# Semiparametric Empirical Likelihood Inference under Two-sample Density Ratio Models

by

Meng Yuan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2021

## Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

| | |
|---|---|
| External Examiner: | Art B. Owen<br>Professor, Department of Statistics<br>Stanford University |
| Supervisor(s): | Pengfei Li<br>Professor<br><br>Changbao Wu<br>Professor |
| Internal Member: | Jerald F. Lawless<br>Distinguished Professor Emeritus<br><br>Yingli Qin<br>Associate Professor |
| Internal-External Member: | Thomas Parker<br>Associate Professor, Department of Economics |

## Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Statement of Contributions

A version of Chapter 2 of this thesis has been prepared as a research paper submitted to a journal for publication. A version of Chapter 3 of this thesis has been published by *The Canadian Journal of Statistics*. A version of Chapter 4 has been prepared as a research paper submitted to a journal for publication. All three papers are co-authored with my supervisors, and my contributions include formulating the research problems, deriving main theorems, implementing the proposed methods through R programming, performing the simulations and real data analysis, and writing the initial drafts. A version of Sections 4.2.1 and 4.2.2 in Chapter 4 serves as the main results in a research paper, which has been accepted for publication by *Annals of the Institute of Statistical Mathematics*. This paper is co-authored with Dr. Pengfei Li, Dr. Chunlin Wang, and Mr. Boxi Lin, where my contributions include deriving main theorems, performing partial simulations, analyzing real data examples, and writing the initial draft.

## Abstract

The semiparametric density ratio model (DRM) provides a flexible and useful platform for combining information from multiple sources. It has been widely used in many fields. This thesis considers several important inference problems under two-sample DRMs.

Chapter 1 serves as an introduction. We review the DRM, empirical likelihood, which is a useful inference tool under the DRM, and some applications of DRMs. We also outline the research problems that will be explored in the subsequent chapters.

How to effectively use auxiliary information and data from multiple sources to enhance statistical inference is an important and active research topic in many fields. In Chapter 2, we consider statistical inference under two-sample DRMs with additional parameters, including the main parameters of interest, defined through and/or additional auxiliary information expressed as estimating equations. We examine the asymptotic properties of the maximum empirical likelihood estimators (MELEs) of the unknown parameters in the DRMs and/or defined through estimating equations, and establish the chi-square limiting distributions for the empirical likelihood ratio (ELR) statistics. We show that the asymptotic variance of the MELEs of the unknown parameters does not decrease if one estimating equation is dropped. Similar properties are obtained for inferences on the cumulative distribution function and quantiles of each of the populations involved. We also propose an ELR test for the validity and usefulness of the auxiliary information. Simulation studies show that correctly specified estimating equations for the auxiliary information result in more efficient estimators and shorter confidence intervals. Two real examples are used for illustrations.

The Youden index is a popular summary statistic for receiver operating characteristic curves. It gives the optimal cutoff point of a biomarker to distinguish the diseased and healthy individuals. In Chapter 3, we model the distributions of a biomarker for individuals in the healthy and diseased groups via a DRM. Based on this model, we propose MELEs of the Youden index and the optimal cutoff point. We further establish the asymptotic normality of the proposed estimators and construct valid confidence intervals for the Youden index and the corresponding optimal cutoff point. The proposed method automatically covers both cases when there is no lower limit of detection and when there is a fixed and finite lower limit of detection for the biomarker. Extensive simulation studies and a real-data example are used to illustrate the effectiveness of the proposed method and its advantages over the existing methods.

The Gini index is a popular inequality measure with many applications in social and economic studies. Chapter 4 studies inference on the Gini indices of two semicontinuous

populations. We characterize the distribution of each semicontinuous population by a mixture of a discrete point mass at zero and a continuous skewed positive component. The DRM is then employed to link the positive components of the two distributions. We propose the MELEs of the two Gini indices and their difference, and further investigate the asymptotic properties of the proposed estimators. The asymptotic results enable us to construct confidence intervals and perform hypothesis tests for the two Gini indices and their difference. We show that the proposed estimators are more efficient than the existing fully nonparametric estimators. The proposed estimators and the asymptotic results are also applicable to cases without excessive zero values. Simulation studies show the superiority of our proposed method over existing methods. Two real-data applications are presented using the proposed methods.

In Chapter 5, we summarize our research contributions and discuss some interesting topics, which are related to our current work, for future research.

**Acknowledgements**

First and foremost, I would like to express my deepest gratitude to my supervisors, Dr. Pengfei Li and Dr. Changbao Wu, for their guidance, support, and encouragement during my Ph.D. studies. I would have never experienced the fun of the statistical world without their continuous inspiration, endless patience, and profound knowledge. Their high standard and rigorous attitude towards academic research pushed me to become a better version of myself. It was my great honor to be their student. They will continue to be my role models of knowledge, professionalism, and wisdom.

I would like to thank all my committee members, Dr. Art B. Owen, Dr. Jerald F. Lawless, Dr. Thomas Parker, and Dr. Yingli Qin for dedicating their time to reviewing my work and providing inspiring comments. Special thanks go to my master supervisor Dr. Paul Marriott for his guidance. I also thank Dr. Chunlin Wang, who discussed with me frequently about my research and enriched my knowledge.

I would like to thank my friends, Junhan, Menglu, Yidan, and Yilin. I am very grateful to have them around to exchange thoughts and share happiness and sadness in the entire journey. I would also like to thank Shuang, Tao, and Xueqiao, who have been accompanied me since my first day in Canada. The time when we got together and had endless talks are delightful and invaluable memories for me. I thank Ms. Mary Lou Dufton and Mr. Greg Preston for their continuous help in administrative assistance. Many thanks also go to all my friends and department staff members who are not listed here for their kindness and help.

Last but not least, I would like to express my heartfelt appreciation to Yuyu for his love, support, and tolerance. Ph.D. studies have never been an easy journey for me, and the COVID-19 has made it even harder. It would never be possible for me to complete the dissertation without the company of Yuyu.

## Dedication

*To my beloved parents, grandma and sister for their unconditional love and support.*
*To the memory of my dearest grandpa.*

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Two-sample Density Ratio Models

This thesis proposes new and effective procedures for several important inference problems under two-sample density ratio models (DRMs). In this section, we introduce the definition of two-sample DRMs, and explain relationships between DRMs and other commonly used statistical models.

Suppose we have two independent random samples $\{X_{01}, \ldots, X_{0n_0}\}$ and $\{X_{11}, \ldots, X_{1n_1}\}$ from two populations with cumulative distribution functions (CDFs) $F_0$ and $F_1$, respectively. The dimension of $X_{ij}$ can be one or greater than one. Let $dF_i$ denote the density of $F_i$ for $i = 0, 1$. The two-sample DRM (Anderson, 1979; Qin, 2017) postulates

$$dF_1(x) = \exp\{\alpha + \boldsymbol{\beta}^\top \boldsymbol{q}(x)\} dF_0(x) = \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} dF_0(x), \tag{1.1}$$

where $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top)^\top$ are unknown parameters for the DRM; $\boldsymbol{Q}(x) = (1, \boldsymbol{q}(x)^\top)^\top$ with the basis function $\boldsymbol{q}(x)$ being a pre-specified, non-trivial function of dimension $d$; and the baseline distribution $F_0$ is unspecified. The DRMs can be broadened to allow for multi-sample cases; see Wang (2017) and the references therein. Throughout this thesis, DRMs refer to (1.1) for two samples.

The DRM in (1.1) links two distribution functions through a parametric form for the log density ratio, which helps to utilize information across two samples. Meanwhile, the baseline distribution $F_0$ remains completely unspecified. The parametric and nonparametric components together make DRMs semiparametric and flexible to embrace many commonly used statistical models.

The DRM has a natural connection to the well-studied logistic regression (Anderson, 1979; Qin and Zhang, 1997) if one treats $D = 0$ and 1 as indicators for the observations from $F_0$ and $F_1$, respectively. On the one hand, conditional on $X = x$, the logistic regression is

$$P(D = 1|X = x) = \frac{\exp\{\alpha^* + \boldsymbol{\beta}^\top \boldsymbol{q}(x)\}}{1 + \exp\{\alpha^* + \boldsymbol{\beta}^\top \boldsymbol{q}(x)\}}. \tag{1.2}$$

On the other hand, if the DRM (1.1) is satisfied, using the Bayes' formula gives

$$P(D = 1|X = x) = \frac{dF_1(x)P(D = 1)}{dF_1(x)P(D = 1) + dF_0(x)P(D = 0)} = \frac{\exp\{\alpha^* + \boldsymbol{\beta}^\top \boldsymbol{q}(x)\}}{1 + \exp\{\alpha^* + \boldsymbol{\beta}^\top \boldsymbol{q}(x)\}}, \tag{1.3}$$

where $\alpha^* = \alpha + \log\{P(D = 1)/P(D = 0)\}$. Hence, the DRM is equivalent to the logistic regression model. Because of that, the inference procedures for logistic regression models, which are extensively investigated in the literature, provide some ideas to explore the properties of DRMs. For example, the equivalence brings in the computational convenience when calculating the estimate of $\boldsymbol{\theta}$ in the DRM.

