c_{o2}.

The second parametric confidence interval for J_3 is based on the asymptotic normality of $\widehat{J}_{3ML}$. Instead of derive the asymptotic variance for J_{3ML} via Delta method, we apply the bootstrap procedures mentioned in chapter 3 to calculate the bootstrap variance estimate $Var(\widehat{J}^*_{3ML})$ of $Var(\widehat{J}_{3ML})$. Let $\{\widehat{J}^{*b}_{3ML} : b = 1, 2, \ldots, B\}$ be the bootstrap replications of $\widehat{J}_{3ML}$. The bootstrap variance can be calculated as $Var(\widehat{J}^*_{3ML}) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{J}^{*b}_{3ML} - \frac{1}{B}\sum_{b=1}^{B}\widehat{J}^{*b}_{3ML}\right)^2$. Then, the $(1-\alpha)100\%$ hybrid bootstrap and maximum likelihood (HBML) based confidence interval for J_3 is defined as follow:

$$\left(\widehat{J}_{3ML} - z_{1-\alpha/2}\sqrt{Var(\widehat{J}^*_{3ML})}, \widehat{J}_{3ML} + z_{1-\alpha/2}\sqrt{Var(\widehat{J}^*_{3ML})}\right). \tag{4.38}$$

We can similarly construct the $(1-\alpha)100\%$ HBML based confidence interval for c_{o1} and c_{o2}.

Inspired by the successful application of the AC adjusted normal approximation (ACNA) confidence interval for the original Youden index, we also propose this confidence interval for J_3. Define the AC adjusted empirical estimate for J_3 to be

$$\begin{aligned}\widehat{J}_{3AC} \;=\; & \frac{1}{2}\max_{c_1}\{\frac{\sum_{i_-=1}^{n_-} I(X_{i_-} \le c_1) + 2}{n_- + 4} - \frac{\sum_{i_0=1}^{n_0} I(W_{i_0} \le c_1) + 2}{n_0 + 4}\} + \\ & \frac{1}{2}\max_{c_2}\{\frac{\sum_{i_0=1}^{n_0} I(W_{i_0} \le c_2) + 2}{n_0 + 4} - \frac{\sum_{i_+=1}^{n_+} I(Y_{i_+} \le c_2) + 2}{n_+ + 4}\}.\end{aligned} \tag{4.39}$$

According to Theorem 4.1, $\widehat{J}_{3AC}$ has the asymptotic normality at given cut-off points c_1 and c_2; however, we face the same difficulty that the asymptotic variance for $\widehat{J}_{3AC}$ is unknown. Again, we apply the bootstrap procedure to estimate the $Var(\widehat{J}_{3AC})$. After obtaining the B bootstrap replications of $\widehat{J}_{3AC}$: $\{\widehat{J}^{*b}_{3AC} : b = 1, 2, \ldots, B\}$, the $(1-\alpha)100\%$ ACNA confidence interval for J_3 is defined as follows:

$$\left(\bar{J}^*_{3AC} - z_{1-\alpha/2}\sqrt{Var(\widehat{J}^*_{3AC})},\ \bar{J}^*_{3AC} + z_{1-\alpha/2}\sqrt{Var(\widehat{J}^*_{3AC})}\right), \tag{4.40}$$

where $\bar{J}^*_{3AC} = \frac{1}{B}\sum_{b=1}^{B}\widehat{J}^{*b}_{3AC}$ and $Var(\widehat{J}^*_{3AC}) = \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{J}^{*b}_{3AC} - \bar{J}^*_{3AC})^2$ is the variance of bootstrap replications.

Additionally, we apply the Bias Corrected and accelerated (BCa) method to construct another nonparametric confidence interval for J_3 as a comparison to the ACNA confidence interval for J_3. To construct the BCa interval, first we need to apply bootstrap technology to obtain B bootstrap replications of $\widehat{J}_{3E}$, denoted by $\widehat{J}_{3E}^{*b} : b = 1, 2, \ldots, B$. Let $\widehat{J}_{3E}^{(*1)}, \widehat{J}_{3E}^{(*2)}, \ldots, \widehat{J}_{3E}^{(*B)}$ be the ordered $\widehat{J}_{3E}^{*b}$. The BCa confidence interval is defined as follow:

$$\left(\widehat{J}_{3E}^{(*[(B+1)\theta_1])}, \widehat{J}_{3E}^{(*[(B+1)\theta_2])}\right), \tag{4.41}$$

where $\theta_1 = \Phi(\zeta_2 + \frac{\zeta_2 + z_{\alpha/2}}{1-\zeta_1(\zeta_2+z_{\alpha/2})})$, and $\theta_2 = \Phi(\zeta_2 + \frac{\zeta_2 + z_{1-\alpha/2}}{1-\zeta_1(\zeta_2+z_{1-\alpha/2})})$. Let $N = n_- + n_0 + n_+$, then the two quantities ζ_1 and ζ_2 can be found by the following equations:

$$\zeta_1 = \sum_{i=1}^{N} \vartheta_i^3 / (6(\sum_{i=1}^{N} \vartheta_i^2)^{3/2}), \text{ and } \zeta_2 = \Phi^{-1}(\frac{1}{B}\sum_{b=1}^{B}(\widehat{J}_{3E}^{*b} < \widehat{J}_{3E})). \tag{4.42}$$

ϑ_i is defined as: $\vartheta_i = \widehat{J}_{3E}^{(\cdot)} - \widehat{J}_{3E}^{-i}$, where $\widehat{J}_{3E}^{-i}$ is calculated by delete the i-th observation in the original samples, and $\widehat{J}_{3E}^{(\cdot)} = \frac{\sum_{i=1}^{N} \widehat{J}_{3E}^{-i}}{N}$. In (4.41), $[(B+1)\theta_1]$ represent the integer part of $(B+1)\theta_1$, and $\widehat{J}_{3E}^{(*[(B+1)\theta_1])}$ represent the $[(B+1)\theta_1]$-th bootstrap replication in $\widehat{J}_{3E}^{(*1)}, \widehat{J}_{3E}^{(*2)}, \ldots, \widehat{J}_{3E}^{(*B)}$. The upper bound $\widehat{J}_{3E}^{(*[(B+1)\theta_2])}$ is similarly defined.

The nonparametric confidence intervals for the optimal thresholds are the bootstrap percentile (BP) based intervals.

4.6 Simulation Studies

Since the ECI and HBML methods rely on the normal assumption, we generate test results for the three groups from the normal distribution. Without loss of generality, we fix $\mu_- = 1$, $\sigma_- = 1$ and $\sigma_+ = 1.2$. The other parameters are selected based on the following considerations:

- To evaluate the performance of the three methods comprehensively, the true values of J_3 are selected as 0.55, 0.65, 0.75, and 0.85. These values indicate that the diagnostic accuracy of the test is at the acceptable, good, very good, or excellent levels, respectively.

- Since J_3 can be interpreted as the average accuracy of the test in discriminating the two pairs of groups, i.e., J_3 is the average of J_{-0} and J_{0+}. So, the values of μ_0, σ_0, c_{o1}, and c_{o2} are determined to achieve specified values of J_{-0} and J_{0+} such that $J_3 = \frac{1}{2}(J_{-0} + J_{0+})$.

The parameter settings are listed in the table below.

Table 4.1. Parameter Setting for Chapter 4

J_3	μ_0	σ_0	μ_+	c_{o1}	c_{o2}	J_{-0}	J_{0+}
0.55	2.8670	1.2358	4.5095	1.9730	3.6738	0.6	0.5
0.55	2.3627	1.0205	4.2229	1.6896	3.3233	0.5	0.6
0.65	3.1259	1.0518	5.0152	2.0611	4.0959	0.7	0.6
0.65	2.5817	0.8843	4.7189	1.7710	3.6368	0.6	0.7
0.75	6.8108	4.1214	11.5606	2.9286	9.5174	0.8	0.7
0.75	4.8503	3.2713	10.2040	2.6769	8.1232	0.7	0.8
0.85	8.5329	3.9086	14.5146	3.1501	12.3324	0.9	0.8
0.85	5.6948	2.9612	12.3094	2.7733	9.9487	0.8	0.9

In above settings, under the same J_3 value, we have two scenarios, let I be the label for the case that the J_{-0} is larger than the J_{0+}, II be the label for the case that the J_{-0} is less than the J_{0+}. Here the sample sizes n_-, n_0, and n_+ are determined to be $(50, 50, 50)$, $(80, 80, 80)$, $(100, 100, 100)$, and $(100, 50, 80)$. Let the label for the four sample size settings above be S_1, S_2, S_3, and S_4, respectively. For each simulation setting, we generate 1000 random samples from the underlying distributions to calculate the empirical average coverage probability and average interval length. In each run, we apply the re-sampling procedure $B = 500$ times.

The simulated results, summarized in Appendix C, lead to the following conclusions. Under normal assumptions for the underlying distribution, the GPQ based ECI for J_3 outperforms all other intervals. The performance of the HBML based confidence interval for J_3 is adept at the selected sample sizes. The ACNA confidence interval for J_3 generally has an acceptable performance and performs better than the BCa interval. However, its performance is poor when $J_3 = 0.55$. The reason is that the $\widehat{J}_{3AC}$ tends to be much larger than the true $J_3 = 0.55$, further, when sample sizes increase, the estimated variance for $\widehat{J}_{3AC}$ decreases. This causes the increment of the lower bound error of this interval resulting that the average coverage probability decreases as sample sizes increase.

The ECI and HBML based confidence interval for optimal cut-off points perform as good as expected. However, the BP interval for the optimal thresholds consistently over estimate, whose coverage probability almost exceeds the confidence level, and the average interval lengths are much wider than the parametric methods.

In practice, if the normal distribution assumptions are valid for the test results, we recommend the ECI for J_3, otherwise, the ACNA interval should be applied.

Chapter 5

NONPARAMETRIC COVARIATES ADJUSTMENT FOR YOUDEN INDEX

5.1 Introduction

In the studies of exploring covariate effects on the accuracy of diagnostic test, the milestone would be the paper of Pepe [42], in which, Pepe introduced three approaches to model the covariates effects on diagnostic accuracy. The first approach is directly setting up a regression model for the test result on the covariates. The second one is to model the AUC on the covariates. And the last one is to directly model the ROC curve on the covariate information. Table 2 in Pepe [42] comprehensively summarized the advantages and disadvantages for the three approaches. Later, numerous literatures discussed the covariates adjustment for the performance of diagnostic test as introduced in Chapter 1. If concerning individual parameters, like the cut-off point, Yao *et al.* [64] indicated that the last two approaches lose the connection with the cut-off value and does not allow the prediction of the sensitivity and specificity at a given cut-off conditional on covariates. Under this consideration, the first approach is more interesting. Faraggi [13] employed the first approach by using simple linear regression models with normal error, and then proposed the corresponding adjusted estimate for AUC and Youden index. Nonetheless, the linear model for the test results is limited to linear form and the homoscedasticity error assumption. Yao *et al.* [64] extended Faraggi's [13] work by using a non-parametric heteroscedastic regression models. Here, we utilize the same models as in Yao *et al.* [64] to establish inferences of covariats adjustment for the Youden index.

5.2 Model and Methods

5.2.1 Heteroscedastic Regression Models for the Test Results

Assume the following non-parametric models for X and Y:

$$X|(Z = z) = \mu_x(z) + \sqrt{\nu_x(z)}\epsilon_x, \tag{5.1}$$

$$Y|(Z = z) = \mu_y(z) + \sqrt{\nu_y(z)}\epsilon_y, \tag{5.2}$$

where Z represents the covariates, the standard errors ϵ_1 and ϵ_2 are independent of each other and have mean zero and standard deviation one, the range of the variance functions $\nu_1(z)$ and $\nu_2(z)$ is restricted in $\Re^+$ and finite for all $z \in \Re$. In addition, let F_{-Z} and F_{+Z} denote the c.d.f. of X and Y at given Z respectively, f_{-Z} and f_{+Z} denote the p.d.f. of X and Y at given Z respectively, $F^*_-(\cdot)$ and $F^*_+(\cdot)$ denote the c.d.f. of ϵ_1 and ϵ_2 respectively, and $f^*_-(\cdot)$ and $f^*_+(\cdot)$ denote the p.d.f. of ϵ_1 and ϵ_2 respectively. Here, the error distributions F^*_- and F^*_+ are assumed to be independent of Z.

5.2.2 Covariate-adjusted Youden Index under Normal Error Assumption

With the covariates Z, both the Youden index and the optimal cut-off point actually are dependent on Z. Let $C(z) = \{c : \max_c [P(Y \geq c|Z = z)) + P(X \leq c|Z = z) - 1]\}$ represent the collection of possible optimal cut-off points at $Z = z$, $c_{o,1}(z) = \inf_c\ C(z)$, and $c_{o,2}(z) = \max_c\ C(z)$. The Youden index at given $Z = z$ is

$$\begin{aligned} J(z) &= \max_c\{P(Y \geq c|Z = z)) + P(X \leq c|Z = z) - 1\} \\ &= P(Y \geq c_o(z)|Z = z) + P(X \leq c_o(z)|Z = z) - 1 \\ &= P(X \leq c_o(z)|Z = z) - P(Y \leq c_o(z)|Z = z) \\ &= F_{-Z}(c_o(z)) - F_{+Z}(c_o(z)), \end{aligned} \tag{5.3}$$

where $c_o(z) = c_{o,1}(z), \text{or } c_{o,2}(z)$. If the errors ϵ_1 and ϵ_2 are assumed to be normally distributed in models (5.1) and (5.2), the Youden index at $Z = z$ can be expressed as

$$J_N(z) = \Phi\left(\frac{\mu_y(z) - c_o(z)}{\sqrt{\nu_y(z)}}\right) - \Phi\left(\frac{\mu_x(z) - c_o(z)}{\sqrt{\nu_x(z)}}\right), \tag{5.4}$$

where $J_N(z)$ stands for $J(z)$ under normal error. With normality and the assumption that $\mu_y(z) > \mu_x(z)$, $c_o(z)$ has the following closed form:

$$c_o(z) = \frac{\mu_x(z)(b^2 - 1) - a + b\sqrt{a^2 + (b^2 - 1)\nu_x(z)ln(b^2)}}{(b^2 - 1)}, \tag{5.5}$$

where $a = \mu_y(z) - \mu_x(z)$, $b = \sqrt{\nu_y(z)}/\sqrt{\nu_x(z)}$.

Under models (5.1)- (5.2), the mean and variance functions μ_x, μ_y, ν_x, and ν_y can be consistently estimated via nonparametric method, namely, the local polynomial regression technique. Let $\widehat{\mu}_x(z)$, $\widehat{\mu}_y(z)$, $\widehat{\nu}_x(z)$, and $\widehat{\nu}_y(z)$ be the local polynomial estimates for μ_x, μ_y, ν_x, and ν_y by using local polynomial method (Fan and Gijbels [12]). Let $\{(z_{i,x}, x_i) : i = 1 \cdots n_-\}$ and $\{(z_{j,y}, y_j) : j = 1 \cdots n_+\}$ be random samples of "non-diseased" subjects and "diseased" subjects from models (5.1)- (5.2), where $z_{i,x}$ and $z_{j,y}$ are the corresponding observed covariate values in the "non-diseased" and "diseased" samples. Then, $\widehat{\mu}_x(z)$ can be obtained by minimizing:

$$\sum_{i=1}^{n_-}\{x_i - \sum_{k=0}^{p}\mathfrak{B}_k(z_{i,x} - z)^k\}^2\mathcal{K}(\frac{z_{i,x} - z}{h_{\mu,x}})/h_{\mu,x}, \tag{5.6}$$

where $\mathcal{K}(\cdot)$ is a well defined symmetric kernel density function, $\mathfrak{B}_k$'s are the regression coefficients to be solved to minimize (5.6), $h_{\mu,x}$ is the bandwidth selected for controlling the smoothing in the regression, and p is the order selected of the polynomial function. Similarly, we could obtain $\widehat{\mu}_y(z)$. With $\widehat{\mu}_x(z)$ and $\widehat{\mu}_y(z)$, we can obtain the squared residuals $\nu_{i,x}$ and $\nu_{j,y}$ in models (5.1) and (5.2) as follows:

$$\nu_{i,x} = \{x_i - \widehat{\mu}_x(z_{i,x})\}^2, \tag{5.7}$$

$$\nu_{j,y} = \{y_j - \widehat{\mu}_y(z_{j,y})\}^2. \tag{5.8}$$

With the observed squared residuals, we can obtain $\widehat{\nu}_y(z)$ via minimizing

$$\sum_{i=1}^{n_-}\{\nu_{i,x} - \sum_{k=0}^{p}\mathfrak{B}_k(z_{i,x} - z)^k\}^2\mathcal{K}(\frac{z_{i,x} - z}{h_{\nu,x}})/h_{\nu,x}, \tag{5.9}$$

where, $h_{\nu,x}$ is the bandwidth selected for obtaining $\widehat{\nu}_x(z)$. Similarly, we can obtain $\widehat{\nu}_y(z)$. The selection of the bandwidths for the mean and variance functions will be discussed in the simulation section.

Let $\widehat{c}_o(z)$ be the plug-in estimate of $c_o(z)$. Then, the covariate-adjusted estimator for the Youden index at given z can be defined as follows:

$$\widehat{J}_N(z) = \Phi\left(\frac{\widehat{\mu}_y(z) - \widehat{c}_o(z)}{\sqrt{\widehat{\nu}_y(z)}}\right) - \Phi\left(\frac{\widehat{\mu}_x(z) - \widehat{c}_o(z)}{\sqrt{\widehat{\nu}_x(z)}}\right). \tag{5.10}$$

5.2.3 Covariate-adjusted Youden Index without Normal Error Assumption

The covariate-adjusted Youden index in section 5.2.2 is a semi-parametric estimate for the Youden index based on regression models with normal error distribution assumption for test results. However, this method may be sensitive to departures from the distributional assumption. Therefore, it is necessary to provide a fully nonparametric covariate-adjustment for the Youden index.

In this section, ϵ_x and ϵ_y in models (5.1) - (5.2) are assumed to be distribution free, i.e., both $F_-^*(\cdot)$ and $F_+^*(\cdot)$ are unknown distributions. Our goal is to estimate $J(z)$ at given z based on these samples.

To estimate $J(z)$ at given z, we have to estimate test values at given $Z = z$ since the mean functions $\mu_x(z)$ and $\mu_y(z)$ and the variance functions $\nu_x(z)$ and $\nu_y(z)$ as well as the error distributions $F_-^*(\cdot)$ and $F_+^*(\cdot)$ are unknown. Estimating the mean and variance functions can be easily implemented by modern non-parametric methods (e.g., local polynomial method). However producing a good estimate for the error distribution is a difficult task in nonparametric heteroscedastic regression models. In stead of using the complex estimation of the error distributions, we employ the following procedure which has been used in Yao *et al.* [64].

1. Find non-parametric estimates $\widehat{\mu}_x$, $\widehat{\mu}_y$, $\widehat{\nu}_x$, and $\widehat{\nu}_y$ for μ_x, μ_y, ν_x, and ν_y by using local polynomial method.

2. Find the standardized residuals:

$$\widehat{\epsilon}_{i,x} = \frac{x_i - \widehat{\mu}_x(z_{i,x})}{\sqrt{\widehat{\nu}_x(z_{i,x})}}, \quad \widehat{\epsilon}_{j,y} = \frac{y_j - \widehat{\mu}_y(z_{j,y})}{\sqrt{\widehat{\nu}_y(z_{j,y})}}.$$

3. Estimate test values at given $Z = z$ as follows:

$$\widehat{x}_{i,z} = \widehat{\mu}_x(z) + \sqrt{\widehat{\nu}_x(z)}\widehat{\epsilon}_{i,x}, \quad \widehat{y}_{j,z} = \widehat{\mu}_y(z) + \sqrt{\widehat{\nu}_y(z)}\widehat{\epsilon}_{j,y}.$$

Then, the nonparametric covariate-adjusted estimator for the Youden index can be defined as follows:

$$\begin{aligned}\widehat{J}_E(z) &= \max_c \left[n_-^{-1} \sum_{i=1}^{n_-} I(\widehat{x}_{i,z} \le c) - n_+^{-1} \sum_{j=1}^{n_+} I(\widehat{y}_{j,z} \le c) \right] \\ &= n_-^{-1} \sum_{i=1}^{n_-} I(\widehat{x}_{i,z} \le \widehat{c}_{oE}(z)) - n_+^{-1} \sum_{j=1}^{n_+} I(\widehat{y}_{j,z} \le \widehat{c}_{oE}(z))\end{aligned}$$

where $I(\cdot)$ is the indicator function, $\widehat{c}_{oE}(z) = \widehat{c}_{oE}^{(1)}(z)$ or $\widehat{c}_{oE}^{(2)}(z)$ with

$$\widehat{c}_{oE}^{(1)}(z) = \inf_c \left\{ c : \max_c \left[n_-^{-1} \sum_{i=1}^{n_-} I(\widehat{x}_{i,z} \le c) - n_+^{-1} \sum_{j=1}^{n_+} I(\widehat{y}_{j,z} \le c) \right] \right\}, \tag{5.11}$$

$$\widehat{c}_{oE}^{(2)}(z) = \max_c \left\{ c : \max_c \left[n_-^{-1} \sum_{i=1}^{n_-} I(\widehat{x}_{i,z} \le c) - n_+^{-1} \sum_{j=1}^{n_+} I(\widehat{y}_{j,z} \le c) \right] \right\}. \tag{5.12}$$

$\widehat{c}_{oE}^{(i)}(z)$'s are the empirical estimates for the optimal cut-off point. It is noted that $\widehat{c}_{oE}^{(1)}(z)$ maximizes the empirical sensitivity and $\widehat{c}_{oE}^{(2)}(z)$ maximizes the empirical specificity. Choosing the empirical estimate in such a way, rather than determining it be the median of possible solutions, could reduce the computational burden in simulation.

5.3 Asymptotic Properties of the Covariate-adjusted Estimators for the Youden Index

We present the asymptotic properties of the covariate-adjusted estimators for the Youden index in this section. Firstly, we explore the asymptotic properties of $\widehat{J}_N(z)$ under the normal error assumption. Then we discuss the properties of $\widehat{J}_E(z)$ when the normality assumption for the error distribution is released.

5.3.1 Asymptotic Properties of $\widehat{J}_N(z)$

Before presenting the asymptotic properties of $\widehat{J}_N(z)$, let us introduce some necessary notations. Let

$$\begin{aligned}
\mathcal{E}_x(z) &= E(\epsilon_x^3|Z=z), \qquad \mathcal{E}_y(z) = E(\epsilon_y^3|Z=z),\\
\mathcal{V}_x(z) &= Var(\epsilon_x^2|Z=z), \qquad \mathcal{V}_y(z) = Var(\epsilon_y^2|Z=z),\\
\mathcal{M}_j(\mathcal{K}) &= \int \mu^j \mathcal{K}(\mu)d\mu, \text{ for all integer } j \geq 0,\\
\mathcal{R}(\mathcal{K}) &= \int \mathcal{K}^2(\mu) < \infty, \qquad S_p = (\mathcal{M}_{j+l}(\mathcal{K}))_{0\leq j,l\leq p}\\
\mathcal{K}^*(\mu) &= e_1^T S_p^{-1}(1,\mu,\ldots,\mu^p)^T\mathcal{K}(\mu),\\
\mathcal{R}(\mathcal{K}^*,\rho) &= \int \mathcal{K}^*(\mu)\mathcal{K}^*(\mu/\rho)d\mu, \text{ for any } 0<\rho<\infty,
\end{aligned}$$

where e_k is the $(p+1)\times 1$ vector with the k-th element being 1 and 0 others, and p is the chosen order in local polynomial regression estimation.

Under some regularity assumptions, for a given z, the local polynomial estimators of the mean and variance functions in model (5.1) and (5.2) are uniformly consistent with rates $O(\mathbb{R}_{xn_-})$ and $O(\mathbb{R}_{yn_+})$, respectively (see Yao *et al.* [64]),

$$\sup_{z\in\mathcal{D}(Z)} |\widehat{\mu}_x(z)-\mu_x(z)| = O(\mathbb{R}_{xn_-}) \ a.s., \quad \sup_{z\in\mathcal{D}(Z)} |\widehat{\nu}_x(z)-\nu_x(z)| = O(\mathbb{R}_{xn_-}) \ a.s., \tag{5.13}$$

$$\sup_{z\in\mathcal{D}(Z)} |\widehat{\mu}_y(z)-\mu_y(z)| = O(\mathbb{R}_{yn_+}) \ a.s., \quad \sup_{z\in\mathcal{D}(Z)} |\widehat{\nu}_y(z)-\nu_y(z)| = O(\mathbb{R}_{yn_+}) \ a.s., \tag{5.14}$$

where $\mathbb{R}_{xn_-} = h_{\mu,x}^{p+1}+\sqrt{log(1/h_{\mu,x})/(n_-h_{\mu,x})}$, $\mathbb{R}_{yn_+} = h_{\mu,y}^{p+1}+\sqrt{log(1/h_{\mu,y})/(n_+h_{\mu,y})}$, $h_{\mu,x}$ and $h_{\mu,y}$ are bandwidth for estimating $\mu_x(z)$ and $\mu_y(z)$, and $\mathcal{D}(Z)$ is the set of possible values of Z. Furthermore, Yao *et al.* [64]) showed that they are asymptotic normal, namely

$$\sqrt{n_-h_{\mu,x}}\,(\widehat{\mu}_x(z)-\mu_x(z),\widehat{\nu}_x(z)-\nu_x(z))^T \xrightarrow{d} N(\mathcal{B}_x(z),\Sigma_x(z)), \tag{5.15}$$

$$\sqrt{n_+h_{\mu,y}}\,(\widehat{\mu}_y(z)-\mu_y(z),\widehat{\nu}_y(z)-\nu_y(z))^T \xrightarrow{d} N(\mathcal{B}_y(z),\Sigma_y(z)), \tag{5.16}$$

where

$$
\begin{aligned}
\mathcal{B}_x(z) &= \{b_{x1}(z), b_{x2}(z)\}^T, & \Sigma_x(z) &= \sigma_{x,ij}(z)_{1\le i,j\le 2},\\
b_{x1} &= \frac{\mathcal{M}_{p+1}\mathcal{K}^*}{(p+1)!} d_x \mu_x^{p+1}(z), & b_{x2} &= \frac{\mathcal{M}_{p+1}\mathcal{K}^*}{(p+1)!} d_x \rho_x^{p+1} \nu_x^{p+1}(z),\\
\sigma_{x,11}(z) &= \frac{\mathcal{R}(\mathcal{K}^*)\nu_x(z)}{\theta(z)}, & \sigma_{x,22}(z) &= \frac{\mathcal{R}(\mathcal{K}^*)\mathcal{V}_x(z)}{\theta(z)\rho_x},\\
\sigma_{x,12}(z) &= \sigma_{x,21}(z) = \frac{\mathcal{R}(\mathcal{K}^*,\rho_x)\mathcal{E}_x(z)}{\theta(z)\rho_x}, & d_x &= \lim \left(n_- h_{\mu,x}^{2p+3}\right)^{1/2}\\
\mathcal{B}_y(z) &= \{b_{y1}(z), b_{y2}(z)\}^T, & \Sigma_y(z) &= \sigma_{y,ij}(z)_{1\le i,j\le 2},\\
b_{y1} &= \frac{\mathcal{M}_{p+1}\mathcal{K}^*}{(p+1)!} d_y \mu_y^{p+1}(z), & b_{y2} &= \frac{\mathcal{M}_{p+1}\mathcal{K}^*}{(p+1)!} d_y \rho_y^{p+1} \nu_y^{p+1}(z),\\
\sigma_{y,11}(z) &= \frac{\mathcal{R}(\mathcal{K}^*)\nu_y(z)}{\theta(z)}, & \sigma_{y,22}(z) &= \frac{\mathcal{R}(\mathcal{K}^*)\mathcal{V}_y(z)}{\theta(z)\rho_y},\\
\sigma_{y,12}(z) &= \sigma_{y,21}(z) = \frac{\mathcal{R}(\mathcal{K}^*,\rho_y)\mathcal{E}_y(z)}{\theta(z)\rho_y}, & d_y &= \lim \left(n_+ h_{\mu,y}^{2p+3}\right)^{1/2},\\
\rho_i &= \lim h_{\nu_i}/h_{\mu,i},\ i = x, y, \text{ and } h_{\nu_i} \text{ is the bandwidth for estimating } \nu_i,
\end{aligned}
$$

and $\theta(z)$ represents the probability density function of Z, if the covariates are treated as a random variable.

Based on above asymptotic properties, utilizing the Cramér-Wold device and Slutsky's theorem, we obtain the following theorems for $\widehat{J}_N(z)$.

Theorem 5.1. *Under assumptions (E1)-(E5) stated in Appendix E, for a given $Z = z$, we have that*

(i) if $\frac{n_+}{n_-} \to \infty$, $\sqrt{n_- h_{\mu,x}}(\widehat{J}_N(z) - J_N(z)) \xrightarrow{d} N(M_1(z), V_1(z))$, where

$$
\begin{aligned}
M_1(z) &= \frac{\partial J_N(z)}{\partial \mu_x(z)} b_{x1}(z) + \frac{\partial J_N(z)}{\partial \nu_x(z)} b_{x2}(z),\\
V_1(z) &= \left(\frac{\partial J_N(z)}{\partial \mu_x(z)}\right)^2 \sigma_{x,11}(z) + \left(\frac{\partial J_N(z)}{\partial \nu_x(z)}\right)^2 \sigma_{x,22}(z)\\
&\quad + \sigma_{x,12}(z)\left(\frac{\partial J_N(z)}{\partial \nu_x(z)}\frac{\partial J_N(z)}{\partial \mu_x(z)} + \frac{\partial J_N(z)}{\partial \mu_x(z)}\frac{\partial J_N(z)}{\partial \nu_x(z)}\right).
\end{aligned}
$$

(ii) if $\frac{n_+}{n_-} \to 0$, $\sqrt{n_+ h_{\mu,y}}(\widehat{J}_N(z) - J_N(z)) \xrightarrow{d} N(M_2(z), V_2(z))$, *where*

$$\begin{aligned} M_2(z) &= \frac{\partial J_N(z)}{\partial \mu_y(z)} b_{y1}(z) + \frac{\partial J_N(z)}{\partial \nu_y(z)} b_{y2}(z), \\ V_2(z) &= \left(\frac{\partial J_N(z)}{\partial \mu_y(z)}\right)^2 \sigma_{y,11}(z) + \left(\frac{\partial J_N(z)}{\partial \nu_y(z)}\right)^2 \sigma_{y,22}(z) \\ &\quad + \sigma_{y,12}(z)\left(\frac{\partial J_N(z)}{\partial \nu_y(z)}\frac{\partial J_N(z)}{\partial \mu_y(z)} + \frac{\partial J_N(z)}{\partial \mu_y(z)}\frac{\partial J_N(z)}{\partial \nu_y(z)}\right), \end{aligned}$$

(iii) if $\frac{n_+}{n_-} \to \lambda$, $0 < \lambda < \infty$, $\sqrt{n_- h_{\mu,x}}(\widehat{J}_N(z) - J_N(z)) \xrightarrow{d} N(M_3(z), V_3(z))$, *where*

$$M_3(z) = M_1(z) + CM_2(z), \quad V_3(z) = V_1(z) + C^2 V_2(z),$$

and $C = \lambda^{-\frac{p+1}{2p+3}}(\frac{d_x}{d_y})^{\frac{1}{2p+3}}$. The detail of the partial derivatives are listed in Appendix E.

Theorem 5.2. *Under assumptions* $(E1\dagger)$*-*$(E5\dagger)$ *and (E6)-(E8) stated in Appendix E, we have*

$$\sup_{z \in \mathcal{D}(Z)} |\widehat{J}_N(z) - J_N(z)| = O(\mathbb{R}_{xn_-} + \mathbb{R}_{yn_+}). \tag{5.17}$$

5.3.2 Asymptotic Properties of $\widehat{J}_E(z)$

Now we explore the asymptotic properties of the empirical estimate $\widehat{J}_E(z)$ of $J(z)$ without normality assumption.

Let

$$\epsilon_{i,x} = \frac{x_i - \mu_x(z_{i,x})}{\sqrt{\nu_x(z_{i,x})}}, \quad \epsilon_{j,y} = \frac{y_j - \mu_y(z_{j,y})}{\sqrt{\nu_y(z_{j,y})}}.$$

and

$$x_{i,z} = \mu_x(z) + \sqrt{\nu_x(z)}\epsilon_{i,x}, \quad y_{j,z} = \mu_y(z) + \sqrt{\nu_y(z)}\epsilon_{j,y}.$$

$$\begin{aligned} \tilde{J}_E(z) &= \max_c \left[n_-^{-1} \sum_{i=1}^{n_-} I(x_{i,z} \le c) - n_+^{-1} \sum_{j=1}^{n_+} I(y_{j,z} \le c) \right] \\ &= n_-^{-1} \sum_{i=1}^{n_-} I(x_{i,z} \le \tilde{c}_{oE}(z)) - n_+^{-1} \sum_{j=1}^{n_+} I(y_{j,z} \le \tilde{c}_{oE}(z)), \end{aligned}$$

where $\tilde{c}_{oE}(z) = \tilde{c}_{oE}^{(1)}(z)$ or $\tilde{c}_{oE}^{(2)}(z)$, and

$$\tilde{c}_{oE}^{(1)}(z) = \inf_c \left\{ c : \max_c \left[n_-^{-1} \sum_{i=1}^{n_-} I(x_{i,z} \le c) - n_+^{-1} \sum_{j=1}^{n_+} I(y_{j,z} \le c) \right] \right\}, \tag{5.18}$$

$$\tilde{c}_{oE}^{(2)}(z) = \max_c \left\{ c : \max_c \left[n_-^{-1} \sum_{i=1}^{n_-} I(x_{i,z} \le c) - n_+^{-1} \sum_{j=1}^{n_+} I(y_{j,z} \le c) \right] \right\}. \tag{5.19}$$

$\tilde{J}_E(z)$ can be treated as a "hypothetical" estimator for $J(z)$ because the mean functions $\mu_x(z)$, $\mu_y(z)$, $\nu_x(z)$, and $\nu_y(z)$ need to be estimated in practice. If these mean functions and variance functions are known, $\widehat{J}_E(z)$ is an asymptotically unbiased estimator for $J(z)$.

Theorem 5.3. *If $n_+/n_- \to \lambda$ for some $0 < \lambda < \infty$, then*

$$E\left[\tilde{J}_E(z)\right] \longrightarrow J(z), \quad \textit{for a given } z. \tag{5.20}$$

Theorem 5.4. *Under the same assumptions for Theorem 5.2 and (E9) stated in Appendix E, if $n_+/n_- \to \lambda$ for some $0 < \lambda < \infty$, then*

$$E\left[\left(\widehat{J}_E(z) - \tilde{J}_E(z)\right)^2\right] \longrightarrow 0, \quad \textit{for a given } z. \tag{5.21}$$

The asymptotic unbiasness of $\widehat{J}_E(z)$ can be obtained from Theorem 5.3 and Theorem 5.4. But the asymptotic normality of the empirical estimator $\widehat{J}_E(z)$ has eluded us so far. It should be the future research for us.

5.4 Confidence Intervals for the Youden Index and Simulation Studies

5.4.1 Confidence Intervals for the Covariate-adjusted Youden Index

Under the normal error assumption for models (5.1) and (5.2), using the asymptotic distribution (from Theorem 5.1) of $\widehat{J}_N(z)$, we can construct a normal approximation-based confidence interval (NA interval) for the YI at given $Z = z$. However, it should be noticed that, from Appendix F, the partial derivatives of $J_N(z)$ with respect to the mean and variance functions in Theorem 1

involve too many unknown parameters, and the two quantities, $d_x = \lim \left(n_- h_{\mu,x}^{2p+3}\right)^{1/2}$ and $d_y = \lim \left(n_+ h_{\mu,y}^{2p+3}\right)^{1/2}$, are unknown. To avoid those complex plugging-ins, we use the bootstrap method to estimate the bias and variance of $\widehat{J}_N(z)$. At given $Z = z$, re-sample the original data B times to obtain B bootstrap replications of $\widehat{J}_N(z)$, notated as $\{\widehat{J}_N(z)^{*b} : b = 1, 2, \ldots, B\}$. Then, the bias of b $J_N(z)$ can be estimated by:

$$\widehat{M}_3^*(z) = \frac{1}{B}\sum_{b=1}^{B}(\widehat{J}_N(z)^{*b} - \widehat{J}_N(z)),$$

and the variance of $J_N(z)$ can be estimated by:

$$\widehat{V}_3^*(z) = \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{J}_N(z)^{*b} - \widehat{J}_N(z))^2.$$

At a given z, the $(1-\alpha)100\%$ normal approximation (NA) confidence interval for $J_N(z)$ is defined as follows:

$$\left(\widehat{J}_N(z) - \widehat{M}_3^*(z) - z_{1-\alpha/2}\sqrt{V_3^*(z)},\ \widehat{J}_N(z) - \widehat{M}_3^*(z) + z_{1-\alpha/2}\sqrt{V_3^*(z)}\right), \tag{5.22}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of the standard normal distribution.

Without the normal error assumption, the confidence interval for the Youden index at given $Z = z$ should be based on the nonparametric estimate $\widehat{J}_E(z)$. Since the asymptotic distribution of $\widehat{J}_E(z)$ is still unknown, we propose a nonparametric interval for $J(z)$ by using bootstrap method. Let

$$\widehat{J}_{AC}(z) \quad = \quad \frac{\sum_{i=1}^{n_-} I(\widehat{x}_{i,z} \le \widehat{c}_{oE}(z)) + 2}{n_- + 4} - \frac{\sum_{j=1}^{n_+} I(\widehat{y}_{j,z} \le \widehat{c}_{oE}(z)) + 2}{n_+ + 4}.$$

$\widehat{J}_{AC}(z)$ is inspired by Agresti and Coull's [1] interval estimate for a binomial proportion which has very good small sample performance. Since $z_{1-\alpha/2}$ is approximately equal to 2 when $\alpha = 0.05$, $\widehat{J}_{AC}(z)$ may be regarded as an adjusted estimate for the difference between two proportions (i.e., $P(X \le c_o(z)|Z = z)$ and $P(Y \le c_o(z)|Z = z)$) by adding two successes and two failures to the pseudo Bernoulli observations. We summarize the bootstrap procedure in the following steps:

(i). Draw a re-sample of size m, $\widehat{x}^*_{i,z}$'s, with replacement from $\widehat{x}_{i,z}$'s and a re-sample of size n, $\widehat{y}^*_{j,z}$'s, with replacement from $\widehat{y}_{j,z}$'s.

(ii). Calculate the bootstrap version of $\widehat{J}_{AC}(z)$

$$\widehat{J}^*_{AC}(z) = \frac{\sum_{j=1}^{n_-} I(\widehat{y}^*_{j,z} \leq \widehat{c}^*_{oE}(z)) + 2}{n_- + 4} - \frac{\sum_{i=1}^{n_+} I(\widehat{x}^*_{i,z} < \widehat{c}^*_{oE}(z)) + 2}{n_+ + 4},$$

where $\widehat{c}^*_{oE}(z)$ is the bootstrap version of $\widehat{c}_{oE}(z)$.

(iii). Repeat step (i) and step (ii) B times to obtain the set of bootstrap replications $\{\widehat{J}^{*b}_{AC}(z) : b = 1, 2, \ldots, B\}$ (it is suggested that $B \geq 200$).

Then, the bootstrap variance estimator $Var(\widehat{J}^*_{AC}(z))$ is defined as

$$Var(\widehat{J}^*_{AC}(z)) = \frac{1}{B-1} \sum_{b=1}^{B} (\widehat{J}^{*b}_{AC}(z) - \bar{J}^*_{AC}(z))^2$$

where $\bar{J}^*_{AC}(z) = \frac{1}{B} \sum_{b=1}^{B} \widehat{J}^{*b}_{AC}(z)$.

Now the ACNA interval for $J(z)$ is defined as follows:

$$\left(\bar{J}^*_{AC}(z) - z_{1-\alpha/2} \sqrt{Var(\widehat{J}^*_{AC}(z))},\ \bar{J}^*_{AC}(z) + z_{1-\alpha/2} \sqrt{Var(\widehat{J}^*_{AC}(z))} \right).$$

5.4.2 Simulation Studies

In this section, we conduct simulation studies to examine the finite sample performances of the proposed methods for estimating the Youden index with adjustment for covariates. In the study, we utilize two sets of models to evaluate the efficiency of our methods.

In the first situation, we consider the following models for the healthy population and diseased population:

$$\begin{aligned} X|Z &= 6 + 1.5Z + 1.5\sin(Z) + \sqrt{0.4 + \Phi(2Z-6)}\epsilon_x, \\ Y|Z &= 7.2 + 1.5Z + 1.5\sin(Z) + \sqrt{Z - 0.8} + \sqrt{1.2 + \Phi(2Z-6)}\epsilon_y, \end{aligned}$$

where both ϵ_x and ϵ_y follow the standard normal distribution, and Φ is the c.d.f. of standard normal distribution. The simulated observations $\{x_i, z_{i,x}\}$ and $\{y_j, z_{j,y}\}$ for the two populations

are generated by drawing Z values from $Uniform(1,5)$ independently and drawing the errors from $N(0,1)$ independently, where $i = 1, \ldots, n_-$ and $j = 1, \ldots, n_+$. We choose the sample sizes to be $n_+ = n_- = 50$ and $n_+ = n_- = 100$ to compare performances of the methods at smaller sample size and larger sample size.