The DRM also includes many commonly used distributions as special cases. Any two distributions in the same exponential family satisfy the DRM (1.1) with certain $\boldsymbol{q}(x)$. We say a distribution belongs to the exponential family if the corresponding probability density function or probability mass function takes the following form (Kay and Little, 1987),

$$f(x; \boldsymbol{\xi}) = A(x)B(\boldsymbol{\xi}) \exp\{\boldsymbol{h}(\boldsymbol{\xi})^\top \boldsymbol{g}(x)\}, \tag{1.4}$$

with the support of the distribution not depending on the parameter $\boldsymbol{\xi}$. Suppose the two distributions $F_0$ and $F_1$ are from the same exponential family with different parameters $\boldsymbol{\xi}_0$ and $\boldsymbol{\xi}_1$. Then

$$\frac{dF_1(x)}{dF_0(x)} = \frac{f(x; \boldsymbol{\xi}_1)}{f(x; \boldsymbol{\xi}_0)} = \exp[\log\{B(\boldsymbol{\xi}_1)/B(\boldsymbol{\xi}_0)\} + \{\boldsymbol{h}(\boldsymbol{\xi}_1) - \boldsymbol{h}(\boldsymbol{\xi}_0)\}^\top \boldsymbol{g}(x)].$$

This suggests that $F_0$ and $F_1$ satisfy the DRM in (1.1) with $\alpha = \log\{B(\boldsymbol{\xi}_1)/B(\boldsymbol{\xi}_0)\}$, $\boldsymbol{\beta} = \boldsymbol{h}(\boldsymbol{\xi}_1) - \boldsymbol{h}(\boldsymbol{\xi}_0)$, and $\boldsymbol{q}(x) = \boldsymbol{g}(x)$. Note that in the DRM (1.1), the baseline $F_0$ is left unspecified and $\boldsymbol{q}(x)$ is the only parametric component that needs to be specified. Hence the DRM assumptions are weaker than the fully parametric model assumptions in (1.4). For example, the basis function $\boldsymbol{q}(x) = \log x$ includes two log-normal distributions with the same variance with respect to the log-scale, as well as two gamma distributions with the same scale parameter; the basis function $\boldsymbol{q}(x) = x$ embraces two normal distributions with different means but a common variance and two exponential distributions with different rates. We refer to Kay and Little (1987), Cai (2014), and Wang (2017) for more examples.

The DRMs are inherently biased sampling models with weight functions involving unknown parameters (Qin, 1998). More precisely, let $F_0$ be the interested but unknown distribution and $F_1$ be the distribution resulted from biased sampling of $F_0$ according to the weight function $w(x; \boldsymbol{\eta})$ with unknown parameter $\boldsymbol{\eta}$. Then the biased sampling model with the weight function $w(x; \boldsymbol{\eta})$ gives the density of $F_1$ by (Rao, 1965)

$$dF_1(x) = \frac{w(x; \boldsymbol{\eta}) dF_0(x)}{\int w(x; \boldsymbol{\eta}) dF_0(x)}.$$

For example, the choice of $w(x; \boldsymbol{\eta}) = x$ is related to a length-biased sampling (Qin, 1993). It is clear that the biased sampling model with the weight function $w(x; \boldsymbol{\eta}) = \exp\{\boldsymbol{\beta}^\top \boldsymbol{q}(x)\}$ satisfy the DRM (1.1).

We wrap up this section with some discussion on the choice of $\boldsymbol{q}(x)$ in the DRM (1.1) in applications. To use the DRM, we need to specify $\boldsymbol{q}(x)$ in advance. If the practitioners believe that a logistic regression model in (1.2) with $\boldsymbol{q}(x) = x$ is adequate to describe the relationship between $D$ and $X$, then they can use the DRM (1.1) with $\boldsymbol{q}(x) = x$. If it is believed that gamma distributions or normal distributions provide good fit to $F_0$ and $F_1$, then they can use the semiparametric DRM (1.1) with $\boldsymbol{q}(x) = (x, \log x)^\top$ or $(x, x^2)^\top$ instead of a parametric model to achieve robustness of inferences. The DRM (1.1) with a particular choice of $\boldsymbol{q}(x)$ can be further checked by the goodness-of-fit test discussed in Qin and Zhang (1997). The details of this test will be provided at the end of Section 1.2.2.

## 1.2 Empirical Likelihood Inference under DRMs

A nice property of the DRM is that it permits elegant inference solutions through empirical likelihood. In this section, we briefly review the empirical likelihood for one-sample problems and then apply it to two-sample problems under the DRM.

### 1.2.1 One sample empirical likelihood

The empirical likelihood is first introduced by Owen (1988) to mimic the parametric likelihood. Since Owen's seminar paper, the empirical likelihood has become remarkably popular because it has many nice properties corresponding to those of parametric likelihood methods, e.g., the empirical likelihood ratio (ELR) confidence region is range-respecting, transformation-invariant, and Bartlett correctable. More importantly, the ELR statistic obeys Wilks' theorem (Owen, 1990; Hall and La Scala, 1990; DiCiccio et al., 1991; Owen, 1991; Qin and Lawless, 1994). In this subsection, we mainly review the empirical likelihood method for making inference on the population mean and the parameters defined through estimating equations. We refer to Owen (2001)'s monograph for a comprehensive review and discussion of the empirical likelihood.

Let $\{X_1, \cdots, X_n\}$ be independent observations from a population with completely unknown CDF $F$. The likelihood function of $F$ is defined as (Owen, 1988)

$$L(F) = \prod_{i=1}^{n} \{F(X_i) - F(X_i^-)\}.$$

Following Owen (1988), the sample-based version of $F$ is given by

$$F^*(x) = \sum_{n=1}^{n} p_i I(X_i \leq x),$$

where $p_i = F(X_i) - F(X_i^-)$ and $I(\cdot)$ is an indicator variable. Note that $p_i$'s should satisfy the constraints

$$p_i > 0 \text{ and } \sum_{i=1}^{n} p_i = 1 \tag{1.5}$$

to ensure that $F^*$ is a CDF. The maximizer of the likelihood function $L(F) = \prod_{i=1}^{n} p_i$ subject to the constraints in (1.5) corresponds to the empirical CDF $F_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \leq x)$. The ELR function is then defined as

$$R(F) = L(F)/L(F_n) = \prod_{i=1}^{n} np_i.$$

Consider the population mean $\mu$ as the parameter of interest. We assume the dimension of $\mu$ is one for simplicity. By using Lagrange multipliers, we profile out $p_i$'s and obtained the profile ELR function of $\mu$ as (Owen, 1988)

$$R_n(\mu) = \sup_{p_1, \cdots, p_n} \left\{ \prod_{i=1}^{n} np_i : p_i > 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i X_i = \mu \right\}.$$

Let $\mu^*$ be the true value of $\mu$. It has been shown that under some mild moment conditions, the ELR statistic $-2 \log R_n(\mu^*)$ asymptotically follows a chi-square distribution with one degree of freedom, which is a nonparametric version of the Wilks' theorem. Similar properties remain valid when the empirical likelihood method is applied to other functionals such as those in Owen (1988, 1990), and to linear regression model (Owen, 1991).

Next, we discuss the empirical likelihood inference for parameters defined through estimating equations (Qin and Lawless, 1994). Suppose the parameter of interest, $\boldsymbol{\psi}$, is of dimension $p$. The information of $\boldsymbol{\psi}$ and $F$ are available through a set of $r \geq p$ functionally independent unbiased estimating equations:

$$E\{\boldsymbol{g}(X; \boldsymbol{\psi})\} = \mathbf{0}$$

with $\boldsymbol{g}(x; \boldsymbol{\psi}) = (g_1(x; \boldsymbol{\psi}), \cdots, g_r(x; \boldsymbol{\psi}))^\top$. In this case, the profile likelihood function of $\boldsymbol{\psi}$ takes the form

$$L_n(\boldsymbol{\psi}) = \sup_{p_1, \cdots, p_n} \left\{ \prod_{i=1}^n p_i : p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \boldsymbol{g}(X_i; \boldsymbol{\psi}) = \mathbf{0} \right\}.$$

The maximum empirical likelihood estimator (MELE) of $\boldsymbol{\psi}$, denoted as $\hat{\boldsymbol{\psi}}$, is the maximizer of $L_n(\boldsymbol{\psi})$ with respect to $\boldsymbol{\psi}$. Qin and Lawless (1994) showed that under some regularity conditions, the estimator $\hat{\boldsymbol{\psi}}$ is consistent and asymptotically normal, and its asymptotic variance cannot decrease if an estimating equation is dropped when $r > p$. They further proved that the ELR statistic $-2 \log\{L_n(\boldsymbol{\psi}^*)/L_n(\hat{\boldsymbol{\psi}})\}$ converges in distribution to a chi-square distribution with $p$ degree of freedom, where $\boldsymbol{\psi}^*$ is the true value of $\boldsymbol{\psi}$.

### 1.2.2 Empirical likelihood under two-sample DRMs

Since the baseline distribution $F_0$ in DRMs (1.1) is unspecified, it is natural to adopt the empirical likelihood for inference under the DRM. Suppose we have two random samples

$$X_{01}, \ldots, X_{0n_0} \sim F_0 \quad \text{and} \quad X_{11}, \ldots, X_{1n_1} \sim F_1,$$

and two CDFs $F_0$ and $F_1$ are linked through the DRM (1.1). Based on the two samples, the full empirical likelihood is

$$\prod_{i=0}^{1} \prod_{j=1}^{n_i} dF_i(X_{ij}),$$

with $dF_i(X_{ij}) = F_i(X_{ij}) - F_i(X_{ij}^-)$. Following the empirical likelihood principle of Owen (2001) and with the help of DRM (1.1), we use the combined sample to estimate the baseline distribution $F_0$ as

$$F_0^*(x) = \sum_{i=0}^{1}\sum_{j=1}^{n_i} p_{ij} I(X_{ij} \le x), \tag{1.6}$$

where $p_{ij} = dF_0(X_{ij})$. The DRM (1.1) and (1.6) together imply that

$$F_1^*(x) = \sum_{i=0}^{1}\sum_{j=1}^{n_i} \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} p_{ij} I(X_{ij} \le x). \tag{1.7}$$

The fact that both $F_0^*$ and $F_1^*$ are CDFs introduces the following constraints:

$$p_{ij} > 0, \quad \sum_{i=0}^{1}\sum_{j=1}^{n_i} p_{ij} = 1, \quad \sum_{i=0}^{1}\sum_{j=1}^{n_i} p_{ij} \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} = 1. \tag{1.8}$$

With (1.6) and (1.7), the likelihood function under the DRM (1.1) then becomes

$$\mathcal{L}_n = \left\{ \prod_{i=0}^{1}\prod_{j=1}^{n_i} p_{ij} \right\} \left[ \prod_{j=1}^{n_1} \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{1j})\} \right]. \tag{1.9}$$

Using Lagrange multipliers and for any given $\boldsymbol{\theta}$, it can be shown that the maximum of $\mathcal{L}_n$ is reached at

$$p_{ij} = n^{-1}\{1 + \lambda[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1]\}^{-1}, \tag{1.10}$$

where the Lagrange multiplier $\lambda$ satisfies

$$\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1}{1 + \lambda[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1]} = 0. \tag{1.11}$$

It is apparent that $\lambda$ depends on the given $\boldsymbol{\theta}$.