In the second situation, we assume the models for the non-diseased and the diseased populations as follows:

$$\begin{aligned} X|Z &= 6 + 1.5Z + 1.5\sin(Z) + \sqrt{0.4 + \Phi(2Z-6)}\epsilon_x, \\ Y|Z &= 8 + 1.5Z + 1.5\sin(Z) + \sqrt{Z - 0.5} + \sqrt{1.5 + \Phi(2Z-6)}\epsilon_y, \end{aligned}$$

where ϵ_x, ϵ_y follow heavy tail symmetric distribution, namely, the student t-distribution with degree of freedom 4. The purpose of this setting is to evaluate the performances of the methods when the underlying distributions are miss-specified.

In the simulation study, we fix the order in local polynomial regression to be 1, namely $p = 1$, to implement the local linear approximations. For the bandwidth selection, Cleveland and Loader [8] mentioned that we can either select fixed bandwidth for all observations in the domain or select varied bandwidths as a function of the observations. Here, we follow the second method to select varied bandwidths, and it is called the nearest-neighborhood bandwidth selection. If the sample size is notated by n, we select a fixed number of observations, which is $\alpha \times n$ and round it up to the nearest integer k_α, in the nearest-neighborhood of each observation to implement the local approximation, where α $(0 < \alpha < 1)$, the smoothing parameter, stands for a fixed proportion of observations. Then, at each observation, the corresponding half bandwidth is the distance from that observation to the nearest k_α-th observation. Accordingly, we can select varied bandwidths via selecting a fixed α. Choosing the sequences of varied bandwidths, $h_{\mu,x}$, $h_{\mu,y}$ $h_{\nu,x}$ and $h_{\nu,y}$ for estimating $\mu_x(z)$, $\mu_y(z)$, $\nu_x(z)$, and $\nu_y(z)$ costs lots of the computing resource. It is an extremely heavy duty for us to use the optimal varied bandwidths if we target at the Youden index. Alternatively, we selected a reasonable path to access the "optimal" varied bandwidths by minimizing the true integrated squared errors. We select varied $h_{\mu,x}$ as the values which minimize $\int [\widehat{\mu}_x(z; h_{\mu,x}) - \mu_x(z)]^2 dz$, and choose $h_{\mu,y}$, $h_{\nu,y}$, and $h_{\nu,y}$ similarly.

With the generated data, we evaluate the performances of the estimator $\widehat{J}_N(z)$ under normal assumption and the nonparametric estimators $\widehat{J}_E(z)$ and $\widehat{J}_{AC}(z)$ by reporting the mean square error (MSE) at given covariate values. We repeat the simulation for each setting 500 times to calculate MSE at different value of z. From Figure F.1, we observe that $\widehat{J}_{AC}(z)$ has the smallest MSE among the three estimators. When sample size increases, the MSE's of all estimators decrease as expected, shown in Figure F.1 (right). For the second model, which assumed the t-distribution for the errors, as expected, the MSE of $\widehat{J}_N(z)$ is significantly larger than those of $\widehat{J}_E(z)$ and $\widehat{J}_{AC}(z)$ (see Figure F.2).

We also examine the 95% level point-wise NA and ACNA confidence intervals for $J(z)$. The usual bootstrap percentile (BP) confidence interval based on the empirical estimator $\widehat{J}_E(z)$ is also included in the comparison. In the simulation study, we calculate the average upper bounds and the average lower bounds of these confidence intervals at given z from 500 Monte Carlo runs, in each Monte Carlo run, we bootstrap the original sample 999 times to obtain more accurate estimate. Figures F.3 and F.4 display the point-wise average confidence bands for $J(z)$. From these figures, we found that, even if the error distribution is correctly specified (in the first simulation setting), the point-wise ACNA band is competitive to the NA band. If the error distribution is mis-specified, the ACNA band outperforms the NA band.

5.5 A Real Application

In this section, we consider the Pima Indians Diabetes Study data originally discussed by Smith *et al.* [53]. In the dataset, nine variables are recorded: Number of times pregnant (V_1), Plasma glucose concentration in an OGTT (V_2), Diastolic blood pressure (mm Hg) (V_3), Triceps skin fold thickness (mm) (V_4), 2-Hour serum insulin (mu U/ml) (V_5), Body mass index ($weightinkg/(heightinm)^2$) (V_6), Diabetes pedigree function (V_7), Age (years) (V_8), Class variable (0 or 1) (V_9). There were 268 cases and 500 controls. Two individuals in the case group had OGTT value 0 and three individuals had OGTT value 0, we deleted these five observations in the data analysis. The OGTT is a classical and standard diagnostic test for Diabetes. Smith and Thompson [54] considered the age as a potential covariate which would influence the OGTT results.

First, we consider the situation without covariate adjustment. The OGTT results from case and control groups are not normally distributed based on the Pearson chi-square test for normality (p-value = 0.001, 0.023 respectively). The empirical estimate for the Youden index is $J_E = 0.446$. This estimated Youden index value indicates that the ability of OGTT for distinguishing diabetes is mediocre.

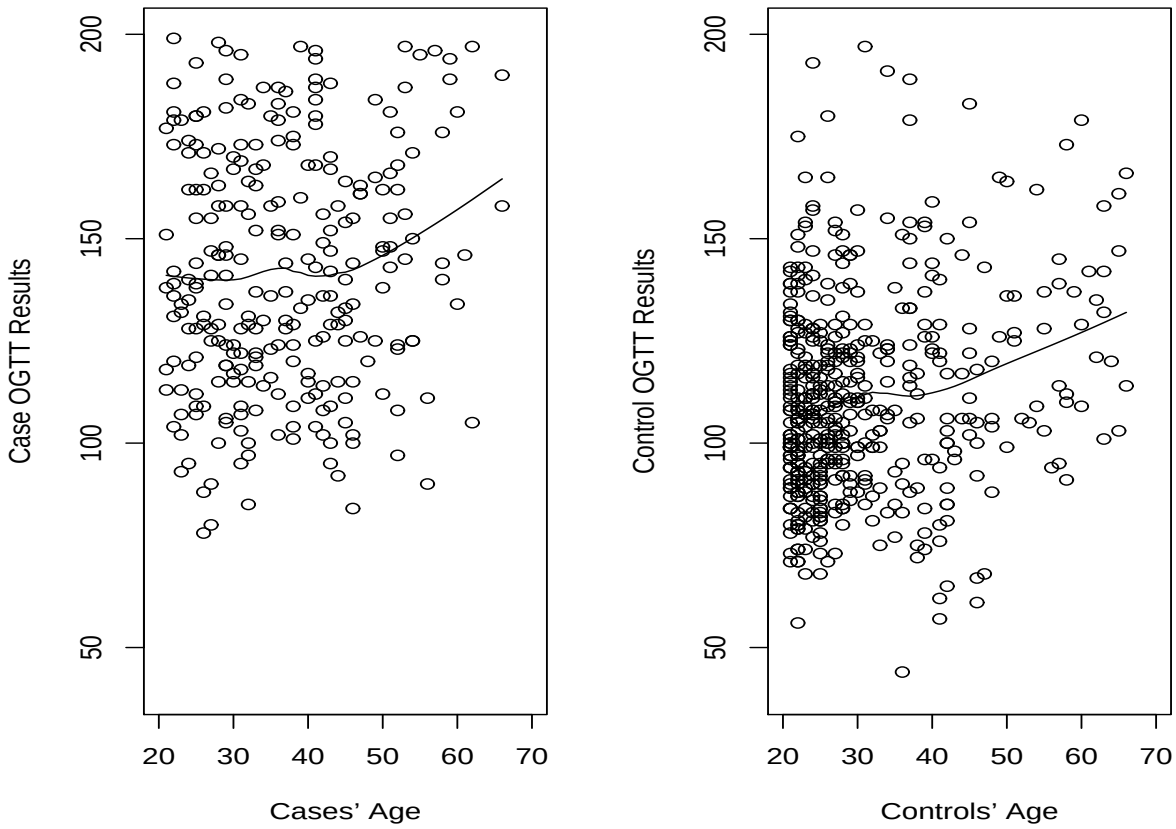

Figure 5.1. The scatter plot of OGTT test vs. Age, left for cases, right for controls. Solid lines are local polynomial estimates for the mean functions.

Now we consider the effect of age in estimating the Youden index. The scatter plots of the OGTT results vs. age among non-diseased and diseased groups (see Figure 5.1) do not indicate a strong linear relationship between OGTT and age. They also indicate variabilities of the OGTT are huge at observed age points (from the scatter plot). Consequently, the linear regression models employed in Faraggi [13] can not be directly applied here. However, the heteroscedastic regression

models (5.1) and (5.2) could work for this data set. Here, we use the OGTT results of subjects aged from 21 years to 66 years, and produce three covariate-adjusted estimates $\widehat{J}_N(z)$, $\widehat{J}_E(z)$ and $\widehat{J}_{AC}(z)$ for the Youden index with 95% point-wise BP band and ACNA band. In this application, the sequences of bandwidths for the mean and the variance functions of the two populations are selected by standard leaving-one-out cross-validation method, targeting on minimizing the prediction errors. We obtain the point-wise confidence intervals via bootstrap the original sample 999 times. From Figure 5.2, it is obvious that the accuracy of diagnosing diabetes by testing the glucose level in blood varies by age. The diagnostic accuracy of OGTT for younger individuals (age $<$ 30 years) could be more precise than that for individuals aged from 30 years to 35 years. There is a small spike which shows a slightly increasing accuracy for 38 years to 40 years old individuals, and then the accuracy decreases slowly to about 50 years. When testing individuals are getting older (age $>$ 50 years), the accuracy of OGTT increases. The confidence bands become wider when ages getting larger, this probably is due to the sparseness of observations with age larger than 60. The differences between the three proposed estimates are not obvious. However, we recommend the nonparametric covariate-adjusted estimates for the Youden index to this data set because it is more flexible and robust than the one with normal error assumption.

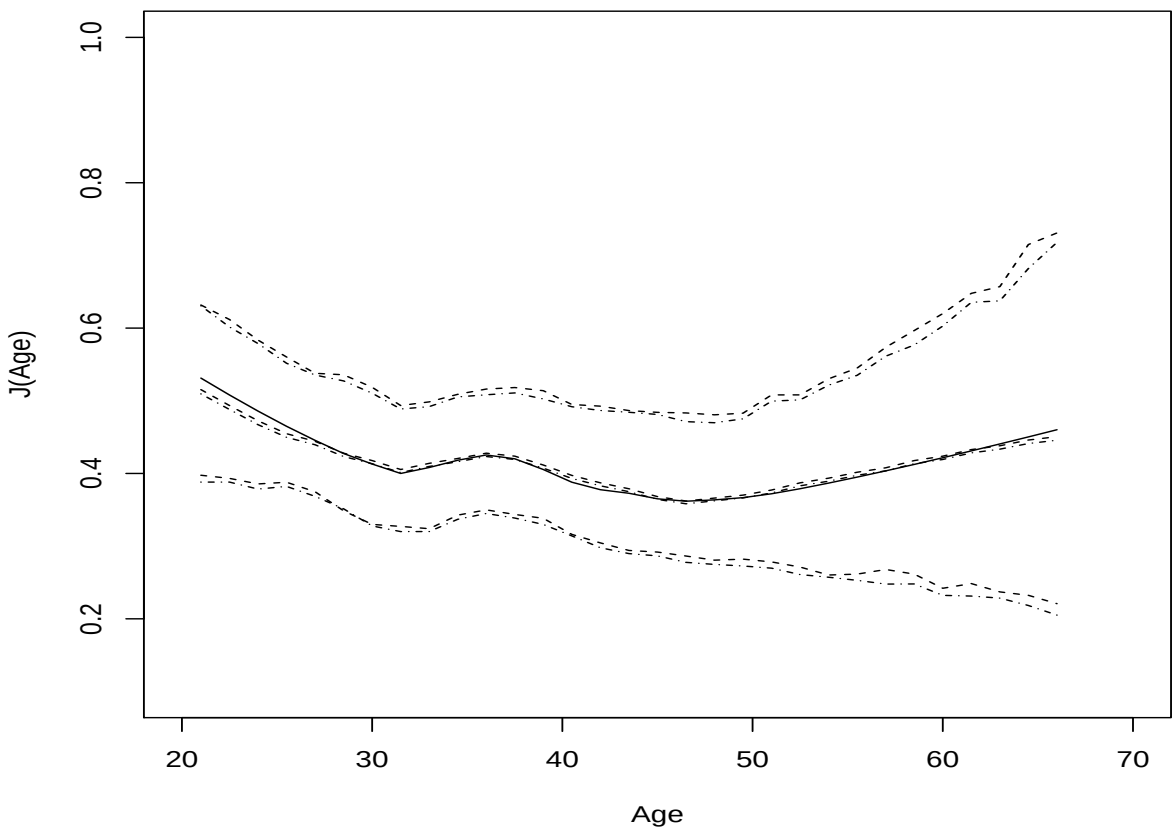

Figure 5.2. Nonparametric estimates for $J(Age)$: $\widehat{J}_N$ (solid), $\widehat{J}_E$ (dashed), and $\widehat{J}_{AC}$ (dotdash). Point-wise confidence bands for $J(Age)$: BP band (dashed), ACNA band (dotdash).

Chapter 6

CONCLUSIONS AND FUTURE WORK

The Youden index, a frequently used summary of the ROC curve, has been found in many applications in different fields, such as medical informatics and bioinformatics. It provides a criterion for evaluating the optimal threshold value of a test for which the summation of sensitivity and specificity is maximized.

In Chapter 2, we introduced a new point of view, the diagnostic curve, to assess the accuracy of medical tests. The proposed point-wise confidence intervals are easy to implement with satisfying performances in most settings, and they are purely nonparametric. In practice, we highly recommend utilizing the diagnostic curve to evaluate the performances of diagnostic tests.

In Chapter 3, we proposed and compared three confidence intervals for the difference between two Youden indices of paired diagnostic tests. When the underlying distributions are normal, the GPQ-based exact confidence interval shows perfect performance even when sample size are extremely small ($m, n \leq 20$). The hybrid bootstrap and maximum likelihood based confidence interval has acceptable performance when sample sizes are large enough ($m, n \geq 50$). While the application of the GPQ-based exact interval depends on the underlying normal distribution assumption, our simulation results show that it has some robust properties against the deviation from the normal distribution assumption. The proposed nonparametric BP interval generally does not have satisfactory performance.

In Chapter 4, we defined the Youden index for three ordinal diagnostic groups and proposed its estimates, examined the nonparametric asymptotic property, and proposed several confidence interval estimations for the Youden index. According to the simulation outcomes, we recommend the proposed parametric confidence intervals when the underlying distributions are normal. Although the AC adjusted normal approximation confidence interval is sensitive to the true value of the Youden index, it is robust and can offer acceptable performances for lots of cases.

In Chapter 5, we proposed nonparametric covariate-adjusted estimates for the Youden index. The simulation study conducted here demonstrated the robustness and effectiveness of the proposed

method, hereby, we suggest applying the nonparametric approach in real applications. Although some asymptotic properties of the nonparametric covariate-adjusted estimator for the Youden index have been obtained, its asymptotic distribution is still an open question.

With multiple tests, or high dimensional covariates, only the adjustment for the accuracy could not fully satisfy the current medical questions. For example, what are the significant variables or tests which play critical roles in helping detecting the diseases? To answer such questions, we need to explore or establish appropriate methodologies to implement the model selection, which is targeting on maximizing the accuracy of the diagnostic tests. Currently, relative researches undergoing focused on the ROC curve (Lin *et al.* [33]), or AUC (Huang *et al.* [24]); however, few of them focused on the Youden index. Perspectively, researches of covariates selection based on the Youden index are of the primary interests.

REFERENCES

[1] Agresti, A., and Coull, B.A., Approximate is better than 'exact' for interval estimation of binomial proportions. *American Statistician*, Vol. 27, pp. 119–126, 1998.

[2] Agresti, A. and Caffo, B., Simple and effective confidence itnervals for proportions and differences of proportions result from adding two successes and two failures. *The Amrican Statistician*, Vol. 54, pp. 280-288, 2000.

[3] Aoki, K., Misumi, J., Kimura, T., Zhao, W., and Xie, T., Evaluation of cutoff levels for screening of gastric cancer using serum pepsinogens and distributions of levels of serum pepsinogens I, II and of PG I/PG II ratios in a gastric cancer case-control study. *Journal of Epidemiology*, Vol. 7, pp. 143–151, 1997.

[4] Braak, H. and Braak, E., Neuropathologic staging of Alzheimer-related changes. *Acta Neuropathologica*, Vol. 82, pp. 239–259, 1991.

[5] Carpenter, J., and Bithell, J., Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, Vol. 19, pp. 1141–1164, 2000.

[6] Castle, P., Lorincz, A.T., Scott, D.R., Sherman, M.E., Glass, A.G., Rush, B.B., Wacholder, S., Burk, R.D., Manos, M.M., Schussler, J.E., Macomber, P., and Schiffman, M., Comparison between prototype hybrid capture 3 and hybrid capture 2 human papillomavirus DNA assays for detection of high-grade cervical interepithelial neoplasia and cancer. *Journal of Clinical Microbiology*, Vol. 9, pp. 4022–4030, 2003

[7] Chernoff, H., Estimation of the mode. *Annals of the Institute of Statistical Mathematics*, Vol. 16, pp. 31-41, 1964.

[8] Cleveland, W.S., and Loader, C. L., Smoothing by local regression: Principles and methods. In Statistical Theory and Computational Aspects of Smoothing, pp. 10–49. *Springer: New York*, 1996

[9] Csörgö, M. and Révész, P., Strong Approximation in Probablity and Statistics. *Academic Press: New York*, 1981.

[10] Demir, A., Yarali, N., Fisgin, T., Duru, F., and Kara, A., Most reliable indices in differentiation between thalassemial trait and iron deficiency anemia. *Pediatrics International*, Vol. 44, pp. 612–616, 2002.

[11] Dodd, L. and Pepe, M.S., Semi-parametric regression for the area under the Receiver Operating Characteristic Curve. *Journal of the Acoustical Society of America*, Vol. 98, pp. 409–417, 2003.

[12] Fan, J. and Gijbels, I., Local Polynomial Modeling and Its Applications. *Chapman & Hall: London*, 1996.

[13] Faraggi, D., Adjusting receiver operating curves and related indices for covariates. *The Statistician*, Vol. 52, pp. 179–192, 2003.

[14] Fluss, R., Faraggi, D., and Reiser, B., Estimation of the Youden index and its associated cutoff point. *Biometrical Journal*. Vol. 47, pp. 458–472, 2005.

[15] Grmec, S. and Gasparovic, V., Comparison of APACHE II, MEES and Glasgow Coma Scale in patients with nontraumatic coma for prediction of mortality. *Critical Care*, Vol. 5, pp. 19–23, 2001.

[16] Gamage, J., Mathew, T., and Weerahandi, S., Generalized p-values and generalized confidence regions for the multivariate Behrens-Fisher problem and MANOVA. *Journal of Multivariate Analysis*, Vol. 88, pp. 177–189, 2004.

[17] Hall, P., On the removal of skewness by transformation. *Journal of the Royal Statistical Society, Series B*, Vol. 54, pp. 221-228, 1992.

[18] Hanning, J., Iyer, H.K., and Patterson, P., Fiducial generalized confidence intervals. *Journal of the American Statistical Association*, Vol. 101, pp. 254-269, 2006.

[19] Hawass, N-E-D., Comparing the sensitivities and specificities of two diagnostic procedures performed on the same group of patients *The British Journal of Radiology*, Vol. 70, pp. 360–366, 1997.

[20] He, X. and Frey, E.C. Three-class ROC analysis–the equal error utility assumption and the optimality of three-class ROC surface using the ideal observer. *IEEE Transactions on Medical Imaging*, Vol. 25, pp. 979–986, 2006.

[21] Hsieh, F.S. and Turnbull, W.B., Nonparametric methods for evaluating diagnostic tests. *School of Opertations, Cornell University*, Technical Report No. 1024, 1992.

[22] Hsieh, F. and Turnbull, W., Nonparametric methods for evaluating diagnostic tests. *Statistia Sinica*, Vol. 6, pp. 47-62, 1996.

[23] Hsueh, H., Liu, J., and Chen, J.J., Unconditional exact tests for equivalence or noninferiority for paired binary endpoints. *Biometrics*, Vol. 57, pp. 478–483, 2001.

[24] Huang, X., Qin, G., and Fang, Y., Optimal combinations of diagnostic tests based on AUC. *Biometrics* Vol. 67, pp. 568–576, 2011.

[25] Iyer, H.K., Wang, C., and Mathew, T., Models and confidence intervals for true values in interlaboratory trials. *Journal of the American Statistical Association*, Vol. 99, pp. 1060–1071, 2004.

[26] Kim, J. and Pollard, D., Cube root asymptotics. *Annals of Statistics*, Vol. 18, pp. 191–219, 1990.

[27] Kim, K., Maximization of the sum of sensitivity and specificity as a diagnostic cutpoint criterion. *Journal of Clinical Epidemiology*, Vol. 61, pp. 516–518, 2008.