Plugging (1.10) into (1.9) and taking the logarithm, the profile log-likelihood function of $\boldsymbol{\theta}$ (Qin, 1998), up to a constant not depending on $\boldsymbol{\theta}$, is given by

$$\ell(\boldsymbol{\theta}) = -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \log\{1 + \lambda[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1]\} + \sum_{j=1}^{n_1} \boldsymbol{\theta}^\top \boldsymbol{Q}(X_{1j}).$$

The MELE of $\boldsymbol{\theta}$ is then defined as

$$\tilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}).$$

Consequently, the MELEs of $p_{ij}$'s are

$$\tilde{p}_{ij} = n^{-1}\{1 + \tilde{\lambda}[\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} - 1]\}^{-1}, \tag{1.12}$$

where $\tilde{\lambda}$ is the Lagrange multiplier corresponding to $\tilde{\boldsymbol{\theta}}$, or equivalently, $\tilde{\lambda}$ is the solution of (1.11) with $\tilde{\boldsymbol{\theta}}$ in the place of $\boldsymbol{\theta}$.

The Lagrange multiplier $\lambda$ defined through (1.11) usually does not have a closed form. Because of that, maximizing $\ell(\boldsymbol{\theta})$ to numerically obtain $\tilde{\boldsymbol{\theta}}$ may not be an easy task. Keziou and Leoni-Aubin (2008) and Cai et al. (2017) pointed out $\tilde{\boldsymbol{\theta}}$ also maximizes the following dual empirical log-likelihood function:

$$\ell_{nd}(\boldsymbol{\theta}) = -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \log\left\{1 + \frac{n_1}{n}\left[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1\right]\right\} + \sum_{j=1}^{n_1} \boldsymbol{\theta}^\top \boldsymbol{Q}(X_{1j}), \tag{1.13}$$

i.e.,

$$\tilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell_{nd}(\boldsymbol{\theta}). \tag{1.14}$$

We now provide some details for the claim in (1.14). We first argue that $\tilde{\lambda} = n_1/n$.

Note that $\tilde{\boldsymbol{\theta}}$ satisfies

$$\begin{aligned}
0 &= \frac{\partial \ell(\tilde{\boldsymbol{\theta}})}{\partial \alpha} \\
&= -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\tilde{\lambda}\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}}{1+\tilde{\lambda}[\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}-1]} - \sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{[\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}-1]\cdot \frac{d\lambda}{d\alpha}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}}{1+\tilde{\lambda}[\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}-1]} + n_1 \\
&= -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\tilde{\lambda}\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}}{1+\tilde{\lambda}[\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}-1]} + n_1 \\
&= -n\tilde{\lambda} + n_1,
\end{aligned}$$

where we have used (1.11) in the third step, and the following fact in the last step:

$$\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\tilde{\lambda}\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}}{1+\tilde{\lambda}[\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}-1]} = \sum_{i=0}^{1}\sum_{j=1}^{n_i} n\tilde{\lambda}\tilde{p}_{ij}\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} = n\tilde{\lambda}.$$

Hence, $\tilde{\lambda} = n_1/n$ and

$$\frac{\partial \ell_{nd}(\tilde{\boldsymbol{\theta}})}{\partial \alpha} = -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\frac{n_1}{n}\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}}{1+\frac{n_1}{n}[\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}-1]} + n_1 = 0. \tag{1.15}$$

For $\partial \ell_{nd}(\tilde{\boldsymbol{\theta}})/\partial \boldsymbol{\beta}$, we notice that

$$0 = \frac{\partial \ell(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}} = \frac{\partial \ell_{nd}(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}} - \sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{[\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}-1]\cdot \frac{d\lambda}{d\boldsymbol{\beta}}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}}{1+\tilde{\lambda}[\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}-1]} = \frac{\partial \ell_{nd}(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}}, \tag{1.16}$$

where we have used (1.11) in the last step. Therefore, (1.15) and (1.16) together imply that

$$\frac{\partial \ell_{nd}(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = 0.$$

That is, $\tilde{\boldsymbol{\theta}}$ is a stationary point of $\ell_{nd}(\boldsymbol{\theta})$. It can be checked that $\ell_{nd}(\boldsymbol{\theta})$ is a concave function of $\boldsymbol{\theta}$. Hence $\tilde{\boldsymbol{\theta}}$ further maximizes $\ell_{nd}(\boldsymbol{\theta})$ and (1.14) is proved.

Note that $\ell_{nd}(\boldsymbol{\theta})$ can be rewritten as

$$\ell_{nd}(\boldsymbol{\theta}) = \ell^*(\alpha^*, \boldsymbol{\beta}) - n_0\log(n_0/n) - n_1\log(n_1/n),$$

where $\ell^*(\alpha^*, \boldsymbol{\beta})$ is the log-likelihood function of logistic regression model in (1.2) with $\alpha^* = \alpha + \log(n_1/n_0)$. This fact and the result in (1.14) together make the computation of $\tilde{\boldsymbol{\theta}}$ very straightforward. For instance, we can directly use the existing R functions such as *glm* for such purpose.

With $\tilde{\boldsymbol{\theta}}$ and the fact that $\tilde{\lambda} = n_1/n$, the MELEs of $p_{ij}$'s are computed as

$$\tilde{p}_{ij} = n^{-1}\left\{1 + \frac{n_1}{n}\left[\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} - 1\right]\right\}^{-1}.$$

We then estimate the CDFs $F_0$ and $F_1$ as

$$\tilde{F}_0(x) = \sum_{i=0}^{1}\sum_{j=1}^{n_i}\tilde{p}_{ij} I(X_{ij} \le x) \quad\text{and}\quad \tilde{F}_1(x) = \sum_{i=0}^{1}\sum_{j=1}^{n_i}\exp\{\tilde{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}\tilde{p}_{ij} I(X_{ij} \le x). \quad (1.17)$$

The estimators of other interesting population quantities, such as population mean and quantiles, can be constructed from $\tilde{F}_0$ and $\tilde{F}_1$.

The two estimates $\tilde{F}_0$ and $\tilde{F}_1$ can also be used to construct a goodness-of-fit test to check the validity of the DRM. Qin and Zhang (1997) defined a Kolmogorov-Smirnov-type test statistic as

$$\Delta_n = \sup_{-\infty \le x \le \infty} \sqrt{n}|\tilde{F}_0(x) - \bar{F}_0(x)|,$$

where $\bar{F}_0(x) = n_0^{-1}\sum_{j=1}^{n_0} I(X_{0j} \le x)$ represents the empirical CDF of $F_0$. We reject the null hypothesis that the DRM (1.1) is satisfied if $\Delta_n$ is greater than some critical value. The limiting distribution of $\Delta_n$ has a complicated form. Qin and Zhang (1997) suggest to use a Bootstrap method to find the critical value.

## 1.3 Literature Review: Applications and Developments of DRMs

In the existing literature, the DRM has been investigated extensively because of its flexibility and efficiency. In this section, we mainly review the applications and developments of DRMs related to the research problems that we will study in later chapters.

Using the equivalence of the logistic regression model and the DRM, Qin (1998) studied the inference problems for the parameters of logistic regression models based on retrospective case-control data. Qin et al. (2015) developed much improved methods for a retrospective case-control study under the logistic regression model by utilizing auxiliary information from public registration databases.

The DRM is widely used to link the related distributions and improve inference on population quantities. Zhang (2000), Chen and Liu (2013), and Cai and Chen (2018) considered the quantile estimation under the DRMs. They showed that the MELEs of quantiles admit Bahadur representation and are more efficient than empirical quantiles. Recently, Zhang et al. (2020) investigated the ELR statistic for quantiles and showed that the ELR-based confidence region of the quantiles is preferable to the Wald-type confidence region. Li et al. (2018) studied the MELE and compared the dual empirical likelihood Wald-type confidence interval (CI) and the profile-ELR-based CI for some specific one-dimensional parameter (see Section 2.1.2 for more details). Their simulation shows that the profile-ELR-based CI has better performance in terms of coverage probability. Fokianos (2004) and Qin and Zhang (2005) discussed density estimation under the DRMs. They show that the density estimators under the DRM are more efficient than the usual kernel density estimators.

The DRM is also applied to multi-sample hypothesis-testing problems. Fokianos et al. (2001) considered the homogeneity test of distributions under the DRM by using a Wald-type test. Keziou and Leoni-Aubin (2008) and Cai et al. (2017) investigated the ELR test for the homogeneity of distributions under the DRM. Wang et al. (2017a, 2018) further developed the ELR statistics for testing the homogeneity of distributions and the equality of population means, respectively, for multiple samples with excessive zeros. The simulation studies in these papers all show that the proposed tests under the DRMs are more powerful than the fully nonparametric tests.

The DRM has gained popularity in the field of receiver operating characteristic (ROC) analysis. Qin and Zhang (2003) investigated the estimation of the ROC curve as well as one of its popular summary statistic, the area under the curve (AUC). Zhang (2006) proposed a Wald-type statistic to test whether the accuracy of a diagnostic test is acceptable in terms of the AUC. Wan and Zhang (2007) constructed a smoothed ROC curve estimator based on kernel technique. Wang and Zhang (2014) built an ELR-based CI for the AUC, which is shown to be more robust than a fully parametric method, and more efficient than a fully nonparametric approach. Wan and Zhang (2008) and Zhang and Zhang (2014) studied the inference problems for the difference of AUCs for two correlated ROC curves.

The DRM has also been employed for inference based on censored samples. Shen et al. (2007) considered the conditional empirical likelihood to make inference with the randomly right-censored data. Ren (2008) performed the inference for various types of censored data using the weighted empirical likelihood approach. Jiang and Tu (2012) compared the performance of conditional and weighted empirical likelihood inference with the randomly right-censored data. Wang et al. (2011) developed the empirical likelihood inference based on the right-censored data with fixed censoring points while Shen et al. (2012) and Wei and Zhou (2016) dealt with the empirical likelihood inference for randomly right-censored data. Cai and Chen (2018) used dual empirical likelihood to perform inference for left- and/or right-censored samples with fixed censoring points.

Diagnosis of the DRMs and selecting the basis function $\boldsymbol{q}(x)$ for the DRMs have been a topic with extensive discussion. The estimators of the quantities of interest may suffer from bias and loss of efficiency under a misspecified DRM (Fokianos and Kaimi, 2006). Qin and Zhang (1997) and Zhang (2002) proposed goodness-of-fit tests to examine the validity of DRMs for a pre-specified $\boldsymbol{q}(x)$. We refer to Section 1.2 for more details. Fokianos et al. (1999) proposed a generalized-moments specification test for the logistic link. Box-Cox family of transformations are suggested by Fokianos and Kaimi (2006) to choose the basis function $\boldsymbol{q}(x)$ in the DRM, which may help reduce the negative effect caused by the model misspecification. Zhang and Chen (2021) proposed to use functional principal component analysis method to choose a data-adaptive basis function. The equivalence between DRMs and logistic regression models also provides a direction for selecting the basis function in the DRMs. Fokianos (2007) adjusted some popular selection criteria, such as Akaike information criterion and Bayesian information criterion, for selecting the basis function under the DRMs.

There are other applications of DRMs, including inference under semiparametric mixture models (Qin, 1999; Zou et al., 2002; Li and Qin, 2011; Li et al., 2017), the modeling of multivariate extremal distributions (de Carvalho and Davison, 2014), and dominance index estimation (Zhuang et al., 2019).

## 1.4 Outline of the Thesis

In this thesis, we use the empirical likelihood to develop new and effective procedures for three important inference problems: (1) inference under two-sample DRMs with additional parameters defined through and/or additional auxiliary information expressed as estimating equations; (2) inference on the Youden index and the optimal cutoff point under two-sample DRMs; (3) inference on the Gini indices of two semicontinuous populations.

With the increasing availability of data sources, utilizing the auxiliary information to enhance statistical inference is of great interest. Inspired by Qin and Lawless (1994), estimating equations would provide a unified platform for the use of auxiliary information and inferences on the main parameters of interest such as the moments and quantiles of population under DRMs. In Chapter 2, we propose general semiparametric inference procedures to utilize the combined information from two samples as well as auxiliary information. We model the CDFs of two populations by a DRM and assume that auxiliary information about the CDFs and interested parameters are expressed through estimating equations. We investigate the theoretical properties for the MELEs of the unknown parameters as well as the ELR statistics on these parameters. The inference procedures on the CDFs and population quantiles are also studied. It should be noted that misspecified auxiliary information could have adverse effect on statistical inference. We further develop an ELR test for checking the validity and usefulness of auxiliary information.

The Youden index is a widely-used summary statistic of the ROC curve and has the advantage of providing a criterion to choose the "optimal" cutoff point of a biomarker to distinguish the diseased and healthy individuals. Inference on the Youden index and the optimal cutoff point has been studied based on either parametric methods or nonparametric methods. The former relies on parametric assumptions on the CDFs of biomarkers in the diseased and healthy groups, while the latter produces inefficient estimators of the optimal cutoff point (Fluss et al., 2005; Bantis et al., 2019; Hsieh and Turnbull, 1996). In Chapter 3, we propose to link the distributions of the biomarkers in the diseased and healthy groups via a semiparametric DRM, and obtain the MELEs of the Youden index and the corresponding optimal cutoff point. The asymptotic properties of the estimators are explored, which enables us to construct valid CIs for the Youden index and the optimal cutoff point. The measurement of a biomarker may be unquantifiable below a limit of detection and missing from the dataset in applications (Ruopp et al., 2008; Bantis et al., 2017). Our proposed method covers both cases with and without a fixed and finite lower limit of detection.

The Gini index is a popular inequality measure with many applications in social and economic studies. Many studies of the Gini index have applied nonparametric methods and often focused on a single population (Hoeffding, 1948; Qin et al., 2010; Peng, 2011; Wang et al., 2016). In applications, two related populations often share some common characteristics, which is ignored by the nonparametric methods. In addition, it is common to encounter semicontinuous data with a mixture of excessive zero values and positive outcomes in practice (Zhou and Cheng, 2008). In Chapter 4, we consider the inference on Gini indices of two semicontinuous populations. We model the distribution of each semicontinuous population by a mixture of a discrete point mass at zero and a continuous positive component, and further adopt a DRM to link the two positive components to utilize the information from both population. Base on these models, we first establish theoretical results for the MELEs of model parameters and a class of functionals. With these preliminary results, we propose the MELEs of two Gini indices and their difference, and study the asymptotic properties of these estimators. The proposed estimators and the asymptotic results are also applicable to cases when there is no excess of zero values.

Chapter 5 concludes the thesis with a brief summary of our research contributions and provides some potential topics worthy of further investigation.

# Chapter 2

# Empirical Likelihood Inference with Estimating Equations under Density Ratio Models

## 2.1 Introduction

### 2.1.1 Problem setup

Chapter 2 is devoted to developing statistical methods for inference problems under two-sample DRMs with estimating equations. Suppose we have two independent random samples $\{X_{01}, \ldots, X_{0n_0}\}$ and $\{X_{11}, \ldots, X_{1n_1}\}$ from two distributions $F_0$ and $F_1$, respectively. The dimension of $X_{ij}$ can be one or greater than one. We assume that the CDFs $F_0$ and $F_1$ are linked through the DRM (1.1), i.e.,

$$dF_1(x) = \exp\{\alpha + \boldsymbol{\beta}^\top \boldsymbol{q}(x)\} dF_0(x) = \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} dF_0(x), \tag{2.1}$$

where $dF_i(x)$ denotes the density of $F_i(x)$ for $i = 0$ and 1; $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top)^\top$ are the unknown parameters for the DRM; $\boldsymbol{Q}(x) = (1, \boldsymbol{q}(x)^\top)^\top$ with $\boldsymbol{q}(x)$ being the basis function of dimension $d$; and the baseline distribution $F_0$ is unspecified. We further assume that the main parameters of interest can be expressed and/or certain auxiliary information about $F_0$, $F_1$, and $\boldsymbol{\theta}$ is available in the form of functionally independent unbiased estimating equations:

$$E_0\{\boldsymbol{g}(X; \boldsymbol{\psi}, \boldsymbol{\theta})\} = \boldsymbol{0}, \tag{2.2}$$

where $E_0(\cdot)$ refers to the expectation operator with respect to $F_0$, the vector of parameters $\boldsymbol{\psi}$ consists of the main parameters of interest and/or nuisance parameters and has dimension $p$, $\boldsymbol{g}(\cdot;\cdot)$ is $r$-dimensional, and $r \geq p$. In this chapter, our goal is twofold:

(1) we develop new and general semiparametric inference procedures for $(\boldsymbol{\psi}, \boldsymbol{\theta})$ and $(F_0, F_1)$ along with their quantiles under the DRM (2.1) with unbiased estimating equations in (2.2);

(2) we propose a new testing procedure on the validity of (2.2) under the DRM (2.1), which leads to a practical validation method on the usefulness of the auxiliary information.

As we discussed in Section 1, the semiparametric DRM in (2.1) provides a flexible and useful platform for combining information from multiple sources. It also enables us to utilize information from both $F_0$ and $F_1$ to improve inferences on the unknown model parameters and the summary population quantities of interest. The estimating equations in (2.2) play two important roles. First, they can be used to define many important summary population quantities such as the ratio of the two population means, the centered and uncentered moments, the generalized entropy class of inequality measures, the CDFs, and the quantiles of each population. See Example 2.1 below and Section 2.6.1 for more examples. Second, they provide a unified platform for the use of auxiliary information. With many data sources being increasingly available, it becomes more feasible to access auxiliary information, and using such information to enhance statistical inference is an important and active research topic in many fields. Calibration estimators, which are widely used in survey sampling, missing data problems and causal inference, rely heavily on the use of auxiliary information; see Wu and Thompson (2020) and the references therein. Many economics problems can be addressed using similar methodology. For instance, knowledge of the moments of the marginal distributions of economic variables from census reports can be used in combination with microdata to improve the parameter estimates of microeconomic models (Imbens and Lancaster, 1994). Examples 2.2 and 2.3 below illustrate the use of auxiliary information through estimating equations in the form of (2.2).