[28] Lai, C-Y., Tian, L., and Schisterman, E.F., Exact confidence interval estimation for the Youden index and its corresponding optimal cut-point. *Computational Statistics & Data Analysis*, In Press, 2010.

[29] Le, C.T., A solution for the most basic optimization problem associated with ROC curve. *Statistical Methods in Medical Research*, Vol. 15, pp. 571–584, 2006.

[30] Li, C., Chen, T. and Liu, P., On the exact interval estimation for the difference in paired areas under the ROC curves. *Statistics in Medicine*, Vol. 27, pp. 224–242, 2008.

[31] Liao, C., and Iyer, H.K., A tolerance interval for the normal distribution with several variance components. *Statistica Sinica*, Vol. 14, pp. 217–229, 2004.

[32] Liao, C., Lin, T., and Iyer, H.K., One- and two-sided tolerance intervals for general balanced mixed models and unbalanced one-way random models. *Technometrics*, Vol. 47, pp. 323–335, 2005.

[33] Lin, H., Zhou, L., Peng, H., and Zhou, X-H., Selection and combination of biomarkers using ROC method for disease classification and prediction. *The Canadian Journal of Statistics* Vol. 39, pp. 324–343, 2011.

[34] Liu,J., Hsueh, H., Hsieh, E., and Chen, J.J., Tests for equivalence or non-inferiority for paired binary data. *Statistics in Medicine*, Vol. 21, pp. 231–245, 2002.

[35] Mckeel, Jr. D.W., Price, J.L., Miller, J.P., Grant, E.A., Xiong, C., Berg, L., and Morris, J.C., Neuropathologic criteria for diagnosing Alzheimer disease in persons with pure dementia fo Alzheimer type. *Journal of Neuropathology and Experimental Neuropathology*, Vol. 63, pp. 1028–1037, 2004.

[36] McNally, R.J., Iyer, H.K., and Mathew, T., Tests for individual and population bioequivalence based on generalized p-values. *Statistics in Medicine*, Vol. 22, pp. 31–53, 2003.

[37] Miller R.G., Efron, B., Brown, Jr B.W., and Moses, L.E., Biostatistics casebook. *John Wiley and Son: New York*, 1980.

[38] Mossman, D., Three-way ROCs. *Medical Decision Making*, Vol. 18, pp. 110-121, 1999.

[39] Nakas, C.T., Alonzo, T.A., nad Yiannoutsos, C.T., Accuracy and cut-off point selection in three-classification probelems using a generalization of the Youden index. *Statistics in Medicine*, Vol. 29, pp. 2946–2955, 2010.

[40] Obuchowski, N.A., Multireader multinodality receiver operating characteristic curver studies: Hypothesis testing and sample size estimation using an analysis of variance approach with dependent observations. *Academic Radiology*, Vol. 2, pp. 522–529, 1995.

[41] Pepe, M.S., A regression modeling framwork for reciever operating characteristics curves in medical diagnostic testing. *Biometrika*, Vol. 84, pp. 595–608, 1997.

[42] Pepe, M.S., Three approaches to regression analysis of reciever operating characteristic curves for continues test results. *Biometrics*, Vol. 54, pp. 124–135, 1998.

[43] Pepe, M.S., An interpretation for the ROC curve and inference using GLM procedures. *Biometrika*, Vol. 56,pp. 352–359, 2000.

[44] Pepe, M.S., The Statistical Evaluation of Medical Tests for Classification and Prediction. *Oxford University Press: New York*, 2003.

[45] Perkins, N., Schisterman, E.F., The Youden index and the optimal cut-point corrected for measurement error. *Biometrical Journal*, Vol. 47, pp. 428–441, 2005.

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

[47] Qin, G., Hsu, Y.-S., and Zhou, X.-H, New confidence intervals for the difference between two sensitivities at a fixed level of specificity. *Statistics in Medicine*, Vol. 25, pp. 3487–3502, 2006.

[48] Schisterman, E.F., and Perkins, N., Confidence interval for Youden index and corresponding optimal cut-point. *Communication in Statistics-Simulation and Computation*, Vol. 36, pp. 549–563, 2007.

[49] Schisterman, E.F., Faraggi, D., Reiser, B., and Hu, J., Youden index and the optimal threshold for markers with mass at zero. *Statistics in Medicine*, Vol. 27, pp. 297–315, 2008.

[50] Schisterman, E., Reiser, B., and Fraaggi, D., ROC analysis for markers with mass at zero. *Statistics in Medicine*, Vol. 23, pp. 623–638, 2006.

[51] Shao, J., Mathematical Statistics. 2nd ed. *Springer-Verlag: New York*, 2003.

[52] Shapiro, D.E., The interpretation of diagnostic tests. *Statistical Methods in Medical Research*, Vol. 8, pp. 113–134, 1998

[53] Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., and Johannes, R.S., Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. *Johns Hopkins APL Technical Digest*, Vol. 10, pp. 262–266, 1988.

[54] Smith, P.J. and Thompson, T.J., Correcting for confounding in analyzing receiver operating characteristic curves. *Biometrical Journal*, Vol. 38, pp. 857–863, 1996.

[55] Tang, N., Tang, M., and Chan, I.S.F., On tests of equivalence via non-unity relative risk for matched-pair design. *Statistics in Medicine*, Vol. 22, pp. 1217–1233, 2003.

[56] Thomas, E.A.C. and Myers, J., Implications of latency data for threshold and nonthreshold models of signal detection. *Journal of Mathematical Psychology*, Vol. 9, pp. 253–285, 1972.

[57] Thompson, M.L. and Zucchini, W., On the statistical analysis of ROC curves. *Statisitcs in Medicine*, Vol. 8, pp. 1277–1290, 1989.

[58] Toledano, A. and Gatsonis, C., Regression analysis of correlated reciever operating characteristic data. *Acdemic Radiology*, Vol. 2, pp. 530–536, 1995.

[59] Tosteson, A.N.A. and Begg, C.B., A general regression methodology for ROC curve estimation. *Medical Decision Making*, Vol. 8, pp. 204–215, 1988.

[60] Weerahandi, S., Generalized confidence intervals. *Journal of the American Statistical Association*, Vol. 88, pp. 899–905, 1993.

[61] Wieand, S., Gail, M., James, B.R., and James, K., A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*, Vol. 76, pp. 585–592, 1989.

[62] Xiong, C., Van Belle, G., Miller, J.P., and Morris, J.C., Measuring and estimating diagnostic accuracy when there are three oridnal diagnostic groups. *Statistics in Medicine*, Vol. 25, pp. 1251–1273, 2006.

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

[64] Yao, F., Craiu, R. V., and Reiser, B., Nonparametric covariate adjustment for receiver operating characteristic curves. *The Canadian Journal of Statisitcs*, Vol. 38, pp. 27–46, 2010.

[65] Yerli, H. Aydin, E. Haberal, N. Harman, A. Kaskati, T. and Alibek, S., Diagnosing common parotid tumours with magnetic resonance imaging including diffusion-weighted imaging vs fine-needle aspiration cytology: a comparative study. *Dento maxillo facial Radiology*, Vol. 6, pp. 349–355, 2010.

[66] Youden, W.J., Index for rating diagnostic tests. *Cancer*, Vol. 3, pp. 32–35, 1950.

[67] Zhou, H., and Qin, G., New non-parametric confidence intervals for the Youden index. *Journal of Biopharmaceutical Statistics*, In press, 2011.

[68] Zhou, X.-H., McClish, D.K., and Obuchowski, N.A., Statistical Methods in Diagnostic Medicine. *John Wiley & Sons, Inc.: New York,* 2002.

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

[70] Zhou, X.-H., Tsao, M., and Qin, G., New intervals for the difference between two independent binomial proportions. *Journal of Statistical Planning and Inference*, Vol. 123, pp. 97–115, 2004.

Appendix A

SIMULATION RESULTS FOR CHAPTER TWO

Notes:

- In each figure, "Normal-Normal" means that the underlying distributions for non-diseased and diseased populations are both normal distributions, respectively.
- In each figure, "Normal-T" means that the underlying distributions for non-diseased and diseased populations are normal distribution and t-distribution with degree of freedom 4, respectively.
- In the left panel of each figure, the solid line is the true $J(c)$. The dashed lines represent the Wald point-wise interval. The dotdash lines represent the AC point-wise interval. The long dash lines represent the TT point-wise interval.
- In the right panel of each figure, the solid line represents the the average coverage probability of the Wald point-wise interval. The dotdash line represents the average coverage probability of the AC point-wise interval. The long dash line represents the average coverage probability of the TT point-wise interval.

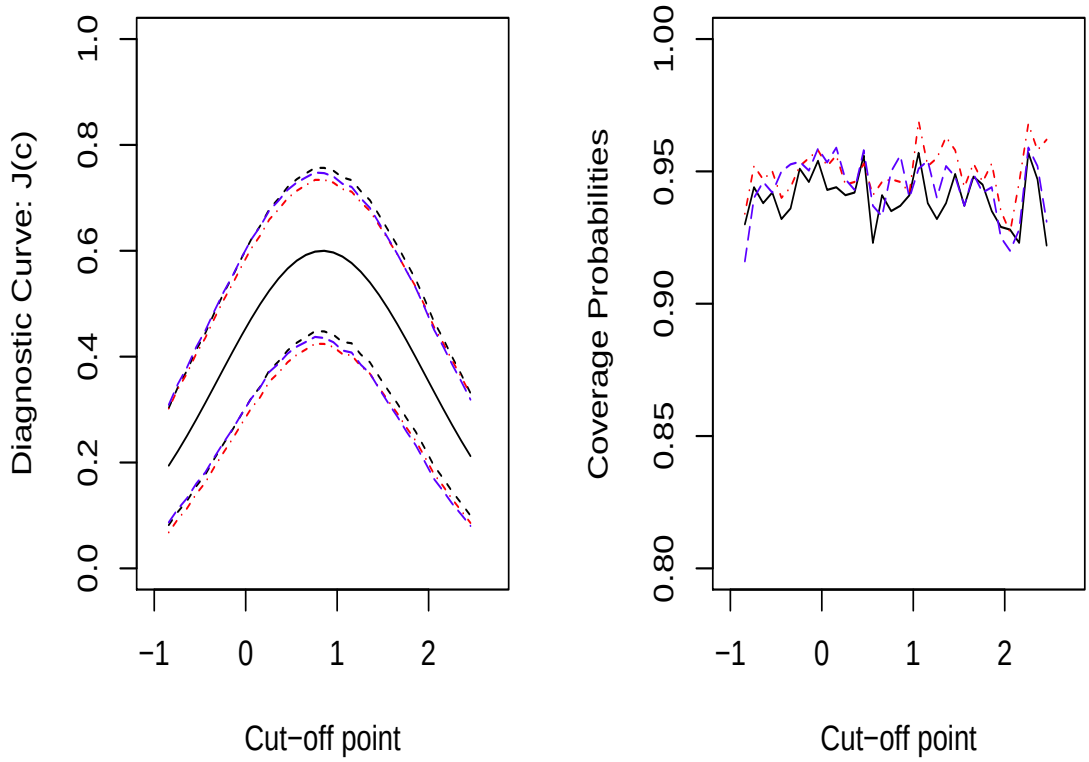

Figure A.1. Normal-Normal, Maximum of DC is 0.60. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability.

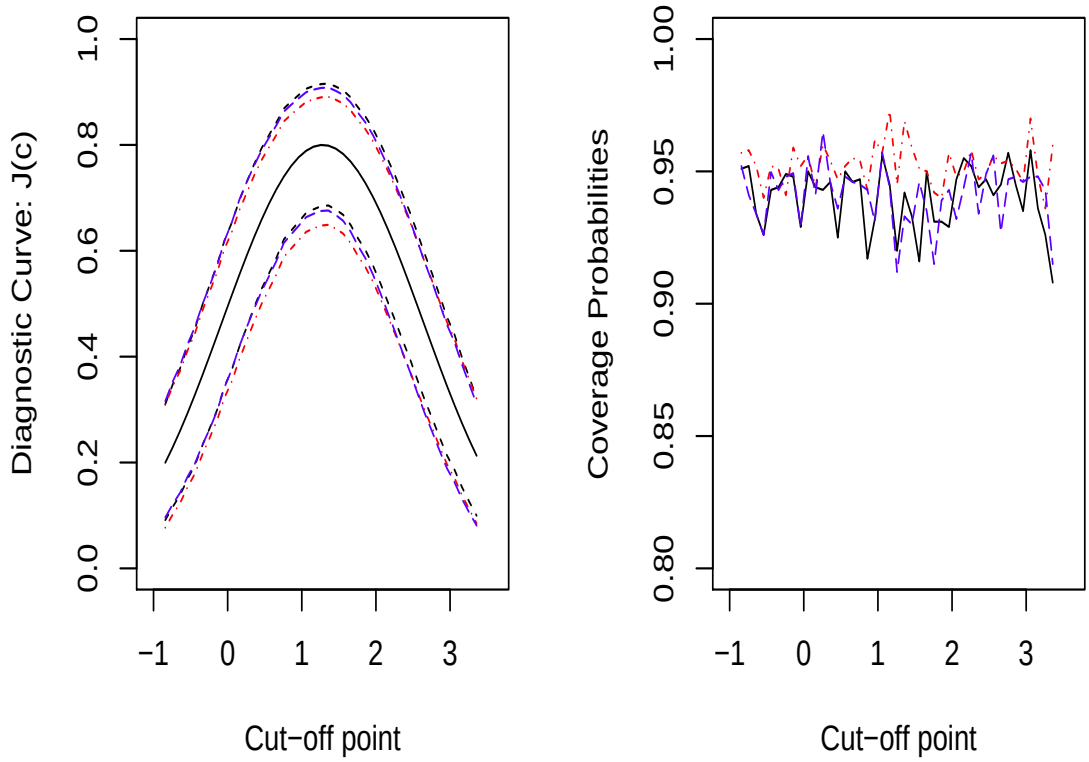

Figure A.2. Normal-Normal, Maximum of DC is 0.80. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability.

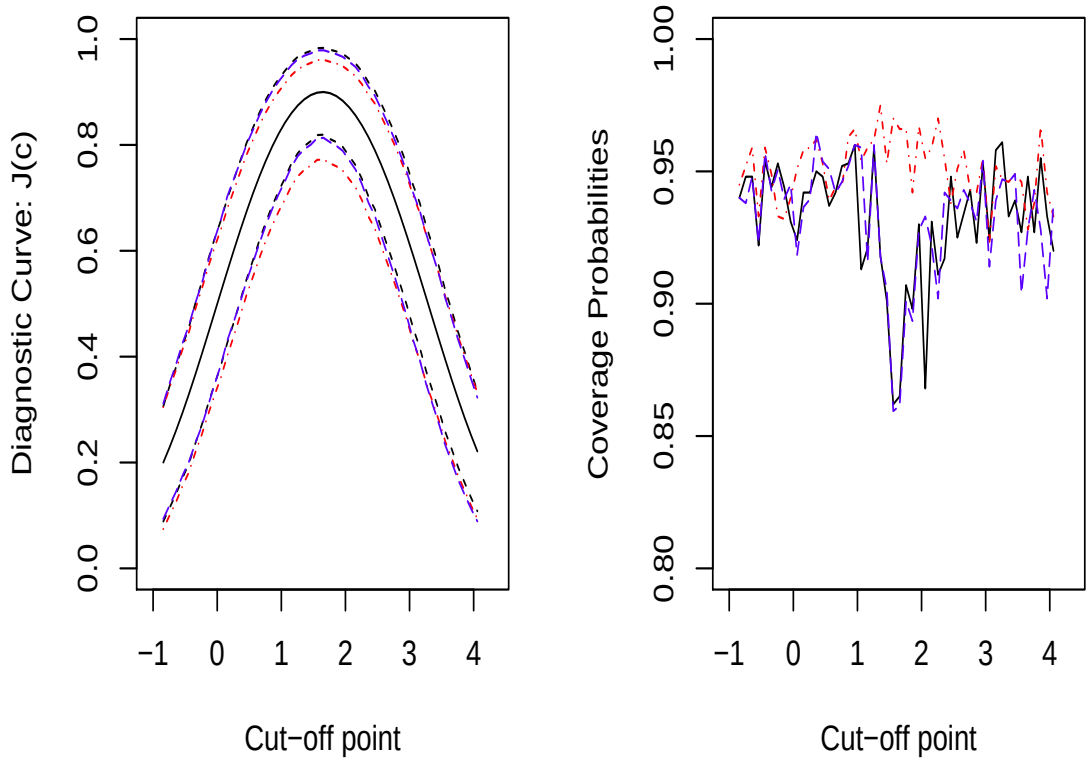

Figure A.3. Normal-Normal, Maximum of DC is 0.90. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability.

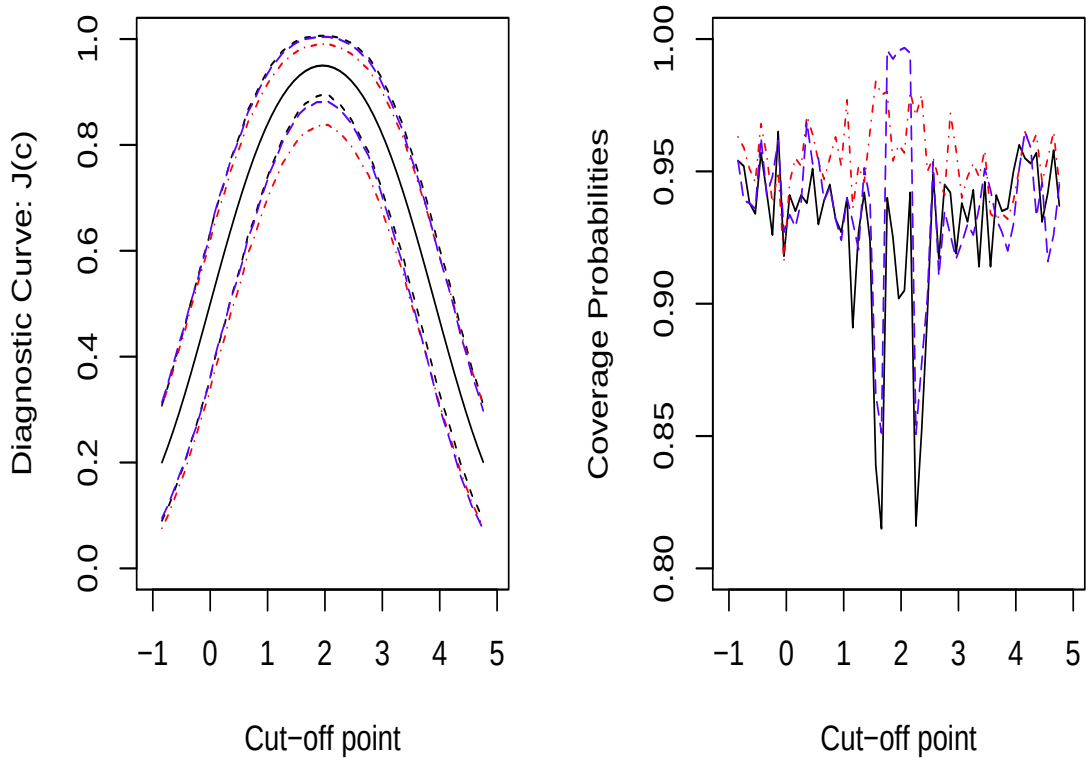

Figure A.4. Normal-Normal, Maximum of DC is 0.95. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability.

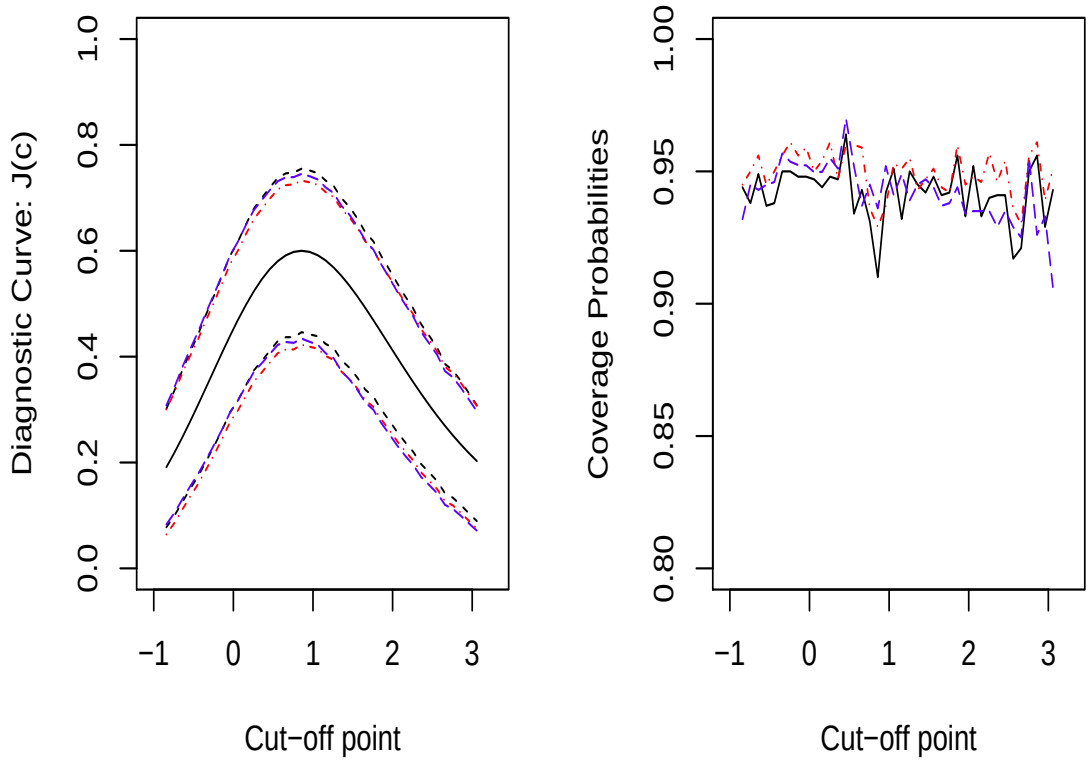

Figure A.5. Normal-T, Maximum of DC is 0.60. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability.