**Example 2.1.** *(The mean ratio of two populations) The ratio of the means of two positive skewed distributions is often of interest in biomedical research (Zhou et al., 1997; Wu et al., 2002). Let $\mu_0$ and $\mu_1$ be the means with respect to $F_0$ and $F_1$, respectively. Further, let $\delta = \mu_1/\mu_0$ denote the mean ratio of the two populations. For inference on $\delta$, a common assumption is that both distributions are lognormal (Zhou et al., 1997; Wu et al., 2002). To* *alleviate the risk of parametric assumptions, we could use the DRM in (2.1) with* $\boldsymbol{q}(x) = \log x$ *or* $\boldsymbol{q}(x) = (\log x, \log^2 x)^\top$ *depending on whether or not the variances with respect to the log-scale are the same. Then, under the DRM* (2.1)*, the mean ratio* $\delta$ *can be defined through the following estimating equation:*

$$g(x; \boldsymbol{\psi}, \boldsymbol{\theta}) = \delta x - x \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\},$$

*with* $\boldsymbol{\psi} = \delta$*. When additional information is available, we may add more estimating equations to improve the estimation efficiency; see Section 2.4.1 for further detail.*

**Example 2.2.** *(Retrospective case-control studies with auxiliary information) Consider a retrospective case-control study with* $D = 1$ *or 0 representing diseased or disease-free status, and* $X$ *representing the collection of risk factors. Note that the two samples are collected retrospectively, given the diseased status. Let* $F_0$ *and* $F_1$ *denote the CDF of* $X$ *given* $D = 0$ *and* $D = 1$*, respectively. Assume that the relationship between* $D$ *and* $X$ *can be modeled by the logistic regression*

$$P(D = 1|x) = \frac{\exp\{\alpha^* + \boldsymbol{\beta}^\top \boldsymbol{q}(x)\}}{1 + \exp\{\alpha^* + \boldsymbol{\beta}^\top \boldsymbol{q}(x)\}},$$

*where* $\alpha^* = \alpha + \log\{P(D = 1)/P(D = 0)\}$*. Then, using the equivalence between the DRM and the logistic regression discussed in Section 1.1,* $F_0$ *and* $F_1$ *satisfy the DRM* (2.1)*.*

*Qin et al. (2015) used covariate-specific disease prevalence information to improve the power of case-control studies. Specifically, let* $X = (Y, Z)^\top$ *with* $Y$ *and* $Z$ *being two risk factors. Assume that we know the disease prevalence at various levels of* $Y$*:* $\phi(a_{l-1}, a_l) = P(D = 1|a_{l-1} < Y \leq a_l)$ *for* $l = 1, \ldots, k$*. Let* $\pi = P(D = 1)$ *be the overall disease prevalence. Using Bayes' formula, the information in the* $\phi(a_{l-1}, a_l)$*'s can be summarized as* $E_0\{\boldsymbol{g}(X; \boldsymbol{\psi}, \boldsymbol{\theta})\} = \boldsymbol{0}$*, where* $\boldsymbol{\psi} = \pi$ *and the* $l$*th component of* $\boldsymbol{g}(x; \boldsymbol{\psi}, \boldsymbol{\theta})$ *is*

$$g_l(x; \boldsymbol{\psi}, \boldsymbol{\theta}) = I(a_{l-1} < x \leq a_l) \left[ \frac{\pi}{1 - \pi} \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} - \frac{\phi(a_{l-1}, a_l)}{1 - \phi(a_{l-1}, a_l)} \right]. \tag{2.3}$$

*Chatterjee et al. (2016) improved the internal study by using summary-level information from an external study. Suppose* $X = (Y^\top, Z^\top)^\top$*, where* $Y$ *is available for both the internal and external studies, while* $Z$ *is available for only the internal study. Assume that the external study provides the true coefficients* $(\alpha_Y^*, \boldsymbol{\beta}_Y^*)$ *for the following logistic regression* *model, which may not be the true model:*

$$h(Y;\alpha_Y,\boldsymbol{\beta}_Y) = P(D=1|Y) = \frac{\exp(\alpha+\boldsymbol{\beta}_Y^\top Y)}{1+\exp(\alpha+\boldsymbol{\beta}_Y^\top Y)}.$$

*This assumption is reasonable when the total sample size $n = n_0 + n_1$ satisfies $n/n_E \to 0$, where $n_E$ is the total sample size in the external study. Further, assume that the joint distribution of $(D, X)$ is the same for both the internal and external studies. Let $h(y) = h(y;\alpha_Y^*,\boldsymbol{\beta}_Y^*)$. In Section 2.6.2, we argue that if the external study is a prospective case-control study, then $E_0\{\boldsymbol{g}(X;\boldsymbol{\psi},\boldsymbol{\theta})\} = \mathbf{0}$, where*

$$\boldsymbol{g}(x;\boldsymbol{\psi},\boldsymbol{\theta}) = [-(1-\pi)h(y) + \pi\exp\{\boldsymbol{\theta}^\top\boldsymbol{Q}(x)\}\{1-h(y)\}](1,y^\top)^\top \tag{2.4}$$

*with $\boldsymbol{\psi} = \pi$; if the external study is a retrospective case-control study, then $E_0\{\boldsymbol{g}(X;\boldsymbol{\theta})\} = \mathbf{0}$, where*

$$\boldsymbol{g}(x;\boldsymbol{\theta}) = [-(1-\pi_E)h(y) + \pi_E\exp\{\boldsymbol{\theta}^\top\boldsymbol{Q}(x)\}\{1-h(y)\}](1,y^\top)^\top \tag{2.5}$$

*with $\pi_E$ being the proportion of diseased individuals in the external study.*

**Example 2.3.** *(A two-sample problem with common mean) Tsao and Wu (2006) considered two populations with a common mean. This type of problem occurs when two "instruments" are used to collect data on a common response variable, and these two instruments are believed to have no systematic biases but differ in precision. The observations from the two instruments then form two samples with a common population mean. In the literature, there has been much interest in using the pooled sample to improve inferences. A common assumption is that the two samples follow normal distributions with a common mean but different variances (Tsao and Wu, 2006). To gain robustness with respect to the parametric assumption, we may use the DRM (2.1) with $\boldsymbol{q}(x) = (x, x^2)^\top$. Under this model, the common-mean assumption can be incorporated via the estimating equation:*

$$E_0\{X\exp\{\boldsymbol{\theta}^\top\boldsymbol{Q}(X)\} - X\} = 0. \tag{2.6}$$

### 2.1.2 Literature review

As we discussed in Section 1.3, the DRM has been studied extensively in the literature due to its flexibility and efficiency. In the following, we discuss several references in more details, which are related to our work in this chapter. Li et al. (2018) studied the MELE and compared two types of CIs for a parameter defined as $\boldsymbol{\psi} = \int u(x;\boldsymbol{\theta})dF_0(x)$, where

$u(\cdot;\cdot)$ is a one-dimensional function. Because of the specific form of $\boldsymbol{\psi}$, their results do not apply to the mean ratio discussed in Example 2.1. Zhang et al. (2020) investigated the ELR statistic for quantiles under the DRM and showed that the ELR-based confidence region of the quantiles is preferable to the Wald-type confidence region. Both Li et al. (2018) and Zhang et al. (2020) did not consider auxiliary information. In summary, the existing literature on DRMs focuses on cases where there is no auxiliary information, and furthermore, there is no general theory available to handle parameters defined through the estimating equations in (2.2).

Using the connection of the DRM to the logistic regression model, Qin et al. (2015) studied the MELE of $\boldsymbol{\theta}$ and the ELR statistic for testing a parameter in $\boldsymbol{\theta}$ under the DRM (2.1) with the unbiased estimating equations in (2.3). Chatterjee et al. (2016) proposed constrained maximum likelihood estimation for the unknown parameters in the internal study using summary-level information from an external study. In Section 2.6.2, we argue that their results are applicable to the MELE of $\boldsymbol{\theta}$ under the DRM (2.1) with the unbiased estimating equations in (2.4) but not to the MELE of $\boldsymbol{\theta}$ under the DRM (2.1) with the unbiased estimating equations in (2.5). Furthermore, they did not consider the ELR statistic for the unknown parameters. Qin et al. (2015) and Chatterjee et al. (2016) focused on how to use auxiliary information to improve inference on the unknown parameters, and they did not check the validity of that information or explore inferences on the CDFs $(F_0, F_1)$ and their quantiles.

### 2.1.3 Our contributions

With two-sample observations from the DRM (2.1), we use the empirical likelihood of Owen (1988, 2001) to incorporate the unbiased estimating equations in (2.2). We show that the MELE of $(\boldsymbol{\psi}, \boldsymbol{\theta})$ is asymptotically normal, and its asymptotic variance will not decrease when an estimating equation in (2.2) is dropped. We also develop an ELR statistic for testing a general hypothesis about $(\boldsymbol{\psi}, \boldsymbol{\theta})$, and show that it has a $\chi^2$ limiting distribution under the null hypothesis. The result can be used to construct the ELR-based confidence region for $(\boldsymbol{\psi}, \boldsymbol{\theta})$. Similar results are obtained for inferences on $(F_0, F_1)$ and their quantiles. Finally, we construct an ELR statistic with the $\chi^2$ null limiting distribution to test the validity of some or all of the estimating equations in (2.2).

We make the following observations:

(1) Our results on the two-sample DRMs contain more advanced development than those in Qin and Lawless (1994) for the one-sample case.

(2) Our inferential framework and theoretical results are very general. The results in Qin et al. (2015) and Chatterjee et al. (2016) for case-control studies are special cases of our theory for an appropriate choice of $\boldsymbol{g}(x;\boldsymbol{\psi},\boldsymbol{\theta})$ in (2.2). Our results are also applicable to cases that are not covered by these two earlier studies, e.g., Example 2.2 with the estimating equations in (2.5) and Example 2.3.

(3) Our proposed ELR statistic, to the best of our knowledge, is the first formal procedure to test the validity of auxiliary information under the DRM or for case-control studies.

(4) Our proposed inference procedures for $(F_0, F_1)$ and their quantiles in the presence of auxiliary information are new to the literature.

The rest of this chapter is organized as follows. In Section 2.2, we develop the empirical likelihood inferential procedures and study the asymptotic properties of the MELE of $(\boldsymbol{\psi},\boldsymbol{\theta})$. We also investigate the ELR statistics for $(\boldsymbol{\psi},\boldsymbol{\theta})$ and for testing the validity of the estimating equations in (2.2). In Section 2.3, we discuss inference procedures for $(F_0, F_1)$ and their quantiles. Simulation results are reported in Section 2.4, and two real-data examples are presented in Section 2.5. For convenience of presentation, more examples for summary quantities, details on extracting the summary-level information from the external case-control study, proofs, and additional simulation results are given in Section 2.6.

## 2.2 Empirical Likelihood and Inference on $(\boldsymbol{\psi},\boldsymbol{\theta})$

In this section, we first use the similar strategy in Section 1.2.2 to develop the empirical likelihood formulation under the DRM (2.1) with the unbiased estimating equations in (2.2). With two samples $\{X_{01},\ldots,X_{0n_0}\}$ and $\{X_{11},\ldots,X_{1n_1}\}$ from $F_0$ and $F_1$, respectively, the full likelihood is

$$\prod_{i=0}^{1}\prod_{j=1}^{n_i} dF_i(X_{ij}).$$

Under the one-sample empirical likelihood formulation of Owen (2001), the baseline distribution function $F_0(x)$ would have been estimated by $F_0^*(x)=\sum_{j=1}^{n_0}p_jI(X_{0j}\le x)$, where $p_j = dF_0(X_{0j})$ for $j=1,\ldots,n_0$. Under the two-sample DRM (2.1), we use the combined sample to estimate the baseline function $F_0(x)$ as

$$F_0^*(x)=\sum_{i=0}^{1}\sum_{j=1}^{n_i}p_{ij}I(X_{ij}\le x), \tag{2.7}$$

where $p_{ij} = dF_0(X_{ij})$ for $i = 0, 1$ and $j = 1, \ldots, n_i$. Note that the size of the combined sample is $n = n_0 + n_1$. With (2.7) and under the DRM (2.1), the empirical likelihood function is given by

$$\mathcal{L}_n = \left\{ \prod_{i=0}^{1} \prod_{j=1}^{n_i} p_{ij} \right\} \left[ \prod_{j=1}^{n_1} \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{1j})\} \right] \tag{2.8}$$

The feasible $p_{ij}$'s satisfy two sets of constraints given by

$$\mathcal{C}_1 = \left\{ (F_0, \boldsymbol{\theta}) : p_{ij} > 0, \ \sum_{i=0}^{1} \sum_{j=1}^{n_i} p_{ij} = 1, \ \sum_{i=0}^{1} \sum_{j=1}^{n_i} p_{ij} \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} = 1 \right\} \tag{2.9}$$

and

$$\mathcal{C}_2 = \left\{ (F_0, \boldsymbol{\psi}, \boldsymbol{\theta}) : \sum_{i=0}^{1} \sum_{j=1}^{n_i} p_{ij} \boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta}) = \mathbf{0} \right\}, \tag{2.10}$$

where the set of constraints $\mathcal{C}_1$ ensures that estimates of $F_0$ and $F_1$ are CDFs and the set of constraints $\mathcal{C}_2$ is induced by the estimating equations in (2.2).

Using the Lagrange multiplier method and for the given $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$, it can be shown that the maximizer of the empirical likelihood function is given by

$$p_{ij} = \frac{1}{n} \frac{1}{1 + \lambda \left[ \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1 \right] + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})},$$

where the Lagrange multipliers $\lambda$ and $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_r)^\top$ are the solutions to the following set of $r + 1$ equations:

$$\sum_{i=0}^{1} \sum_{j=1}^{n_i} \frac{\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1}{1 + \lambda \left[ \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1 \right] + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})} = 0, \tag{2.11}$$

$$\sum_{i=0}^{1} \sum_{j=1}^{n_i} \frac{\boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})}{1 + \lambda \left[ \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1 \right] + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})} = \mathbf{0}. \tag{2.12}$$

The profile empirical log-likelihood of $(\boldsymbol{\psi}, \boldsymbol{\theta})$ is given by

$$\ell_n(\boldsymbol{\psi}, \boldsymbol{\theta}) = -\sum_{i=0}^{1} \sum_{j=1}^{n_i} \log \left\{ 1 + \lambda \left[ \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1 \right] + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta}) \right\} + \sum_{j=1}^{n_1} \boldsymbol{\theta}^\top \boldsymbol{Q}(X_{1j}).$$

The MELEs of $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ are then defined as $(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}) = \arg\max_{\boldsymbol{\psi},\boldsymbol{\theta}} \ell_n(\boldsymbol{\psi}, \boldsymbol{\theta})$.

We now establish the asymptotic distribution of $(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})$. Let $\boldsymbol{\eta} = (\boldsymbol{\psi}^\top, \boldsymbol{\theta}^\top)^\top$ be the vector of parameters and $\boldsymbol{u} = (\lambda, \boldsymbol{\nu}^\top)^\top$ be the vector of Lagrange multipliers. We use $\boldsymbol{\psi}^*$ and $\boldsymbol{\theta}^*$ to denote the true values of $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$. We refer $\boldsymbol{\eta}^* = (\boldsymbol{\psi}^{*\top}, \boldsymbol{\theta}^{*\top})^\top$ to the true value of $\boldsymbol{\eta}$. We further define $\lambda^* = n_1/n$, and

$$
\begin{aligned}
&\omega(x;\boldsymbol{\theta}) = \exp\left\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\right\}, \quad \omega(x) = \omega(x;\boldsymbol{\theta}^*), \qquad h(x) = 1 + \lambda^*\left\{\omega(x) - 1\right\},\\
&h_1(x) = \frac{\lambda^*\omega(x)}{h(x)}, \quad \boldsymbol{G}(x;\boldsymbol{\eta}) = (\omega(x;\boldsymbol{\theta}) - 1, \boldsymbol{g}(x;\boldsymbol{\theta},\boldsymbol{\beta})^\top)^\top, \quad \boldsymbol{G}(x) = \boldsymbol{G}(x;\boldsymbol{\eta}^*),\\
&\boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}} = (1-\lambda^*)E_0\left\{h_1(X)\boldsymbol{Q}(x)\boldsymbol{Q}(x)^\top\right\},\\
&\boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{u}} = \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{\theta}}^\top = E_0\left\{\frac{\partial \boldsymbol{G}(X;\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta}}\right\}^\top - E_0\left\{h_1(X)\boldsymbol{Q}(x)\boldsymbol{G}(X)^\top\right\},\\
&\boldsymbol{A}_{\boldsymbol{\psi}\boldsymbol{u}} = \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{\psi}}^\top = E_0\left\{\frac{\partial \boldsymbol{G}(X;\boldsymbol{\eta}^*)}{\partial \boldsymbol{\psi}}\right\}^\top, \quad \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}} = E_0\left\{\frac{\boldsymbol{G}(X)\boldsymbol{G}(X)^\top}{h(X)}\right\}.
\end{aligned}
$$

Noting that $\omega(\cdot)$, $h(\cdot)$, $h_1(\cdot)$ and $G(\cdot)$ depend on $\boldsymbol{\psi}^*$ and/or $\boldsymbol{\theta}^*$, we drop these redundant parameters for notational simplicity.

**Theorem 2.1.** *Assume that the regularity conditions in Section 2.6.3 are satisfied. As the total sample size $n = n_0 + n_1$ goes to infinity, we have*

$$
n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \to N\left(\boldsymbol{0}, \boldsymbol{J}^{-1}\right)
$$

*in distribution, where*

$$
\boldsymbol{J} = \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top, \quad \boldsymbol{U} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{A}_{\boldsymbol{\psi}\boldsymbol{u}} \\ \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{u}} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{V} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}} \end{pmatrix}.
$$

In the absence of the constraints $\mathcal{C}_2$ in (2.10), we can maximize the empirical likelihood function in (2.8) with respect only to the CDF constraints $\mathcal{C}_1$ in (2.9) to obtain the MELE $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. As we discussed in Section 1.2.2, the MELE $\tilde{\boldsymbol{\theta}}$ equivalently maximizes the following dual likelihood:

$$
\ell_{nd}(\boldsymbol{\theta}) = -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \log\left\{1 + \lambda^*\left[\exp\left\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\right\} - 1\right]\right\} + \sum_{j=1}^{n_1}\left\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{1j})\right\}. \tag{2.13}
$$

That is, $\tilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell_{nd}(\boldsymbol{\theta})$.

**Corollary 2.1.** *Under the conditions of Theorem 2.1,*

*(a) if $r = p$, the asymptotic variance of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ is the same as that of $n^{1/2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$;*

*(b) if $r > p$, the asymptotic variance matrix of $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*)$ cannot decrease if one estimating equation in (2.2) is dropped.*

We provide some further comments on the results presented in Corollary 2.1. First, when the dimensions of the parameters $\boldsymbol{\psi}$ and the estimating equations are equal, we can solve

$$\int \boldsymbol{g}(X; \boldsymbol{\psi}, \tilde{\boldsymbol{\theta}}) d\tilde{F}_0(x) = \mathbf{0}$$

to get the estimator $\tilde{\boldsymbol{\psi}}$ of $\boldsymbol{\psi}$, where $\tilde{F}_0(x)$ is the MELE of $F_0$ without the constraints $\mathcal{C}_2$ in (2.10), and is defined in (1.17). Because of the result in Corollary 2.1(a), the estimators $\tilde{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\psi}}$ share the same asymptotic property. Second, Corollary 2.1(b) indicates that additional auxiliary information leads to more efficient estimation of $\boldsymbol{\eta}$.

The proposed semiparametric method provides a way to find the point estimator of the unknown parameters, which has the asymptotic normality analogue to the parametric estimator. The semiparametric framework also creates a natural platform for hypothesis tests using the ELR statistic. We consider a general null hypothesis

$$H_0 : \boldsymbol{H}(\boldsymbol{\eta}) = \mathbf{0},$$

where the function $\boldsymbol{H}(\cdot)$ is $q \times 1$ with $q \leq p + d + 1$, and the derivative of this function is of rank $q$. This null hypothesis forms a third set of constraints

$$\mathcal{C}_3 = \left\{ \boldsymbol{\eta} = (\boldsymbol{\psi}^\top, \boldsymbol{\theta}^\top)^\top : \boldsymbol{H}(\boldsymbol{\eta}) = \mathbf{0} \right\}.$$

The ELR statistic for testing $H_0$ is then defined as

$$R_n = 2 \left\{ \sup_{\boldsymbol{\psi}, \boldsymbol{\theta}} \ell_n(\boldsymbol{\psi}, \boldsymbol{\theta}) - \sup_{\boldsymbol{\eta} \in \mathcal{C}_3} \ell_n(\boldsymbol{\psi}, \boldsymbol{\theta}) \right\}.$$

The next theorem establishes the asymptotic distribution of the ELR statistic $R_n$ under the null hypothesis $H_0$.