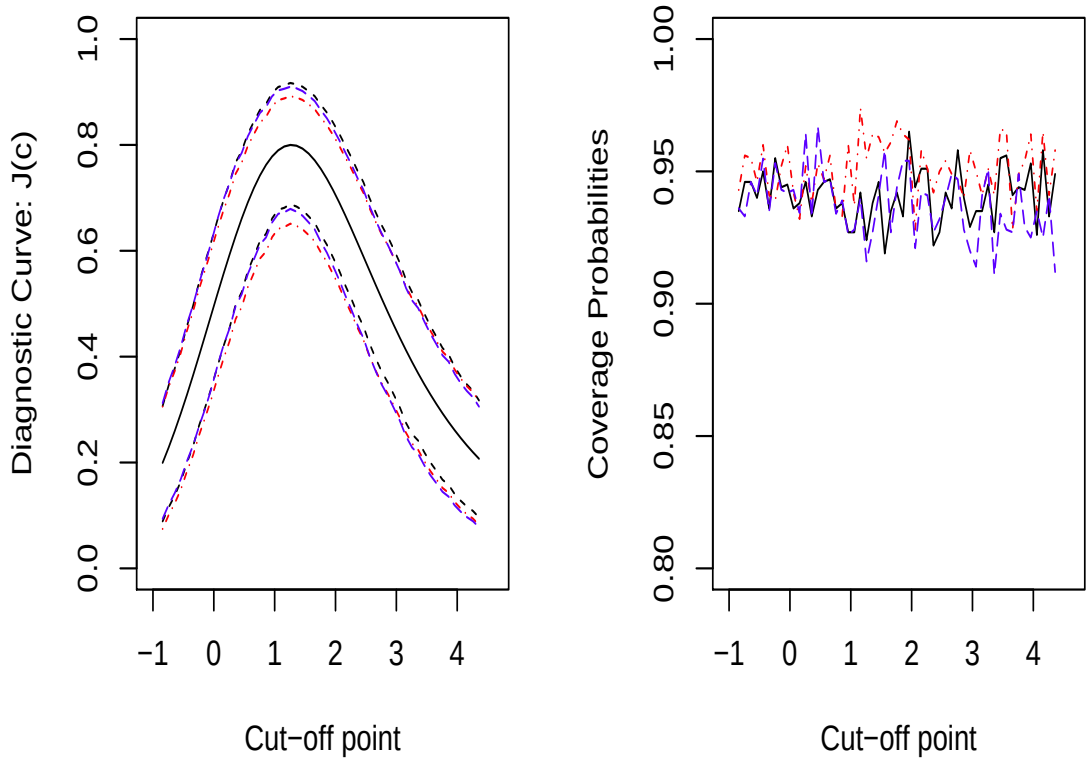

Figure A.6. Normal-T, Maximum of DC is 0.80. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability.

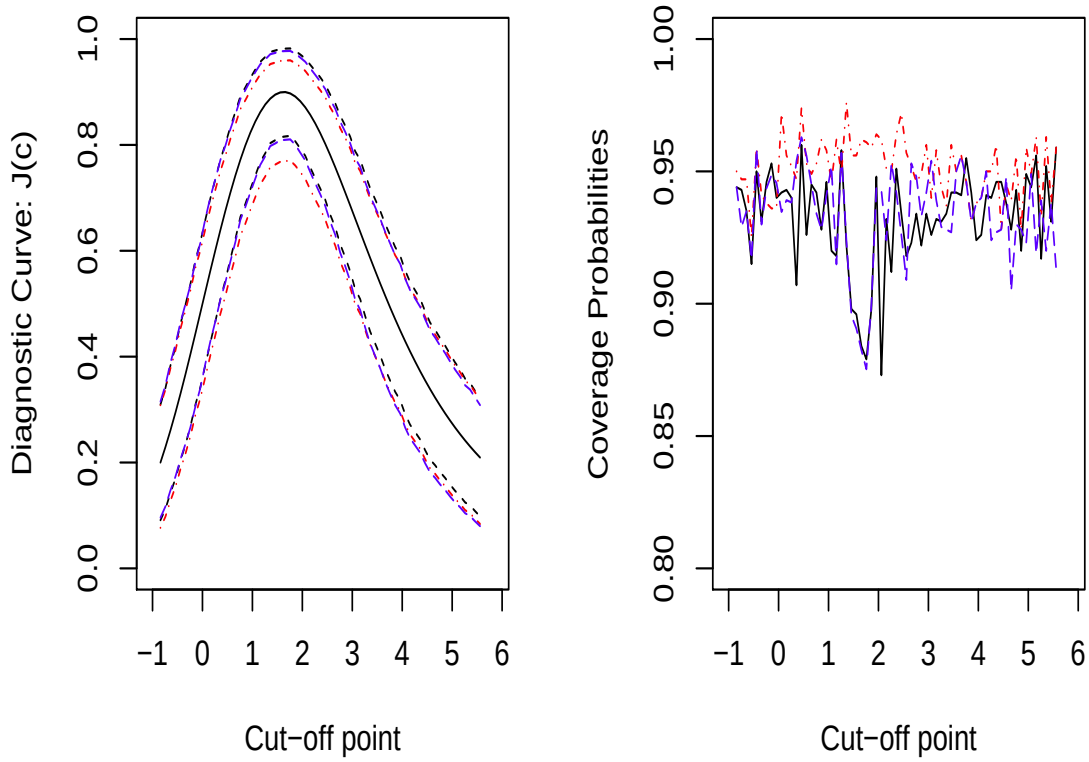

Figure A.7. Normal-T, Maximum of DC is 0.90. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability.

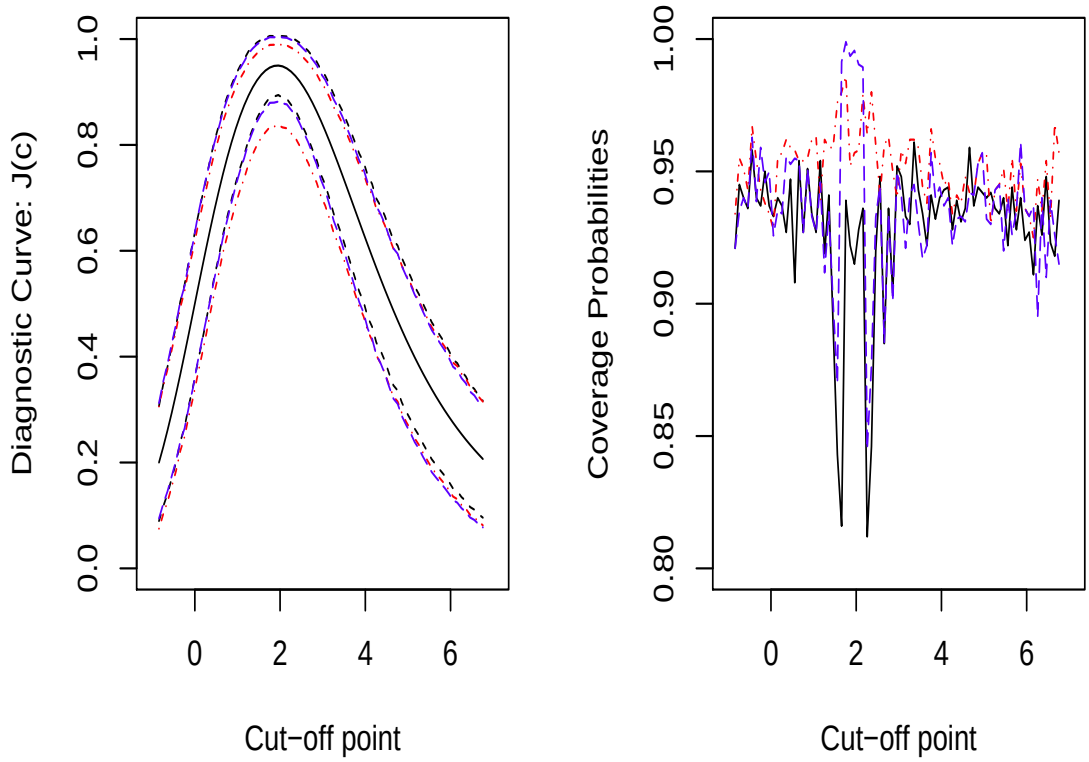

Figure A.8. Normal-T, Maximum of DC is 0.95. Left Panel: Point-wise C.I. for DC. Right Panel: Coverage Probability.

Appendix B

SIMULATION RESULTS FOR CHAPTER THREE

Notes:

- ECI stands for exact confidence interval.
- HBML stands for hybrid bootstrap and maximum likelihood based confidence interval.
- BP stands for bootstrap percentile confidence interval.

Table B.1. The coverage probabilities and average interval lengths of 95% Confidence Intervals for the difference in paired YIs: Scenario 1.

				ECI	HBML	BP
n_+	n_-	σ_{x12}	σ_{y12}	Cov. (Length)	Cov. (Length)	Cov. (Length)
10	10	0.1	0.1	0.9582 (0.5771)	0.8770 (0.3954)	0.7148 (0.2976)
			0.5	0.9674 (0.5363)	0.8730 (0.3662)	0.7122 (0.2844)
			0.9	0.9666 (0.4423)	0.8648 (0.3142)	0.7006 (0.2557)
		0.5	0.5	0.9736 (0.4790)	0.8646 (0.3352)	0.6946 (0.2666)
			0.9	0.9704 (0.3798)	0.8408 (0.2735)	0.6924 (0.2388)
		0.9	0.9	0.9758 (0.2521)	0.8280 (0.1973)	0.6762 (0.1952)
	20	0.1	0.1	0.9724 (0.4746)	0.8884 (0.3558)	0.8380 (0.2933)
			0.5	0.9618 (0.4231)	0.8908 (0.3198)	0.8408 (0.2753)
			0.9	0.9632 (0.3096)	0.8796 (0.2572)	0.8318 (0.2410)
		0.5	0.5	0.9720 (0.4037)	0.8878 (0.2981)	0.8360 (0.2618)
			0.9	0.9598 (0.2863)	0.8732 (0.2291)	0.8278 (0.2284)
		0.9	0.9	0.9610 (0.2248)	0.8534 (0.1750)	0.8188 (0.1929)
	50	0.1	0.1	0.9624 (0.4384)	0.8820 (0.3192)	0.8064 (0.2566)
			0.5	0.9524 (0.3747)	0.8872 (0.2794)	0.8082 (0.2414)
			0.9	0.9652 (0.2436)	0.8826 (0.2016)	0.8194 (0.2030)
		0.5	0.5	0.9622 (0.3679)	0.8794 (0.2685)	0.8128 (0.2348)
			0.9	0.9636 (0.2265)	0.8722 (0.1863)	0.8180 (0.1951)
		0.9	0.9	0.9616 (0.2026)	0.8566 (0.1585)	0.8032 (0.1743)
	100	0.1	0.1	0.9428 (0.4241)	0.8808 (0.3049)	0.7522 (0.2329)
			0.5	0.9516 (0.3613)	0.8740 (0.2613)	0.7408 (0.2138)
			0.9	0.9640 (0.2249)	0.8778 (0.1766)	0.7354 (0.1730)
		0.5	0.5	0.9546 (0.3495)	0.8698 (0.2535)	0.7204 (0.2076)
			0.9	0.9574 (0.2108)	0.8704 (0.1668)	0.7260 (0.1668)
		0.9	0.9	0.9686 (0.1946)	0.8518 (0.1505)	0.7102 (0.1558)
20	20	0.1	0.1	0.9478 (0.3693)	0.9176 (0.3072)	0.9122 (0.2895)
			0.5	0.9524 (0.3396)	0.9108 (0.2849)	0.9088 (0.2777)
			0.9	0.9490 (0.2847)	0.9040 (0.2420)	0.9074 (0.2469)
		0.5	0.5	0.9562 (0.3065)	0.9074 (0.2588)	0.9072 (0.2617)
			0.9	0.9624 (0.2442)	0.9032 (0.2112)	0.9080 (0.2295)
		0.9	0.9	0.9682 (0.1624)	0.8914 (0.1516)	0.9122 (0.1943)
	50	0.1	0.1	0.9496 (0.3031)	0.9174 (0.2630)	0.9060 (0.2562)
			0.5	0.9536 (0.2656)	0.9196 (0.2347)	0.9028 (0.2409)
			0.9	0.9528 (0.2013)	0.9126 (0.1802)	0.9018 (0.2054)
		0.5	0.5	0.9462 (0.2490)	0.9148 (0.2202)	0.8994 (0.2325)
			0.9	0.9544 (0.1813)	0.9114 (0.1619)	0.9008 (0.1957)
		0.9	0.9	0.9536 (0.1343)	0.8968 (0.1283)	0.8932 (0.1726)
	100	0.1	0.1	0.9588 (0.2802)	0.9160 (0.2438)	0.8772 (0.2332)
			0.5	0.9560 (0.2413)	0.9192 (0.2135)	0.8782 (0.2172)
			0.9	0.9576 (0.1662)	0.9144 (0.1502)	0.8744 (0.1759)
		0.5	0.5	0.9510 (0.2387)	0.9132 (0.2044)	0.8746 (0.2103)
			0.9	0.9512 (0.1550)	0.9124 (0.1383)	0.8696 (0.1688)
		0.9	0.9	0.9520 (0.1338)	0.8964 (0.1184)	0.8630 (0.1542)

Table B.2. Table (B.1) Continue.

				ECI	HBML	BP
n_+	n_-	σ_{x12}	σ_{y12}	Cov. (Length)	Cov. (Length)	Cov. (Length)
50	50	0.1	0.1	0.9500 (0.2256)	0.9356 (0.2075)	0.9394 (0.2210)
			0.5	0.9542 (0.2084)	0.9360 (0.1912)	0.9390 (0.2100)
			0.9	0.9402 (0.1777)	0.9372 (0.1631)	0.9380 (0.1882)
		0.5	0.5	0.9604 (0.1892)	0.9322 (0.1728)	0.9338 (0.1752)
			0.9	0.9478 (0.1529)	0.9228 (0.1413)	0.9390 (0.1752)
		0.9	0.9	0.9594 (0.1063)	0.9222 (0.1000)	0.9426 (0.1479)
	100	0.1	0.1	0.9446 (0.1957)	0.9348 (0.1815)	0.9290 (0.1961)
			0.5	0.9538 (0.1732)	0.9388 (0.1620)	0.9384 (0.1840)
			0.9	0.9520 (0.1335)	0.9362 (0.1272)	0.9338 (0.1572)
		0.5	0.5	0.9522 (0.1626)	0.9354 (0.1513)	0.9364 (0.1758)
			0.9	0.9552 (0.1195)	0.9362 (0.1135)	0.9364 (0.1491)
		0.9	0.9	0.9432 (0.0913)	0.9346 (0.0877)	0.9378 (0.1315)
100	100	0.1	0.1	0.9500 (0.1573)	0.9428 (0.1503)	0.9434 (0.1688)
			0.5	0.9514 (0.1452)	0.9450 (0.1378)	0.9398 (0.1600)
			0.9	0.9500 (0.1229)	0.9474 (0.1178)	0.9396 (0.1430)
		0.5	0.5	0.9510 (0.1310)	0.9362 (0.1253)	0.9376 (0.1518)
			0.9	0.9492 (0.1066)	0.9404 (0.1020)	0.9422 (0.1333)
		0.9	0.9	0.9582 (0.0742)	0.9384 (0.0721)	0.9386 (0.1120)

Table B.3. The coverage probabilities and average interval lengths of 95% Confidence Intervals for the difference in paired YIs: Scenario 2.

				ECI	HBML	BP
n_+	n_-	σ_{x12}	σ_{y12}	Cov. (Length)	Cov. (Length)	Cov. (Length)
10	10	0.1	0.1	0.9716 (0.6756)	0.8706 (0.5357)	0.8778 (0.4742)
			0.5	0.9620 (0.6275)	0.8740 (0.5030)	0.8704 (0.4585)
			0.9	0.9602 (0.5484)	0.8740 (0.4570)	0.8748 (0.4341)
		0.5	0.5	0.9642 (0.5703)	0.8642 (0.4689)	0.8680 (0.4434)
			0.9	0.9730 (0.4907)	0.8682 (0.4185)	0.8700 (0.4211)
		0.9	0.9	0.9658 (0.3831)	0.8772 (0.3596)	0.8822 (0.4043)
	20	0.1	0.1	0.9726 (0.5772)	0.8976 (0.4785)	0.8682 (0.4405)
			0.5	0.9614 (0.5142)	0.8944 (0.4393)	0.8588 (0.4203)
			0.9	0.9606 (0.4193)	0.8936 (0.3814)	0.8566 (0.3963)
		0.5	0.5	0.9658 (0.4895)	0.8880 (0.4174)	0.8574 (0.4094)
			0.9	0.9706 (0.3797)	0.8876 (0.3538)	0.8512 (0.3838)
		0.9	0.9	0.9728 (0.3318)	0.8866 (0.3149)	0.8546 (0.3700)
	50	0.1	0.1	0.9634 (0.4383)	0.8876 (0.4291)	0.7946 (0.3856)
			0.5	0.9518 (0.3747)	0.8878 (0.3868)	0.8000 (0.3672)
			0.9	0.9662 (0.2436)	0.9000 (0.3137)	0.7974 (0.3372)
		0.5	0.5	0.9630 (0.3779)	0.8772 (0.3723)	0.7860 (0.3572)
			0.9	0.9654 (0.2365)	0.8966 (0.2993)	0.7950 (0.3304)
		0.9	0.9	0.9622 (0.2050)	0.8952 (0.2803)	0.7938 (0.3235)
	100	0.1	0.1	0.9436 (0.4241)	0.8882 (0.4087)	0.7714 (0.3570)
			0.5	0.9510 (0.3683)	0.8780 (0.3629)	0.7672 (0.3375)
			0.9	0.9648 (0.2239)	0.8902 (0.2854)	0.7770 (0.3072)
		0.5	0.5	0.9542 (0.3565)	0.8794 (0.3558)	0.7604 (0.3317)
			0.9	0.9592 (0.2128)	0.8890 (0.2789)	0.7716 (0.3041)
		0.9	0.9	0.9704 (0.1966)	0.8808 (0.2657)	0.7574 (0.2976)
20	20	0.1	0.1	0.9470 (0.3733)	0.9192 (0.4062)	0.8954 (0.3978)
			0.5	0.9520 (0.3486)	0.9138 (0.3799)	0.9040 (0.3838)
			0.9	0.9486 (0.2947)	0.9100 (0.3427)	0.8916 (0.3633)
		0.5	0.5	0.9556 (0.3165)	0.9102 (0.3535)	0.8904 (0.3698)
			0.9	0.9616 (0.2542)	0.9090 (0.3101)	0.8954 (0.3495)
		0.9	0.9	0.9704 (0.1724)	0.9236 (0.2627)	0.9018 (0.3310)
	50	0.1	0.1	0.9494 (0.3131)	0.9168 (0.3445)	0.8700 (0.3434)
			0.5	0.9546 (0.2756)	0.9178 (0.3135)	0.8702 (0.3274)
			0.9	0.9522 (0.2043)	0.9246 (0.2638)	0.8770 (0.3037)
		0.5	0.5	0.9456 (0.2590)	0.9152 (0.2990)	0.8688 (0.3181)
			0.9	0.9560 (0.1825)	0.9258 (0.2463)	0.8694 (0.2947)
		0.9	0.9	0.9550 (0.1434)	0.9178 (0.2192)	0.8622 (0.2820)
	100	0.1	0.1	0.9576 (0.2902)	0.9188 (0.3213)	0.8782 (0.3173)
			0.5	0.9552 (0.2513)	0.9198 (0.2862)	0.8678 (0.2992)
			0.9	0.9580 (0.1662)	0.9282 (0.2281)	0.8728 (0.2711)
		0.5	0.5	0.9520 (0.2407)	0.9184 (0.2772)	0.8804 (0.2941)
			0.9	0.9518 (0.1550)	0.9226 (0.2182)	0.8612 (0.2655)
		0.9	0.9	0.9526 (0.1338)	0.9284 (0.2033)	0.8634 (0.2587)

Table B.4. Table (B.3) Continue.