**Theorem 2.2.** *Assume that the conditions of Theorem 2.1 hold. Under $H_0$, as $n \to \infty$, the ELR statistic $R_n \to \chi^2_q$ in distribution.*

The result of Theorem 2.2 is very general due to the general form of the function $\boldsymbol{H}(\cdot)$. First, it is applicable to testing problems that focus on some of the parameters in $\boldsymbol{\eta}$. For example, if we wish to test $H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0$, we can choose $\boldsymbol{H}(\boldsymbol{\eta}) = \boldsymbol{\psi} - \boldsymbol{\psi}_0$. Let $R_n^*(\boldsymbol{\psi})$ be the ELR function of $\boldsymbol{\psi}$. That is,

$$R_n^*(\boldsymbol{\psi}) = 2\left\{\sup_{\boldsymbol{\psi},\boldsymbol{\theta}} \ell_n(\boldsymbol{\psi},\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}} \ell_n(\boldsymbol{\psi},\boldsymbol{\theta})\right\}.$$

Then $R_n^*(\boldsymbol{\psi}_0)$ has a chi-squared null limiting distribution with $p$ degrees of freedom. Second, the result can be used to construct confidence regions for some of the parameters in $\boldsymbol{\eta}$. For example, we can construct an ELR-based confidence region for the parameter $\boldsymbol{\psi}$ at the nominal level $1 - a$ as

$$\{\boldsymbol{\psi} : R_n^*(\boldsymbol{\psi}) \le \chi^2_{q,1-a}\}, \tag{2.14}$$

where $\chi^2_{q,1-a}$ is the $100(1-a)$th quantile of the $\chi^2_q$ distribution.

The use of valid auxiliary information leads to improved inference on $\boldsymbol{\eta}$. However, if the information is not properly specified in terms of unbiased estimating functions, the resulting estimator of $\boldsymbol{\eta}$ may be biased (Qin et al., 2015). Our last major theoretical result is to construct an ELR statistic for testing the validity and usefulness of the auxiliary information. Let

$$W_n = 2\left\{\sup_{(\boldsymbol{\eta},F_0)\in\mathcal{C}_1} \log\mathcal{L}_n - \sup_{(\boldsymbol{\eta},F_0)\in\mathcal{C}_1\cap\mathcal{C}_2} \log\mathcal{L}_n\right\} = 2\left\{\ell_{nd}(\tilde{\boldsymbol{\theta}}) - \ell_n(\hat{\boldsymbol{\psi}},\hat{\boldsymbol{\theta}})\right\}. \tag{2.15}$$

**Theorem 2.3.** *Under the conditions of Theorem 2.1 and as $n \to \infty$, we have $W_n \to \chi^2_{r-p}$ in distribution if (2.2) is correctly specified.*

We can also test the validity of some but not all of the estimating equations in (2.2). To do so, we partition the estimating equations in (2.2) into two parts:

$$\boldsymbol{g}(x;\boldsymbol{\psi},\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{g}_1(x;\boldsymbol{\psi},\boldsymbol{\theta}) \\ \boldsymbol{g}_2(x;\boldsymbol{\psi},\boldsymbol{\theta}) \end{pmatrix},$$

where $\boldsymbol{g}_1(\cdot)$ and $\boldsymbol{g}_2(\cdot)$ are of dimension $r - m$ and $m$ with $r - m \ge p$. We are interested in testing $H_0 : E_0\{\boldsymbol{g}_2(X;\boldsymbol{\psi},\boldsymbol{\theta})\} = \boldsymbol{0}$. Let $\ell_{n1}(\boldsymbol{\psi},\boldsymbol{\theta})$ be the profile empirical log-likelihood of $(\boldsymbol{\psi},\boldsymbol{\theta})$ that uses the auxiliary information only through $E_0\{\boldsymbol{g}_1(x;\boldsymbol{\psi},\boldsymbol{\theta})\} = \boldsymbol{0}$. That is,

$$\ell_{n1}(\boldsymbol{\psi},\boldsymbol{\theta}) = -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \log\left\{1 + \lambda\left[\exp\{\boldsymbol{\theta}^\top\boldsymbol{Q}(X_{ij})\} - 1\right] + \boldsymbol{\nu}_1^\top\boldsymbol{g}_1(X_{ij};\boldsymbol{\psi},\boldsymbol{\theta})\right\} + \sum_{j=1}^{n_1}\boldsymbol{\theta}^\top\boldsymbol{Q}(X_{1j}),$$

where $\lambda$ and $\boldsymbol{\nu}_1$ are the solution to

$$\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\exp\{\boldsymbol{\theta}^\top\boldsymbol{Q}(X_{ij})\}-1}{1+\lambda\left[\exp\{\boldsymbol{\theta}^\top\boldsymbol{Q}(X_{ij})\}-1\right]+\boldsymbol{\nu}_1^\top\boldsymbol{g}_1(X_{ij};\boldsymbol{\psi},\boldsymbol{\theta})}=0,$$

$$\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\boldsymbol{g}(X_{ij};\boldsymbol{\psi},\boldsymbol{\theta})}{1+\lambda\left[\exp\{\boldsymbol{\theta}^\top\boldsymbol{Q}(X_{ij})\}-1\right]+\boldsymbol{\nu}_1^\top\boldsymbol{g}_1(X_{ij};\boldsymbol{\psi},\boldsymbol{\theta})}=\mathbf{0}.$$

Then the ELR statistic for testing $H_0: E_0\{\boldsymbol{g}_2(X;\boldsymbol{\psi},\boldsymbol{\theta})\}=\mathbf{0}$ can be constructed similar to (2.15) as

$$W_n^*=2\left\{\sup_{\boldsymbol{\psi},\boldsymbol{\theta}}\ell_{n1}(\boldsymbol{\psi},\boldsymbol{\theta})-\sup_{\boldsymbol{\psi},\boldsymbol{\theta}}\ell_n(\boldsymbol{\psi},\boldsymbol{\theta})\right\}.$$

**Corollary 2.2.** *Under the conditions of Theorem 2.1 and as $n\to\infty$, we have $W_n^*\to\chi^2_m$ if $E_0\{\boldsymbol{g}_2(X;\boldsymbol{\psi},\boldsymbol{\theta})\}=\mathbf{0}$ is true.*

## 2.3 Inferences on CDFs and Quantiles

In this section, we discuss inferences on the CDFs $F_0$ and $F_1$ and their quantiles. For convenience of presentation, we assume that the dimension of $X_{ij}$ is one.

We first construct point estimators of $F_0$ and $F_1$. Let $\hat\lambda$ and $\hat{\boldsymbol{\nu}}$ be the solutions to (2.11) and (2.12) with $(\boldsymbol{\psi},\boldsymbol{\theta})$ replaced by $(\hat{\boldsymbol{\psi}},\hat{\boldsymbol{\theta}})$. The MELEs of $p_{ij}$ are then given as

$$\hat p_{ij}=\frac{1}{n}\frac{1}{1+\hat\lambda\left[\exp\{\hat{\boldsymbol{\theta}}^\top\boldsymbol{Q}(X_{ij})\}-1\right]+\hat{\boldsymbol{\nu}}^\top\boldsymbol{g}(X_{ij};\hat{\boldsymbol{\psi}},\hat{\boldsymbol{\theta}})}.$$

The MELEs of $F_0$ and $F_1$ are then defined as

$$\hat F_0(x)=\sum_{i=0}^{1}\sum_{j=1}^{n_i}\hat p_{ij}I(X_{ij}\le x)\quad\text{and}\quad\hat F_1(x)=\sum_{i=0}^{1}\sum_{j=1}^{n_i}\hat p_{ij}\exp\{\hat{\boldsymbol{\theta}}^\top\boldsymbol{Q}(X_{ij})\}I(X_{ij}\le x).$$

We now present results on the asymptotic properties of the MELEs $\hat F_0(x)$ and $\hat F_1(x)$ of the two population CDFs $F_0(x)$ and $F_1(x)$. Let

$$\boldsymbol{W}=\boldsymbol{V}^{-1}\boldsymbol{U}^\top\boldsymbol{J}^{-1}\boldsymbol{U}\boldsymbol{V}^{-1}-\begin{pmatrix}\mathbf{0}&\mathbf{0}\\\mathbf{0}&\boldsymbol{A}_{\boldsymbol{uu}}^{-1}\end{pmatrix},\ \ \boldsymbol{B}_0^*(x)=\begin{pmatrix}\boldsymbol{B}_{0\boldsymbol{\theta}}(x)\\\boldsymbol{B}_{0\boldsymbol{u}}(x)\end{pmatrix},\ \ \boldsymbol{B}_1^*(x)=\begin{pmatrix}\boldsymbol{B}_{1\boldsymbol{\theta}}(x)\\\boldsymbol{B}_{1\boldsymbol{u}}(x)\end{pmatrix},$$

where

$$\boldsymbol{B}_{0\boldsymbol{\theta}}(x) = E_0\left\{h_1(X)\boldsymbol{Q}(X)I(X \le x)\right\}, \qquad \boldsymbol{B}_{0\boldsymbol{u}}(x) = E_0\left\{\frac{\boldsymbol{G}(X)}{h(X)}I(X \le x)\right\},$$

$$\boldsymbol{B}_{1\boldsymbol{\theta}}(x) = \frac{\lambda^*-1}{\lambda^*}E_0\left\{h_1(X)\boldsymbol{Q}(X)I(X \le x)\right\}, \quad \boldsymbol{B}_{1\boldsymbol{u}}(x) = E_0\left\{\frac{\omega(X)\boldsymbol{G}(X)}{h(X)}I(X \le x)\right\}.$$

Recall that $\tilde{F}_0(x)$ and $\tilde{F}_1(x)$ are the MELEs of $F_0$ and $F_1$ under the DRM when there is no auxiliary information, and are defined (1.17). We refer to Qin and Zhang (1997) for the asymptotic properties of $\tilde{F}_0(x)$ and $\tilde{F}_1(x)$. Denote $x \wedge y = \min(x, y)$.