				ECI	HBML	BP
n_+	n_-	σ_{x12}	σ_{y12}	Cov. (Length)	Cov. (Length)	Cov. (Length)
50	50	0.1	0.1	0.9500 (0.2256)	0.9394 (0.2683)	0.9024 (0.2842)
			0.5	0.9564 (0.2084)	0.9344 (0.2510)	0.9010 (0.2734)
			0.9	0.9366 (0.1777)	0.9278 (0.2240)	0.8900 (0.2578)
		0.5	0.5	0.9618 (0.1892)	0.9318 (0.2322)	0.8958 (0.2624)
			0.9	0.9474 (0.1529)	0.9326 (0.2025)	0.8872 (0.2454)
		0.9	0.9	0.9600 (0.1063)	0.9346 (0.1680)	0.8818 (0.2286)
	100	0.1	0.1	0.9454 (0.1957)	0.9384 (0.2334)	0.9168 (0.2510)
			0.5	0.9530 (0.1732)	0.9418 (0.2130)	0.9128 (0.2383)
			0.9	0.9524 (0.1337)	0.9370 (0.1805)	0.9082 (0.2202)
		0.5	0.5	0.9526 (0.1626)	0.9394 (0.2020)	0.9080 (0.2309)
			0.9	0.9546 (0.1195)	0.9348 (0.1669)	0.9104 (0.2120)
		0.9	0.9	0.9434 (0.0913)	0.9406 (0.1457)	0.9036 (0.2008)
100	100	0.1	0.1	0.9582 (0.1974)	0.9480 (0.1926)	0.9252 (0.2132)
			0.5	0.9572 (0.1839)	0.9420 (0.1795)	0.9238 (0.2041)
			0.9	0.9580 (0.1638)	0.9442 (0.1601)	0.9168 (0.1921)
		0.5	0.5	0.9620 (0.1696)	0.9414 (0.1662)	0.9222 (0.1954)
			0.9	0.9592 (0.1464)	0.9436 (0.1445)	0.9196 (0.1829)
		0.9	0.9	0.9506 (0.1201)	0.9454 (0.1194)	0.9158 (0.1695)

Table B.5. The coverage probabilities and average interval lengths of 95% Confidence Intervals for the difference in paired YIs: Scenario 3.

				ECI	HBML	BP
n_+	n_-	σ_{x12}	σ_{y12}	Cov. (Length)	Cov. (Length)	Cov. (Length)
10	10	0.1	0.1	0.9558 (0.7807)	0.8982 (0.7140)	0.9216 (0.6697)
			0.5	0.9550 (0.7314)	0.9040 (0.6448)	0.9296 (0.6256)
			0.9	0.9662 (0.6485)	0.9058 (0.5431)	0.9268 (0.5503)
		0.5	0.5	0.9714 (0.6756)	0.9110 (0.4529)	0.9348 (0.5785)
			0.9	0.9634 (0.5386)	0.9114 (0.4529)	0.9304 (0.4931)
		0.9	0.9	0.9720 (0.3480)	0.9182 (0.2819)	0.9268 (0.3739)
	20	0.1	0.1	0.9610 (0.7256)	0.9124 (0.6343)	0.9370 (0.6058)
			0.5	0.9576 (0.6284)	0.9118 (0.5532)	0.9428 (0.5514)
			0.9	0.9476 (0.4684)	0.9206 (0.4306)	0.9544 (0.4644)
		0.5	0.5	0.9514 (0.5756)	0.9120 (0.5027)	0.9446 (0.5128)
			0.9	0.9642 (0.4694)	0.9212 (0.3607)	0.9546 (0.4159)
		0.9	0.9	0.9566 (0.2873)	0.9180 (0.2527)	0.9632 (0.3388)
	50	0.1	0.1	0.9524 (0.6538)	0.9032 (0.5687)	0.9130 (0.5332)
			0.5	0.9510 (0.5416)	0.9144 (0.4746)	0.9306 (0.4674)
			0.9	0.9538 (0.3548)	0.9356 (0.3190)	0.9522 (0.3574)
		0.5	0.5	0.9462 (0.5209)	0.9140 (0.4484)	0.9218 (0.4464)
			0.9	0.9628 (0.3123)	0.9342 (0.2816)	0.9512 (0.3318)
		0.9	0.9	0.9574 (0.2648)	0.9126 (0.2232)	0.9404 (0.2857)
	100	0.1	0.1	0.9562 (0.6324)	0.8994 (0.5428)	0.9058 (0.5000)
			0.5	0.9542 (0.5217)	0.9098 (0.4480)	0.9210 (0.4335)
			0.9	0.9526 (0.3026)	0.9310 (0.2708)	0.9370 (0.3054)
		0.5	0.5	0.9570 (0.4605)	0.8982 (0.4282)	0.9056 (0.4163)
			0.9	0.9548 (0.2837)	0.9232 (0.2458)	0.9404 (0.2879)
		0.9	0.9	0.9662 (0.2463)	0.9114 (0.2129)	0.9252 (0.2591)
20	20	0.1	0.1	0.9466 (0.5716)	0.9262 (0.5359)	0.9480 (0.5354)
			0.5	0.9514 (0.5147)	0.9334 (0.4823)	0.9484 (0.4969)
			0.9	0.9534 (0.4309)	0.9216 (0.4060)	0.9510 (0.4375)
		0.5	0.5	0.9640 (0.4521)	0.9366 (0.4227)	0.9590 (0.4562)
			0.9	0.9524 (0.3235)	0.9324 (0.3316)	0.9682 (0.3877)
		0.9	0.9	0.9574 (0.2232)	0.9416 (0.2077)	0.9804 (0.3018)
	50	0.1	0.1	0.9468 (0.4747)	0.9332 (0.4553)	0.9448 (0.4610)
			0.5	0.9616 (0.4087)	0.9360 (0.3891)	0.9496 (0.4140)
			0.9	0.9658 (0.3024)	0.9358 (0.2882)	0.9612 (0.3348)
		0.5	0.5	0.9620 (0.3841)	0.9306 (0.3558)	0.9512 (0.3895)
			0.9	0.9550 (0.2561)	0.9374 (0.2430)	0.9622 (0.3041)
		0.9	0.9	0.9546 (0.1885)	0.9336 (0.1732)	0.9664 (0.2534)
	100	0.1	0.1	0.9610 (0.4479)	0.9264 (0.4215)	0.9412 (0.4264)
			0.5	0.9554 (0.3683)	0.9258 (0.3481)	0.9406 (0.3737)
			0.9	0.9526 (0.2417)	0.9518 (0.2305)	0.9662 (0.2790)
		0.5	0.5	0.9534 (0.3538)	0.9252 (0.3295)	0.9384 (0.3593)
			0.9	0.9564 (0.2137)	0.9416 (0.2018)	0.9626 (0.2604)
		0.9	0.9	0.9508 (0.1748)	0.9328 (0.1602)	0.9562 (0.2295)

Table B.6. Table (B.5) Continue.

				ECI	HBML	BP
n_+	n_-	σ_{x12}	σ_{y12}	Cov. (Length)	Cov. (Length)	Cov. (Length)
50	50	0.1	0.1	0.9540 (0.3623)	0.9384 (0.3522)	9.9524 (0.3736)
			0.5	0.9462 (0.3263)	0.9480 (0.3155)	0.9512 (0.3461)
			0.9	0.9616 (0.2750)	0.9402 (0.2662)	0.9534 (0.3035)
		0.5	0.5	0.9542 (0.2831)	0.9374 (0.2741)	0.9558 (0.3162)
			0.9	0.9512 (0.2231)	0.9450 (0.2153)	0.9618 (0.2676)
		0.9	0.9	0.9476 (0.1368)	0.9470 (0.1329)	0.9772 (0.2075)
	100	0.1	0.1	0.9468 (0.3139)	0.9438 (0.3054)	0.9550 (0.3294)
			0.5	0.9574 (0.2703)	0.9420 (0.2629)	0.9528 (0.2967)
			0.9	0.9634 (0.2049)	0.9506 (0.2009)	0.9636 (0.2430)
		0.5	0.5	0.9516 (0.2465)	0.9486 (0.2385)	0.9548 (0.2781)
			0.9	0.9576 (0.1714)	0.9406 (0.1671)	0.9630 (0.2198)
		0.9	0.9	0.9486 (0.1178)	0.9516 (0.1148)	0.9756 (0.1813)
100	100	0.1	0.1	0.9478 (0.2545)	0.9458 (0.2516)	0.9540 (0.2771)
			0.5	0.9500 (0.2288)	0.9436 (0.2258)	0.9566 (0.2566)
			0.9	0.9512 (0.1936)	0.9442 (0.1898)	0.9578 (0.2240)
		0.5	0.5	0.9438 (0.1993)	0.9476 (0.1967)	0.9534 (0.2352)
			0.9	0.9602 (0.1566)	0.9426 (0.1538)	0.9646 (0.1982)
		0.9	0.9	0.9476 (0.0956)	0.9464 (0.0939)	0.9754 (0.1523)

Table B.7. The coverage probabilities and average interval lengths of 95% ECI for the difference in paired YIs under mixture normal setting, for all scenarios.

				Scen1	Scen2	Scen3
n_+	n_-	σ_{x12}	σ_{y12}	Cov. (Length)	Cov. (Length)	Cov. (Length)
10	10	0.1	0.1	0.9770 (0.6234)	0.9682 (0.7031)	0.9824 (0.8182)
			0.5	0.9742 (0.5751)	0.9738 (0.6464)	0.9692 (0.7547)
			0.9	0.9654 (0.4939)	0.9746 (0.5805)	0.9622 (0.6533)
		0.5	0.5	0.9732 (0.5214)	0.9752 (0.5961)	0.9742 (0.6689)
			0.9	0.9644 (0.4249)	0.9438 (0.5065)	0.9636 (0.5444)
		0.9	0.9	0.9500 (0.9072)	0.9482 (0.4108)	0.9584 (0.3775)
20	20	0.1	0.1	0.9676 (0.4173)	0.9614 (0.4753)	0.9614 (0.5638)
			0.5	0.9722 (0.3822)	0.9636 (0.4396)	0.9716 (0.5065)
			0.9	0.9642 (0.3293)	0.9600 (0.3895)	0.9710 (0.4383)
		0.5	0.5	0.9580 (0.3422)	0.9548 (0.3976)	0.9546 (0.4455)
			0.9	0.9552 (0.2805)	0.9324 (0.3410)	0.9524 (0.3617)
		0.9	0.9	0.9156 (0.1985)	0.9168 (0.2702)	0.9416 (0.2464)
50	50	0.1	0.1	0.9632 (0.2585)	0.9484 (0.2947)	0.9566 (0.3506)
			0.5	0.9586 (0.2365)	0.9500 (0.2720)	0.9490 (0.3152)
			0.9	0.9416 (0.2011)	0.9188 (0.2400)	0.9380 (0.2702)
		0.5	0.5	0.9524 (0.2091)	0.9356 (0.2458)	0.9468 (0.2756)
			0.9	0.9310 (0.1708)	0.9064 (0.2107)	0.9618 (0.2224)
		0.9	0.9	0.8826 (0.1210)	0.8628 (0.1666)	0.9460 (0.1496)

Appendix C

SIMULATION RESULTS FOR CHAPTER FOUR

Notes:

- S_1 represents $(n_-, n_0, n_+) = (50, 50, 50)$.
- S_2 represents $(n_-, n_0, n_+) = (80, 80, 80)$.
- S_3 represents $(n_-, n_0, n_+) = (100, 100, 100)$.
- S_4 represents $(n_-, n_0, n_+) = (100, 50, 80)$.
- I represents $J_{-0} > J_{0+}$.
- II represents $J_{-0} < J_{0+}$.
- ECI stands for exact confidence interval.
- HBML stands for hybrid bootstrap and maximum likelihood based confidence interval.
- ACNA stands for the AC adjusted normal approximation confidence interval.
- BCa stands for the Bias Corrected and accelerated interval.
- BP stands for bootstrap percentile confidence interval.

Table C.1. The coverage probabilities and average interval lengths of 95% Confidence Intervals for J_3.

			ECI	HBML	ACNA	BCa
J_3	Size	Scen.	Cov. (Length)	Cov. (Length)	Cov. (Length)	Cov. (Length)
0.55	S_1	I	0.945 (0.1513)	0.936 (0.1496)	0.894 (0.1639)	0.855 (0.1672)
		II	0.953 (0.1604)	0.938 (0.1592)	0.868 (0.1680)	0.883 (0.1739)
	S_2	I	0.956 (0.1205)	0.945 (0.1194)	0.850 (0.1367)	0.864 (0.1361)
		II	0.942 (0.1276)	0.939 (0.1269)	0.840 (0.1409)	0.865 (0.1412)
	S_3	I	0.951 (0.1079)	0.943 (0.1068)	0.838 (0.1251)	0.874 (0.1244)
		II	0.950 (0.1143)	0.938 (0.1134)	0.846 (0.1288)	0.875 (0.1292)
	S_4	I	0.957 (0.1276)	0.933 (0.1275)	0.853 (0.1463)	0.856 (0.1739)
		II	0.950 (0.1345)	0.941 (0.1332)	0.884 (0.1477)	0.890 (0.1490)
0.65	S_1	I	0.949 (0.1539)	0.933 (0.1502)	0.935 (0.1543)	0.869 (0.1639)
		II	0.945 (0.1577)	0.923 (0.1541)	0.917 (0.1571)	0.874 (0.1677)
	S_2	I	0.950 (0.1221)	0.960 (0.1201)	0.928 (0.1305)	0.869 (1.1351)
		II	0.947 (0.1254)	0.942 (0.1233)	0.912 (0.1317)	0.879 (0.1374)
	S_3	I	0.956 (0.1090)	0.940 (0.1074)	0.903 (0.1189)	0.874 (0.1217)
		II	0.942 (0.1122)	0.932 (0.1108)	0.896 (0.1208)	0.884 (0.1249)
	S_4	I	0.947 (0.1318)	0.934 (0.1298)	0.916 (0.1385)	0.886 (0.1439)
		II	0.957 (0.1341)	0.943 (0.1308)	0.931 (0.1376)	0.860 (0.1425)
0.75	S_1	I	0.941 (0.1039)	0.937 (0.1018)	0.927 (0.1289)	0.905 (0.1390)
		II	0.947 (0.1084)	0.928 (0.1048)	0.956 (0.1299)	0.894 (0.1402)
	S_2	I	0.953 (0.0819)	0.949 (0.0806)	0.940 (0.1067)	0.907 (0.1114)
		II	0.946 (0.0857)	0.937 (0.0842)	0.960 (0.1077)	0.916 (0.1124)
	S_3	I	0.958 (0.0732)	0.945 (0.0717)	0.946 (0.0969)	0.932 (0.1009)
		II	0.951 (0.0765)	0.945 (0.0748)	0.947 (0.0979)	0.901 (0.1020)
	S_4	I	0.940 (0.0983)	0.933 (0.0958)	0.951 (0.1253)	0.912 (0.1358)
		II	0.940 (0.1013)	0.924 (0.0986)	0.947 (0.1248)	0.907 (0.1356)
0.85	S_1	I	0.950 (0.1023)	0.921 (0.0960)	0.864 (0.1072)	0.887 (0.1190)
		II	0.942 (0.1042)	0.936 (0.0972)	0.900 (0.1060)	0.883 (0.1180)
	S_2	I	0.956 (0.0804)	0.940 (0.0774)	0.931 (0.0893)	0.920 (0.0972)
		II	0.950 (0.0820)	0.930 (0.0789)	0.938 (0.0887)	0.909 (0.0966)
	S_3	I	0.961 (0.0717)	0.945 (0.0694)	0.947 (0.0821)	0.903 (0.0869)
		II	0.944 (0.0730)	0.942 (0.0700)	0.949 (0.0814)	0.905 (0.0873)
	S_4	I	0.950 (0.0986)	0.927 (0.0931)	0.927 (0.1054)	0.913 (0.1174)
		II	0.955 (0.0989)	0.919 (0.0938)	0.935 (0.1031)	0.893 (0.1149)

Table C.2. The coverage probabilities and average interval lengths of 95% Confidence Intervals for c_{o1}.

			ECI	HBML	BP
J_3	Sample Size	Scenario	Cov. (Length)	Cov. (Length)	Cov. (Length)
0.55	S_1	I	0.961 (0.4666)	0.949 (0.4620)	0.970 (1.0251)
		II	0.956 (0.4826)	0.940 (0.4563)	0.973 (1.0467)
	S_2	I	0.949 (0.3614)	0.948 (0.3546)	0.984 (0.9291)
		II	0.962 (0.3745)	0.957 (0.3607)	0.980 (0.9374)
	S_3	I	0.939 (0.3214)	0.943 (0.3161)	0.976 (0.8620)
		II	0.947 (0.3326)	0.940 (0.3227)	0.986 (0.8861)
	S_4	I	0.942 (0.3864)	0.929 (0.3787)	0.978 (0.9597)
		II	0.949 (0.4117)	0.945 (0.3936)	0.977 (0.9783)
0.65	S_1	I	0.946 (0.4136)	0.944 (0.4085)	0.970 (0.8858)
		II	0.953 (0.3933)	0.949 (0.3811)	0.979 (0.8835)
	S_2	I	0.960 (0.3237)	0.957 (0.3214)	0.968 (0.7982)
		II	0.950 (0.3060)	0.945 (0.2992)	0.981 (0.7829)
	S_3	I	0.952 (0.2866)	0.956 (0.2849)	0.978 (0.7425)
		II	0.959 (0.2739)	0.961 (0.2692)	0.977 (0.7367)
	S_4	I	0.954 (0.3518)	0.947 (0.3498)	0.970 (0.8368)
		II	0.945 (0.3435)	0.943 (0.3367)	0.969 (0.8360)
0.75	S_1	I	0.935 (0.7558)	0.920 (0.7317)	0.981 (1.9014)
		II	0.950 (0.6778)	0.930 (0.6526)	0.978 (1.5820)
	S_2	I	0.943 (0.5924)	0.927 (0.5803)	0.985 (1.6091)
		II	0.956 (0.5310)	0.943 (0.5192)	0.987 (1.3701)
	S_3	I	0.940 (0.5299)	0.939 (0.5200)	0.983 (1.4854)
		II	0.957 (0.4734)	0.950 (0.4607)	0.987 (1.2636)
	S_4	I	0.948 (0.5612)	0.942 (0.5516)	0.978 (1.7407)
		II	0.946 (0.5049)	0.943 (0.4944)	0.979 (1.4269)
0.85	S_1	I	0.960 (0.8305)	0.945 (0.8117)	0.958 (1.8705)
		II	0.944 (0.6783)	0.933 (0.6567)	0.975 (1.4952)
	S_2	I	0.962 (0.6508)	0.957 (0.6436)	0.972 (1.5958)
		II	0.943 (0.5333)	0.932 (0.5236)	0.977 (1.2938)
	S_3	I	0.940 (0.5817)	0.941 (0.5769)	0.975 (1.5057)
		II	0.959 (0.4758)	0.952 (0.4675)	0.974 (1.1829)
	S_4	I	0.952 (0.6319)	0.953 (0.6315)	0.955 (1.7405)
		II	0.949 (0.5130)	0.941 (0.5101)	0.974 (1.3903)

Table C.3. The coverage probabilities and average interval lengths of 95% Confidence Intervals for c_{o2}.

			ECI	HBML	BP
J_3	Sample Size	Scenario	Cov. (Length)	Cov. (Length)	Cov. (Length)
0.55	S_1	I	0.948 (0.5849)	0.968 (0.5529)	0.967 (1.2373)
		II	0.949 (0.4648)	0.977 (0.4493)	0.969 (1.0383)
	S_2	I	0.945 (0.4502)	0.967 (0.4342)	0.978 (1.1351)
		II	0.948 (0.3616)	0.979 (0.3543)	0.980 (0.9374)
	S_3	I	0.953 (0.3994)	0.974 (0.3881)	0.979 (1.0618)
		II	0.962 (0.3225)	0.979 (0.3176)	0.979 (0.8565)
	S_4	I	0.967 (0.3864)	0.975 (0.3787)	0.982 (1.2429)
		II	0.945 (0.4211)	0.974 (0.4105)	0.980 (1.0007)
0.65	S_1	I	0.965 (0.4718)	0.979 (0.4561)	0.980 (1.0460)
		II	0.953 (0.4124)	0.973 (0.4073)	0.971 (0.8883)
	S_2	I	0.943 (0.3677)	0.976 (0.3608)	0.980 (0.9250)
		II	0.947 (0.3239)	0.975 (0.3208)	0.980 (0.8009)
	S_3	I	0.952 (0.3255)	0.977 (0.3204)	0.980 (0.8692)
		II	0.950 (0.2886)	0.979 (0.2864)	0.981 (0.7457)
	S_4	I	0.954 (0.4278)	0.979 (0.4166)	0.972 (1.0319)
		II	0.959 (0.3817)	0.985 (0.3775)	0.972 (0.8720)
0.75	S_1	I	0.950 (0.8215)	0.958 (0.7893)	0.983 (2.0064)
		II	0.944 (0.7866)	0.949 (0.7675)	0.966 (1.7026)
	S_2	I	0.946 (0.6466)	0.953 (0.6302)	0.986 (1.6871)
		II	0.964 (0.6224)	0.975 (0.6087)	0.982 (1.4536)
	S_3	I	0.950 (0.5774)	0.960 (0.5623)	0.979 (1.5660)
		II	0.951 (0.5557)	0.966 (0.5468)	0.986 (1.3656)
	S_4	I	0.956 (0.6712)	0.962 (0.6526)	0.978 (1.8531)
		II	0.954 (0.6543)	0.964 (0.6481)	0.972 (1.6259)
0.85	S_1	I	0.952 (0.8399)	0.952 (0.8142)	0.961 (1.9056)
		II	0.961 (0.8634)	0.971 (0.8539)	0.929 (1.5496)
	S_2	I	0.959 (0.6619)	0.969 (0.6512)	0.984 (1.6372)
		II	0.948 (0.6776)	0.963 (0.6712)	0.950 (1.3958)
	S_3	I	0.952 (0.5880)	0.958 (0.5773)	0.984 (1.5124)
		II	0.957 (0.6039)	0.973 (0.5970)	0.960 (1.2993)
	S_4	I	0.959 (0.6903)	0.976 (0.6790)	0.973 (1.7405)
		II	0.952 (0.7274)	0.973 (0.7273)	0.940 (1.5607)

Appendix D

PROOF FOR CHAPTER FOUR

D.1 Proof of Theorem 4.1

Notice that,

$$\widehat{J}_{3E} = \frac{1}{2}\left[\max_{c_1}\{F_{n_-}(c_1) - F_{n_0}(c_1)\} + \max_{c_2}\{F_{n_0}(c_2) - F_{n_+}(c_2)\}\right]. \tag{D.1}$$

The empirical maximum can be separately achieved via maximizing on c_1 targeting on $\widehat{F}_-(c_1) - \widehat{F}_0(c_1)$ and maximizing on c_2 targeting on $\widehat{F}_0(c_2) - \widehat{F}_+(c_2)$. Then the almost surely convergence of $\widehat{J}_{3E}$, $\widehat{c}_{o1E}$, and $\widehat{c}_{o2E}$ can be obtained by using lemma 2.1 (page 19) in Hsieh and Turnbull [21].

Now define functional G_3 as

$$\begin{aligned} &G_3(F_1, F_2, F_3, c_1, c_2, c_{o1}, c_{o2}) = \\ &[(F_1(c_1) - F_2(c_1)) - (F_1(c_{o1}) - F_2(c_{o1}))] + [(F_2(c_2) - F_3(c_2)) - (F_2(c_{o2}) - F_3(c_{o2}))], \end{aligned} \tag{D.2}$$

and define functional G_2 as

$$G_2(F_1, F_2, c, c_o) = (F_1(c) - F_2(c)) - (F_1(c_o) - F_2(c_o)). \tag{D.3}$$

From Theorem 4.41 in Csörgö and Révész [9], it follows that

$$\begin{aligned} &G_3(\widehat{F}_-, \widehat{F}_0, \widehat{F}_+, c_1, c_2, c_{o1}, c_{o2}) - G_3(F_-, F_0, F_+, c_1, c_2, c_{o1}, c_{o2}) = \\ &G_2(\widehat{F}_-, \widehat{F}_0, c_1, c_{o1}) - G_2(F_-, F_0, c_1, c_{o1}) + \\ &G_2(\widehat{F}_0, \widehat{F}_+, c_2, c_{o2}) - G_2(F_0, F_+, c_2, c_{o2}) = \\ &\frac{1}{\sqrt{n_-}}[B_1(F_-(c_1)) - B_1(F_-(c_{o1}))] - \frac{1}{\sqrt{n_0}}[B_2(F_0(c_1)) - B_2(F_0(c_{o1}))] + \\ &\frac{1}{\sqrt{n_0}}[B_2(F_0(c_2)) - B_2(F_0(c_{o2}))] - \frac{1}{\sqrt{n_+}}[B_3(F_+(c_2)) - B_3(F_+(c_{o2}))] + \\ &O(n_0^{-1}log(n_0)), \qquad a.s. \end{aligned} \tag{D.4}$$

where $B_1(\cdot)$, $B_2(\cdot)$, and $B_3(\cdot)$ are independent Brownian bridge processes on $[0,1]$.

According to Chernoff [7] and Hsieh and Turnbull [22], under (A4.3) and if c_1 is close to c_{o1}, the distribution of $G_2(\widehat{F}_-, \widehat{F}_0, c_1, c_{o1}) - G_2(F_-, F_0, c_1, c_{o1})$ is approximately

$$n_0^{-1/2}\left[\lambda_1^{-2} f_-(c_{o1}) + f_0(c_{o1})\right]^{1/2} Br_1(c_1 - c_{o1}), \tag{D.5}$$

where Br_1 is a two-sided standard Brownian motion on $(-\infty, \infty)$.

Similarly, if c_2 is close to c_{o2}, the distribution of $G_2(\widehat{F}_0, \widehat{F}_+, c_2, c_{o2}) - G_2(F_0, F_+, c_2, c_{o2})$ is approximately

$$n_0^{-1/2}\left[f_0(c_{o2}) + \lambda_2^2 f_+(c_{o2})\right]^{1/2} Br_2(c_2 - c_{o2}), \tag{D.6}$$

where Br_2 is also a two-sided standard Brownian motion on $(-\infty, \infty)$.

By Taylor expansion, under (A4.3), we have that

$$G_2(F_-, F_0, c_1, c_{o1}) \approx \frac{1}{2}(f'_-(c_{o1}) - f'_0(c_{o1}))(c_1 - c_{o1})^2, \tag{D.7}$$

and

$$G_2(F_0, F_+, c_2, c_{o2}) \approx \frac{1}{2}(f'_0(c_{o2}) - f'_+(c_{o2}))(c_2 - c_{o2})^2. \tag{D.8}$$

Notice that

$$\begin{aligned} &\max_{c_1,c_2}\{G_3(\widehat{F}_-, \widehat{F}_0, \widehat{F}_+, c_1, c_2, c_{o1}, c_{o2})\} = \\ &\max_{c_1}\{G_2(\widehat{F}_-, \widehat{F}_0, c_1, c_{o1})\} + \max_{c_2}\{G_2(\widehat{F}_0, \widehat{F}_+, c_2, c_{o2})\} \end{aligned} \tag{D.9}$$

Further, we have

$$\begin{aligned} &\max_{c_1}\{G_2(\widehat{F}_-, \widehat{F}_0, c_1, c_{o1})\} = \\ &\max_{c_1}\{G_2(\widehat{F}_-, \widehat{F}_0, c_1, c_{o1}) - G_2(F_-, F_0, c_1, c_{o1}) + G_2(F_-, F_0, c_1, c_{o1})\}, \end{aligned} \tag{D.10}$$

and

$$\max_{c_2}\{G_2(\widehat{F}_0,\widehat{F}_+,c_2,c_{o2})\} = \tag{D.11}$$
$$\max_{c_2}\{G_2(\widehat{F}_0,\widehat{F}_+,c_2,c_{o2}) - G_2(F_0,F_+,c_2,c_{o2}) + G_2(F_0,F_+,c_2,c_{o2})\}.$$

From (D.5) - (D.8), and (A4.4), it follows that (D.10) converges to

$$\max_{c_1}\{n_0^{-1/2}\left[\lambda_1^{-2}f_-(c_{o1}) + f_0(c_{o1})\right]^{1/2} Br_1(c_1 - c_{o1}) - \frac{d_1}{2}(c_1 - c_{o1})^2\} = \tag{D.12}$$
$$R_1 n_0^{-\frac{2}{3}} \max_{t_1}(Br(t_1) - t_1^2), \tag{D.13}$$

and (D.11) converges to

$$\max_{c_2}\{n_0^{-1/2}\left[f_0(c_{o2}) + \lambda_2^2 f_+(c_{o2})\right]^{1/2} Br_2(c_2 - c_{o2}) - \frac{d_2}{2}(c_2 - c_{o2})^2\} = \tag{D.14}$$
$$R_2 n_0^{-\frac{2}{3}} \max_{t_2}(Br(t_2) - t_2^2) \tag{D.15}$$

in distribution, respectively. In which, $t_1 = (c_1 - c_{o1})/\chi_1$, $\chi_1 = (\frac{4\Lambda_1}{n_0 d_1^2})^{1/3}$, $\Lambda_1 = (\lambda_1^{-2}f_-(c_{o1}) + f_0(c_{o1}))$, and $R_1 = \frac{d_1}{2}(\frac{4\Lambda_1}{d_1^2})^{\frac{2}{3}}$; $t_2 = (c_2 - c_{o2})/\chi_2$, $\chi_2 = (\frac{4\Lambda_2}{n_0 d_2^2})^{1/3}$, $\Lambda_2 = (f_0(c_{o2}) + \lambda_2^2 f_+(c_{o2}))$, and $R_2 = \frac{d_2}{2}(\frac{4\Lambda_2}{d_2^2})^{\frac{2}{3}}$.

Therefore, we have that

$$\sqrt{4n_0}(\widehat{J}_{3E} - J_3) = \tag{D.16}$$
$$\sqrt{n_0}\left[\widehat{F}_-(c_{o1}) - \widehat{F}_0(c_{o1}) - (F_-(c_{o1}) - F_0(c_{o1}))\right] +$$
$$\sqrt{n_0}\left[\widehat{F}_0(c_{o2}) - \widehat{F}_+(c_{o2}) - (F_0(c_{o2}) - F_+(c_{o2}))\right] +$$
$$\max_{c_1,c_2}\{\sqrt{n_0}G_3(\widehat{F}_-,\widehat{F}_0,\widehat{F}_+,c_1,c_2,c_{o1},c_{o2})\} =$$
$$\sqrt{n_0}\left[\widehat{F}_-(c_{o1}) - \widehat{F}_0(c_{o1}) - (F_-(c_{o1}) - F_0(c_{o1}))\right] + \tag{D.17}$$
$$\sqrt{n_0}\left[\widehat{F}_0(c_{o2}) - \widehat{F}_+(c_{o2}) - (F_0(c_{o2}) - F_+(c_{o2}))\right] +$$
$$\max_{c_1}\{\sqrt{n_0}G_2(\widehat{F}_-,\widehat{F}_0,c_1,c_{o1})\} + \max_{c_2}\{\sqrt{n_0}G_2(\widehat{F}_0,\widehat{F}_+,c_2,c_{o2})\}$$

converges in distribution to the one stated in Theorem 4.1.

Appendix E

PROOF FOR CHAPTER FIVE

Denote the neighborhood of z by $\mathcal{N}(z)$ for given $Z = z$. Here, we first list the assumptions which are necessary for Theorems 4.1-4.4.

The assumptions E1-E5 are cited from Yao *et al.* [64] as belows:

" (E1) $\theta(\cdot)$ is continuous in $\mathcal{N}(z)$ and $\theta(z) > 0$.

(E2) $\nu_x(z) > 0$, $\mu_x^{(p+1)}(\cdot)$, $\nu_x^{(p+1)}(\cdot)$, $\mathcal{E}_x(\cdot)$, and $\mathcal{V}_x(\cdot)$ are continuous in $\mathcal{N}(z)$.

(E3) $h_{\mu,x} \to 0$, $n_- h_{\mu,x} \to \infty$, $n_- h_{\mu,x}^{2p+3} \to d_x^2$ for some $d_x > 0$, $h_{\nu,x}/h_{\mu,x} \to \rho_x$ for some $0 < \rho_x < \infty$, as $n_- \to \infty$.

(E4) $\nu_y(z) > 0$, $\mu_y^{(p+1)}(\cdot)$, $\nu_y^{(p+1)}(\cdot)$, $\mathcal{E}_y(\cdot)$, and $\mathcal{V}_y(\cdot)$ are continuous in $\mathcal{N}(z)$.

(E5) $h_{\mu,y} \to 0$, $n_+ h_{\mu,y} \to \infty$, $n_+ h_{\mu,y}^{2p+3} \to d_y^2$ for some $d_y > 0$, $h_{\nu,y}/h_{\mu,y} \to \rho_y$ for some $0 < \rho_y < \infty$, as $n \to \infty$. "

These five assumptions will lead us to the asymptotic normality of the local polynomial estimators, $\widehat{\mu}_x$, $\widehat{\mu}_y$, $\widehat{\nu}_x$, and $\widehat{\nu}_y$ for the mean functions and variance functions in model (5.1) and (5.2).

The following assumptions: (E6)-(E8), and (E1†)-(E5†) are also cited from Yao *et al.* [64] as:

" (E6) $\mathcal{K}^*$ is uniform continuous, absolutely integrable with respect to Lebesgue measure on $\Re$ and of bounded variation, $\mathcal{K}^*(\mu) \to 0$ as $|\mu| \to \infty$, $\int \{|\mu log(|\mu|)|\}^{1/2} |d\mathcal{K}^*(\mu)| < \infty$.

(E7) $E(|X|^s) < \infty$, $\sup_{z \in \Re_Z} \int |x|^s P(Z, X) dy < \infty$ for some $s \geq 2$, where $P(Z, X)$ is the joint density of (Z, X).

(E8) $E(|Y|^s) < \infty$, $\sup_{z \in \Re_Z} \int |y|^s P(Z, Y) dy < \infty$ for some $s \geq 2$, where $P(Z, Y)$ is the joint density of (Z, Y).

(E1†) $\theta(\cdot) > 0$ and $\theta^{(p+1)}(\cdot)$ is bounded and continuous on $\Re_Z$.

(E2†) On the domain $\Re_Z$, $\nu_x(\cdot) > \delta_x$ for some $\delta_x > 0$ and is bounded, $\mu_x(\cdot)$ is bounded, $\mu_x^{(p+1)}(\cdot)$, $\nu_x^{(p+1)}(\cdot)$, $\mathcal{E}_x(\cdot)$ and $\mathcal{V}_x(\cdot)$ are bounded and continuous.

(E3†) $\Sigma_{n_-} h_{\mu,x}^{\Delta_x} < \infty$ for some $\Delta_x > 0$, $n_-^{2\rho_x - 1} h_{\mu,x} \to \infty$ for some $\rho_x < 1 - s^{-1}$, where $s > 2$ satisfies (E7).

(E4†) On the domain $\Re_Z$, $\nu_y(\cdot) > \delta_y$ for some $\delta_y > 0$ and is bounded, $\mu_y(\cdot)$ is bounded, $\mu_y^{(p+1)(\cdot)}$, $\nu_y^{(p+1)(\cdot)}$, $\mathcal{E}_y(\cdot)$ and $\mathcal{V}_y(\cdot)$ are bounded and continuous.

(E5†) $\Sigma_{n_+} h_{\mu,y}^{\Delta_y} < \infty$ for some $\Delta_y > 0$, $n_+^{2\rho_y - 1} h_{\mu,y} \to \infty$ for some $\rho_y < 1 - s^{-1}$, where $s > 2$ satisfies (E8). ”

Notice that, (E6)-(E8) and (E1†)-(E5†) are necessary assumptions to derive the strong uniformly consistency of the local polynomial estimators, $\widehat{\mu}_x$, $\widehat{\mu}_y$, $\widehat{\nu}_x$, and $\widehat{\nu}_y$ for the mean functions and variance functions in model (5.1) and (5.2).

(E9) $F^*(\cdot)$ and $G^*(\cdot)$ are continuous and monotone increasing on their domains.

E.1 Proof of Theorem 5.1

Assumptions (E1)-(E5) would lead us the asymptotic normality of the local polynomial estimators, $\widehat{\mu}_x$, $\widehat{\mu}_y$, $\widehat{\nu}_x$, and $\widehat{\nu}_y$, for μ_x, μ_y, ν_x, and ν_y. Theorem 4.1 directly follows from lemma 1 in Yao *et al.* [64] and a simple application of the Cramér-Wold device. The partial derivatives of $J_N(z)$ with respect to the mean and variance functions in Theorem 1 are:

$$
\begin{aligned}
\frac{\partial J_N(z)}{\partial \mu_x(z)} &= -\frac{1}{\sqrt{\nu_x(z)}}\phi\left(\frac{c_o(z)-\mu_x(z)}{\sqrt{\nu_x(z)}}\right) \\
&\quad + \frac{\partial c_o(z)}{\partial \mu_x(z)}\left[\frac{1}{\sqrt{\nu_x(z)}}\phi\left(\frac{c_o(z)-\mu_x(z)}{\sqrt{\nu_x(z)}}\right) - \frac{1}{\sqrt{\nu_y(z)}}\phi\left(\frac{\mu_y(z)-c_o(z)}{\sqrt{\nu_y(z)}}\right)\right] \\
\frac{\partial J_N(z)}{\partial \mu_y(z)} &= \frac{1}{\sqrt{\nu_y(z)}}\phi\left(\frac{\mu_y(z)-c_o(z)}{\sqrt{\nu_y(z)}}\right) \\
&\quad + \frac{\partial c_o(z)}{\partial \mu_y(z)}\left[\frac{1}{\sqrt{\nu_x(z)}}\phi\left(\frac{c_o(z)-\mu_x(z)}{\sqrt{\nu_x(z)}}\right) - \frac{1}{\sqrt{\nu_y(z)}}\phi\left(\frac{\mu_y(z)-c_o(z)}{\sqrt{\nu_y(z)}}\right)\right] \\
\frac{\partial J_N(z)}{\partial \nu_x(z)} &= -\frac{1}{2}(c_o(z)-\mu_x(z))\nu_x^{-3/2}\phi\left(\frac{c_o(z)-\mu_x(z)}{\sqrt{\nu_x(z)}}\right) \\
&\quad + \frac{\partial c_o(z)}{\partial \nu_x(z)}\left[\frac{1}{\sqrt{\nu_x(z)}}\phi\left(\frac{c_o(z)-\mu_x(z)}{\sqrt{\nu_x(z)}}\right) - \frac{1}{\sqrt{\nu_y(z)}}\phi\left(\frac{\mu_y(z)-c_o(z)}{\sqrt{\nu_y(z)}}\right)\right] \\
\frac{\partial J_N(z)}{\partial \nu_y(z)} &= -\frac{1}{2}(\mu_y(z)-c_o(z))\nu_y^{-3/2}\phi\left(\frac{\mu_y(z)-c_o(z)}{\sqrt{\nu_y(z)}}\right) \\
&\quad + \frac{\partial c_o(z)}{\partial \nu_y(z)}\left[\frac{1}{\sqrt{\nu_x(z)}}\phi\left(\frac{c_o(z)-\mu_x(z)}{\sqrt{\nu_x(z)}}\right) - \frac{1}{\sqrt{\nu_y(z)}}\phi\left(\frac{\mu_y(z)-c_o(z)}{\sqrt{\nu_y(z)}}\right)\right],
\end{aligned}
$$

and the partial derivatives of $c_o(z)$ with respect to the mean and variance functions are

$$\frac{\partial c_o(z)}{\partial \mu_x(z)} = \frac{b^2 \pm ab(rad)^{-1/2}(-1)}{b^2-1}$$
$$\frac{\partial c_o(z)}{\partial \mu_y(z)} = \frac{-1 \pm ab(rad)^{-1/2}}{b^2-1}$$
$$\frac{\partial c_o(z)}{\partial \nu_x(z)} = -\frac{(\mu_y(z)-\mu_x(z))\nu_y(z)}{(\nu_y(z)-\nu_x(z))^2} \pm$$
$$\left[\frac{\frac{1}{2}\nu_y^{1/2}(z)\nu_x^{-1/2}(z)rad^{1/2} + \frac{1}{2}(\nu_x\nu_y)^{1/2}rad^{-1/2}\left(-ln\frac{\nu_y(z)}{\nu_x(z)} - \frac{\nu_y(z)}{\nu_x(z)} + 1\right)(\nu_y(z)-\nu_x(z))}{(\nu_y(z)-\nu_x(z))^2}\right.$$
$$\left. + \frac{(\nu_x(z)\nu_y(z))rad^{1/2}}{(\nu_y(z)-\nu_x(z))^2}\right]$$
$$\frac{\partial c_o(z)}{\partial \nu_y(z)} = \frac{(\mu_y(z)-\mu_x(z))\nu_x(z)}{(\nu_y(z)-\nu_x(z))^2} \pm$$
$$\left[\frac{\frac{1}{2}\nu_x^{1/2}(z)\nu_y^{-1/2}(z)rad^{1/2} + \frac{1}{2}(\nu_x\nu_y)^{1/2}rad^{-1/2}\left(-ln\frac{\nu_y(z)}{\nu_x(z)} + (\nu_y(z)-\nu_x(z)\frac{1}{\nu_y(z)})\right)(\nu_y(z)-\nu_x(z))}{(\nu_y(z)-\nu_x(z))^2}\right.$$
$$\left. + \frac{(\nu_x(z)\nu_y(z))rad^{1/2}}{(\nu_y(z)-\nu_x(z))^2}\right],$$

where $rad = a^2 + (b^2-1)\nu_x(z)ln(b^2)$, a and b are defined in section 2.2.

E.2 Proof of Theorem 5.2

Assumptions $(E1\dagger)$-$(E5\dagger)$ and (E6)-(E8) would lead us to the strong uniform consistency of the local polynomial estimators, $\widehat{\mu}_x$, $\widehat{\mu}_y$, $\widehat{\nu}_x$, and $\widehat{\nu}_y$, for μ_x, μ_y, ν_x, and ν_y. Then, theorem 4.2 follows from Slutsky's theorem and lemma 2 in Yao *et al.* [64].

E.3 Proof of Theorem 5.3

Let

$$\begin{aligned} J(c;z) &= F_-(X \le c|Z=z) - F_+(Y \le c|Z=z) \equiv F_-(c;z) - F_+(c;z) \\ \widetilde{J}_E(c;z) &= \widetilde{F}_-(c;z) - \widetilde{F}_+(c;z) \end{aligned}$$

where $\widetilde{F}_-(c;z) = \frac{\sum_{i=1}^{n_-} I(x_{i,z} \le c)}{n_-}$, and $\widetilde{F}_+(c;z) = \frac{\sum_{j=1}^{n_+} I(y_{j,z} \le c)}{n_+}$.

By the law of the iterated logarithm (LIL) for empirical process, we have that $\sup_c |\widetilde{F}_-(c;z) - F_-(c;z)| = O(\sqrt{loglogn_-/2n_-})$ $a.s.$, and $\sup_c |\widetilde{F}_+(c;z) - F_+(c;z)| = O(\sqrt{loglogn_+/2n_+})$ $a.s.$. If $n_-/n_+ \to \lambda$, then

$$\begin{aligned}\sup_c |\widetilde{J}_E(c;z) - J(c;z)| &\leq \sup_c |\widetilde{F}_-(c;z) - F_-(c;z)| + \sup_c |\widetilde{F}_+(c;z) - F_+(c;z)| \\ &= O(\sqrt{\log\log n_-/2n_-} + \sqrt{\log\log n_+/2n_+}), a.s.\end{aligned}$$

which indicates the strong convergence of $\widetilde{J}_E(c;z)$ to $J(c;z)$ uniformly on c for a given z. Consequently, for a given z, $\widetilde{c}_{oE}(z)$ converges to $c_o(z)$ almost surely. Straightforwardly, applying the Lebesgue dominated convergence theorem, $E[\widetilde{J}_E(z)]$ converges to $J(z)$ for a given z.

E.4 Proof of Theorem 5.4

Let

$$\widehat{J}_E(c;z) = \widehat{F}_-(c;z) - \widehat{F}_+(c;z),$$

where $\widehat{F}_-(c;z) = \frac{\sum_{i=1}^{n_-} I(\widehat{x}_{i,z} \leq c)}{n_-}$, and $\widehat{F}_+(c;z) = \frac{\sum_{j=1}^{n_+} I(\widehat{y}_{j,z} \leq c)}{n_+}$.

First of all we need to show the uniform consistency of $\widehat{J}(c;z)$ on c for a given z. From the strong uniform consistency of $\widehat{\mu}_x(z)$, $\widehat{\mu}_y(z)$, $\widehat{\nu}_x(z)$, and $\widehat{\nu}_y(z)$, it follows that for a given z,

$$\begin{aligned} I(\widehat{x}_{i,z} \leq c) - I(x_{i,z} \leq c) &\longrightarrow 0, \ a.s. \\ I(\widehat{y}_{j,z} \leq c) - I(y_{j,z} \leq c) &\longrightarrow 0, \ a.s. \end{aligned}$$

uniformly on c for all i. Therefore, for a given z.

$$\begin{aligned} |\widehat{J}_E(c;z) - \widetilde{J}_E(c;z)| &\leq \left| n_-^{-1} \sum_{i=1}^{n_-} (I(\widehat{x}_{i,z} \leq c) - I(x_{i,z} \leq c)) \right| + \left| n_+^{-1} \sum_{j=1}^{n_+} (I(\widehat{y}_{j,z} \leq c) - I(y_{j,z} \leq c)) \right| \\ &\longrightarrow 0, \ a.s. \end{aligned}$$

uniformly on c. Hence, for given $Z = z$,

$$\sup_c |\widehat{J}_E(c;z) - J(c;z)| \leq \sup_c |\widehat{J}_E(c;z) - \widetilde{J}_E(c;z)| + \sup_c |\widetilde{J}_E(c;z) - J(c;z)| \longrightarrow 0, \ a.s.$$

Consequently, for a given z, $\widehat{c}_{oE}(z)$ converges to $c_o(z)$ almost surely.

Now define $\widehat{\delta}_{i,z} = \widehat{x}_{i,z} - \widehat{c}_{oE}(z)$, $\delta_{i,z} = x_{i,z} - \widetilde{c}_{oE}(z)$, $\widehat{\omega}_{j,z} = \widehat{y}_{j,z} - \widehat{c}_{oE}(z)$, and $\omega_{j,z} = y_{j,z} - \widetilde{c}_{oE}(z)$. We have

$$\begin{aligned} E[\{\widehat{J}_E(z) - \widetilde{J}_E(z)\}^2] &= E\left[n_-^{-1}\sum_{i=1}^{n_-}\Big(I(\widehat{\delta}_{i,z} \le 0) - I(\delta_{i,z} \le 0)\Big)\right. \\ &\quad \left. -n_+^{-1}\sum_{j=1}^{n_+}\left(I(\widehat{\omega}_{j,z} \le 0) - I(\omega_{j,z} \le 0)\right)\right]^2 \\ &\le 2\left[E(T_1^2) + E(T_2^2)\right], \end{aligned}$$

where $T_1 = n_-^{-1}\sum_{i=1}^{n_-}\Big(I(\widehat{\delta}_{i,z} \le 0) - I(\delta_{i,z} \le 0)\Big)$, and $T_2 = n_+^{-1}\sum_{j=1}^{n_+}\left(I(\widehat{\omega}_{j,z} \le 0) - I(\omega_{j,z} \le 0)\right)$. Let us explore ET_1^2 first.

$$\begin{aligned} ET_1^2 &= \frac{1}{n_-^2}E\left[\sum_{i=1}^{n_-}\Big(I(\widehat{\delta}_{i,z} \le 0) - I(\delta_{i,z} \le 0)\Big)^2\right. \\ &\quad + \sum_{i \ne i'}\Big(I(\widehat{\delta}_{i,z} \le 0)I(\widehat{\delta}_{i',z} \le 0) + I(\delta_{i,z} \le 0)I(\delta_{i',z} \le 0) \\ &\quad \left. -I(\widehat{\delta}_{i,z} \le 0)I(\delta_{i',z} \le 0) - I(\delta_{i,z} \le 0)I(\widehat{\delta}_{i',z} \le 0)\Big)\right] \\ &\le \frac{1}{n_-} + \frac{1}{n_-^2}\sum_{i \ne i'}\Big[P(\widehat{\delta}_{i,z} \le 0, \widehat{\delta}_{i',z} \le 0) + P(\delta_{i,z} \le 0, \delta_{i',z} \le 0) \\ &\quad -P(\widehat{\delta}_{i,z} \le 0, \delta_{i',z} \le 0) - P(\delta_{i,z} \le 0, \widehat{\delta}_{i',z} \le 0)\Big] \end{aligned}$$

By the strong uniform consistency of $\widehat{\mu}_x(z)$, $\widehat{\mu}_y(z)$, $\widehat{\nu}_x(z)$, and $\widehat{\nu}_y(z)$ and the strong consistency of $\widehat{c}_{oE}(z)$ and $\widetilde{c}_{oE}(z)$, we have that for a given z,

$$\widehat{\delta}_{i,z} \longrightarrow x_{i,z} - c_o(z) \ a.s., \quad \delta_{i,z} \longrightarrow x_{i,z} - c_o(z) \ a.s., \text{ for all } i.$$

So,

$$P(\widehat{\delta}_{i,z} \le 0, \widehat{\delta}_{i',z} \le 0) + P(\delta_{i,z} \le 0, \delta_{i',z} \le 0) - P(\widehat{\delta}_{i,z} \le 0, \delta_{i',z} \le 0) - P(\delta_{i,z} \le 0, \widehat{\delta}_{i',z} \le 0) \longrightarrow 0$$

for all $i \ne i'$. Therefore, $ET_1^2 \longrightarrow 0$ as $n_- \longrightarrow \infty$. Similarly, we can show $ET_2^2 \longrightarrow 0$ as $n_+ \longrightarrow \infty$. Hence, $E\left[\widehat{J}_E(z) - \widetilde{J}_E(z)\right]^2 \longrightarrow 0$.

Appendix F

SIMULATION RESULTS FOR CHAPTER FIVE

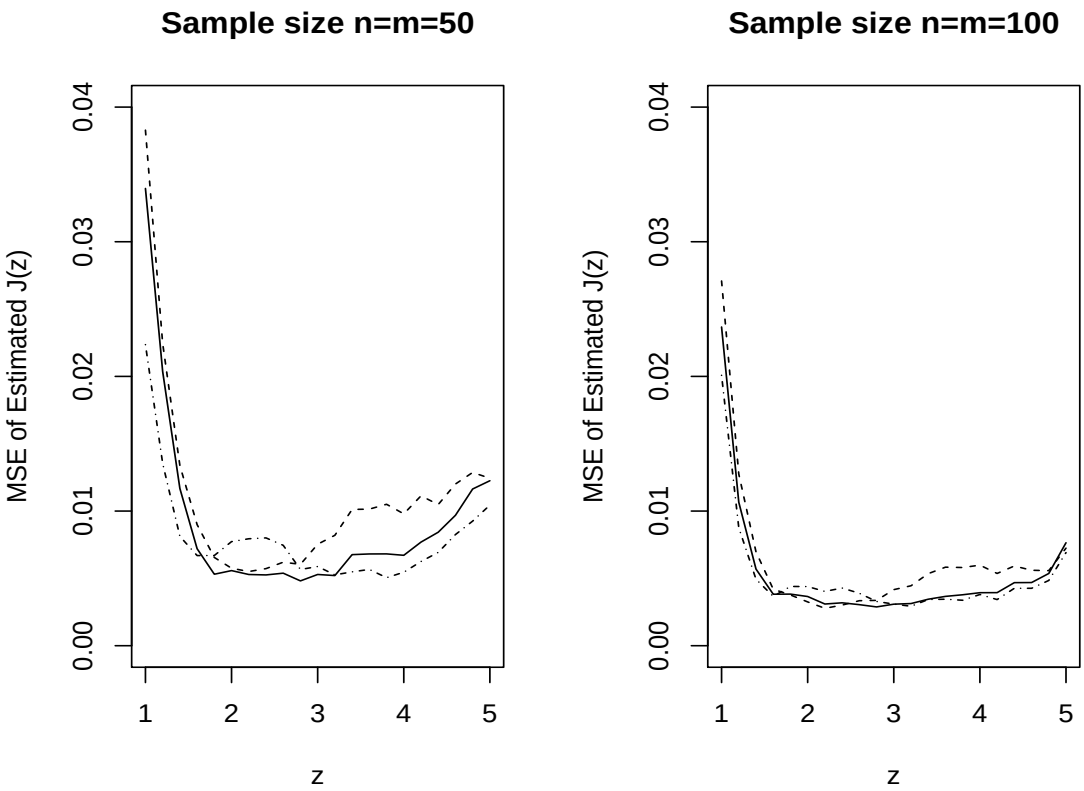

Figure F.1. The MSE's of the estimators when ϵ_1 and ϵ_2 follow the standard normal distribution: solid line for $\widehat{J}_N$, dashed line for $\widehat{J}_E$, and dotdash line for $\widehat{J}_{AC}$.

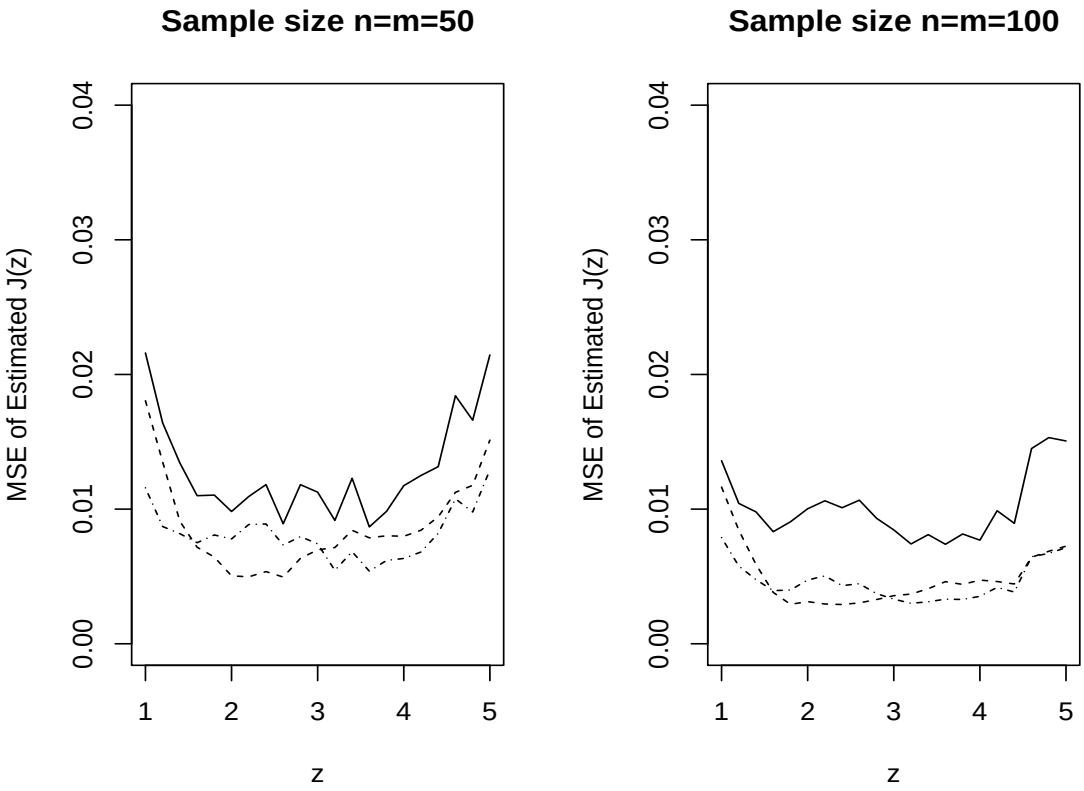

Figure F.2. The MSE's of the estimators when ϵ_1 and ϵ_2 follow t-distribution with degree of freedom 4: solid line for $\widehat{J}_N$, dashed line for $\widehat{J}_E$, and dotdash line for $\widehat{J}_{AC}$.

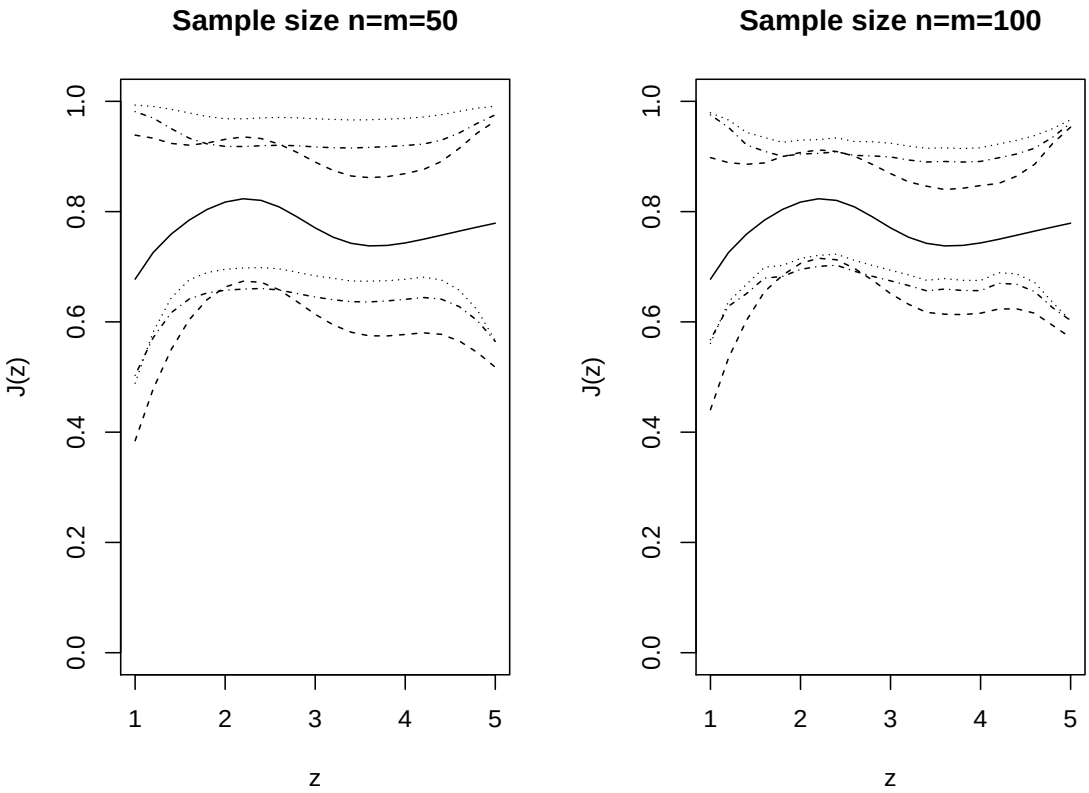

Figure F.3. ϵ_1 and ϵ_2 follow the standard normal distribution. The point-wise confidence bands for $J(z)$: NA band (dashed), BP band (dotted), and ACNA band (dotdash). Solid line is the curve for the true values of $J(z)$.

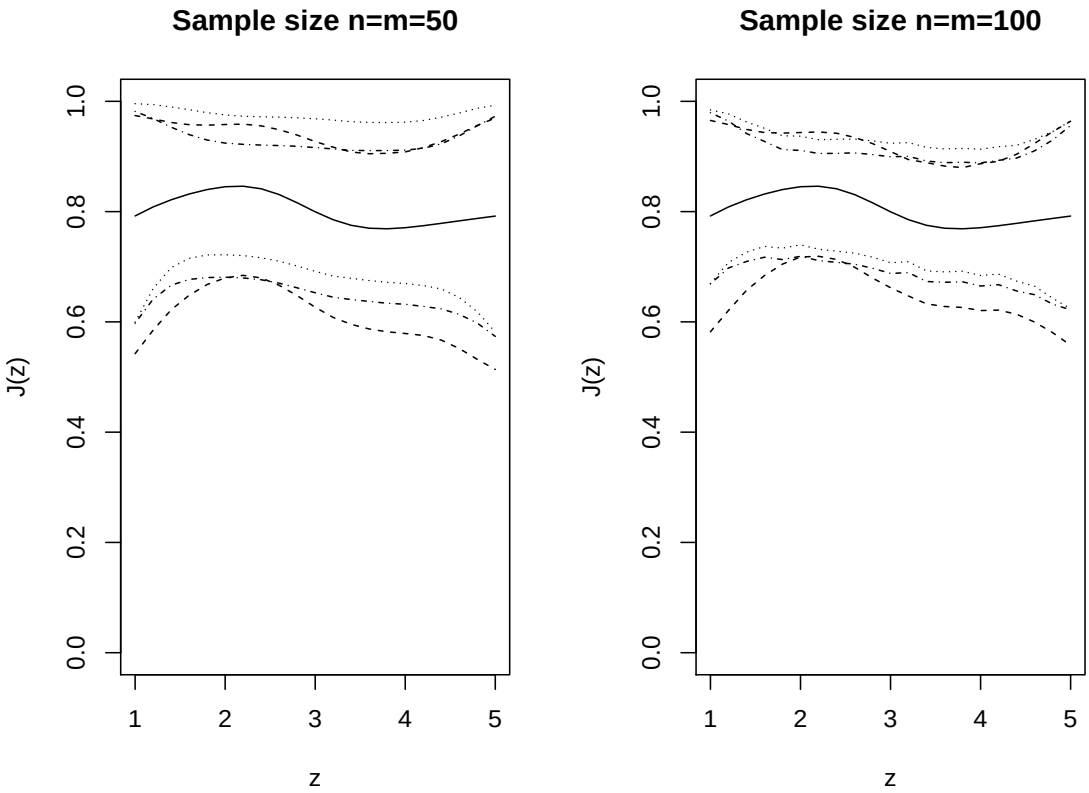

Figure F.4. ϵ_1 and ϵ_2 follow t-distribution with degree of freedom 4. The point-wise confidence bands for $J(z)$: NA band (dashed), BP band (dotted), and ACNA band (dotdash). Solid line is the curve for the true values of $J(z)$.