**Theorem 2.4.** *Assume that the conditions of Theorem 2.1 are satisfied.*

*(a) For any $l, s \in \{0,1\}$ and real numbers $x$ and $y$ in the support of $F_0$, as $n \to \infty$,*

$$\sqrt{n}\begin{pmatrix} \hat{F}_l(x) - F_l(x) \\ \hat{F}_s(y) - F_s(y) \end{pmatrix} \to N\Big(\mathbf{0}, \boldsymbol{\Sigma}_{ls}(x,y)\Big),$$

*where*

$$\boldsymbol{\Sigma}_{ls}(x,y) = \begin{pmatrix} \sigma_{ll}(x,x) & \sigma_{ls}(x,y) \\ \sigma_{sl}(y,x) & \sigma_{ss}(y,y) \end{pmatrix}$$

*with*

$$\sigma_{ij}(x,y) = E_0\left\{\frac{\omega^{i+j}(X)I(X \le x \wedge y)}{h(X)}\right\} - F_i(x)F_j(y) + \boldsymbol{B}_i^*(x)^\top \boldsymbol{W}\boldsymbol{B}_j^*(y)$$

*for any $i, j \in \{l, s\}$.*

*(b) If $r = p$, the asymptotic variance-covariance matrix $\boldsymbol{\Sigma}_{ls}(x,y)$ reduces to the same one of $\sqrt{n}\big(\tilde{F}_l(x) - F_l(x), \tilde{F}_s(x) - F_s(x)\big)^\top$.*

*(c) If $r > p$, the asymptotic variance matrix $\boldsymbol{\Sigma}_{ls}(x,y)$ cannot decrease if one estimating equation in (2.2) is dropped.*

Theorem 2.4 indicates that the MELEs $\hat{F}_0(x)$ and $\hat{F}_1(x)$ have asymptotic properties similar to those of $\hat{\boldsymbol{\eta}}$. That is, they are asymptotically normally distributed; they are asymptotically equivalent to $\tilde{F}_0(x)$ and $\tilde{F}_1(x)$ when $r = p$; and they become more efficient when $r > p$.

In the second half of this section we discuss the estimation of the quantiles of $F_i(x)$ for $i = 0$ and 1. For any $\tau \in (0, 1)$, we define the $\tau$th-quantile of $F_i$ as $\xi_{i,\tau} = \inf\{x : F_i(x) \geq \tau\}$ and its MELE as

$$\hat{\xi}_{i,\tau} = \inf\{x : \hat{F}_i(x) \geq \tau\}. \tag{2.16}$$

Similarly, the estimator of $\xi_{i,\tau}$ based on $\tilde{F}_i(x)$ is defined as

$$\tilde{\xi}_{i,\tau} = \inf\{x : \tilde{F}_i(x) \geq \tau\}. \tag{2.17}$$

See Zhang (2000) and Chen and Liu (2013) for the asymptotic properties of $\tilde{\xi}_{i,\tau}$. We refer to $\hat{\xi}_{i,\tau}$ as the "DRM-EE" quantile estimators and $\tilde{\xi}_{i,\tau}$ as the "DRM" quantile estimators.

The Bahadur representation is a useful tool for studying the asymptotic properties of quantile estimators. In the following theorem, we show that the DRM-EE quantile estimators are Bahadur representable. Let $f_i(x)$ be the probability density function of $F_i(x)$ for $i = 0$ and 1.

**Theorem 2.5.** *Assume that the conditions of Theorem 2.1 are satisfied. Further, for $i = 0, 1$ and any $\tau \in (0, 1)$, assume that $f_i(x)$ is continuous and positive at $x = \xi_{i,\tau}$. Then $\hat{\xi}_{i,\tau}$ admits the Bahadur representation*

$$\hat{\xi}_{i,\tau} = \xi_{i,\tau} + \frac{\tau - \hat{F}_i(\xi_{i,\tau})}{f_i(\xi_{i,\tau})} + O_p\big(n^{-3/4}(\log n)^{1/2}\big).$$

The following theorem shows that the DRM-EE quantile estimators have asymptotic properties similar to those of the MELEs of $\boldsymbol{\eta}$, $F_0(x)$, and $F_1(x)$.

**Theorem 2.6.** *Assume that the conditions in Theorem 2.5 hold for $x = \xi_{l,\tau_l}$ and $x = \xi_{s,\tau_s}$.*

*(a) As $n \to \infty$,*

$$\sqrt{n}\begin{pmatrix} \hat{\xi}_{l,\tau_l} - \xi_{l,\tau_l} \\ \hat{\xi}_{s,\tau_s} - \xi_{s,\tau_s} \end{pmatrix} \to N(\mathbf{0}, \boldsymbol{\Omega}_{ls}),$$

*where*

$$\boldsymbol{\Omega}_{ls} = \begin{pmatrix} \sigma_{ll}(\xi_{l,\tau_l}, \xi_{s,\tau_s})/f_l^2(\xi_{l,\tau_l}) & \sigma_{ls}(\xi_{l,\tau_l}, \xi_{s,\tau_s})/f_l(\xi_{l,\tau_l})f_s(\xi_{s,\tau_s}) \\ \sigma_{sl}(\xi_{s,\tau_s}, x)/f_s(\xi_{s,\tau_s})f_l(\xi_{l,\tau_l}) & \sigma_{ss}(\xi_{s,\tau_s}, \xi_{s,\tau_s})/f_s^2(\xi_{s,\tau_s}) \end{pmatrix}.$$

*(b) If $r = p$, the asymptotic variance matrix $\boldsymbol{\Omega}_{ls}$ of the DRM-EE quantile estimators is the same as that for the DRM quantile estimators;*

*(c) if $r > p$, the asymptotic variance matrix $\Omega_{ls}$ of the DRM-EE quantile estimators cannot decrease if one estimating equation in (2.2) is dropped.*

Using the results of Theorems 2.4 and 2.6, we may construct confidence regions and/or test hypotheses on the CDFs at some fixed points and for quantiles through the Wald-type statistics. However, methods based on the Wald-type statistics require a consistent estimator of the corresponding asymptotic variance. It is more attractive to use the results in Corollary 2.2 to construct the ELR-based confidence region for the CDFs at some fixed points and for quantiles.

Suppose we are interested in constructing a $(1-a)$-level CI for a CDF at some fixed point $x_0$ for $i = 0$ or 1. Denote the parameter of interest as $\zeta = F_i(x_0)$. Let

$$g_1^*(x;\boldsymbol{\theta},\zeta) = \begin{cases} I(x \le x_0) - \zeta, & i = 0 \\ \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} I(x \le x_0) - \zeta, & i = 1 \end{cases}.$$

We further define $\ell_{n1}^*(\boldsymbol{\psi},\boldsymbol{\theta},\zeta)$ to be the profile empirical log-likelihood of $(\boldsymbol{\psi},\boldsymbol{\theta},\zeta)$ under the DRM (2.1) with the unbiased estimating equations in (2.2) and $E_0\{g_1^*(X;\boldsymbol{\theta},\zeta)\}$=0. Then the ELR function of $\zeta$ is defined as

$$R_{n1}(\zeta) = 2\{\ell_n(\hat{\boldsymbol{\psi}},\hat{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\psi},\boldsymbol{\theta}} \ell_{n1}^*(\boldsymbol{\psi},\boldsymbol{\theta},\zeta)\}.$$

We can similarly define the ELR function for a quantile $\xi$ at the quantile level $\tau$ for $i = 0$ or 1, i.e., $\xi = \xi_{i,\tau}$. Let

$$g_2^*(x;\boldsymbol{\theta},\xi) = \begin{cases} I(x \le \xi) - \tau, & i = 0 \\ \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} I(x \le \xi) - \tau, & i = 1 \end{cases}.$$

We further define $\ell_{n2}^*(\boldsymbol{\psi},\boldsymbol{\theta},\xi)$ to be the profile empirical log-likelihood of $(\boldsymbol{\psi},\boldsymbol{\theta},\xi)$ under the DRM (2.1) with the unbiased estimating equations in (2.2) and $E_0\{g_2^*(X;\boldsymbol{\theta},\xi)\} = 0$. Then the ELR function of $\xi$ is defined as

$$R_{n2}(\xi) = 2\{\ell_n(\hat{\boldsymbol{\psi}},\hat{\boldsymbol{\theta}}) - \sup_{\boldsymbol{\psi},\boldsymbol{\theta}} \ell_{n2}^*(\boldsymbol{\psi},\boldsymbol{\theta},\xi)\}.$$

Using Corollary 2.2, we have the following results for $R_{n1}(\zeta^*)$ and $R_{n2}(\xi^*)$, where $\zeta^*$ and $\xi^*$ are the true values of $\zeta$ and $\xi$.

**Corollary 2.3.** *Under the conditions of Theorem 2.1, as $n \to \infty$, both $R_{n1}(\zeta^*)$ and $R_{n2}(\xi^*)$ converge in distribution to $\chi_1^2$.*

Corollary 2.3 enables us to construct the ELR-based CI for $\zeta$ and $\xi$. For example, the ELR-based CI for $\xi$ with level $1-a$ can be constructed as $\{\xi : R_{n2}(\xi) \le \chi^2_{1,1-a}\}$.

## 2.4 Simulation Studies

We conducted simulation studies to investigate three aspects of the proposed semiparametric inference procedures:

(1) The performance of the inference procedures for $\boldsymbol{\psi}$;

(2) The power of the ELR test for the validity and usefulness of the auxiliary information;

(3) The performance of the inference procedures for the population quantiles.

We consider four combinations of sample sizes $(n_0, n_1)$: $(50, 50)$, $(50, 150)$, $(100, 100)$, and $(200, 200)$. For each simulation setting, the number of simulation runs is 2,000.

### 2.4.1 Simulation studies for inferences on $\boldsymbol{\psi}$

**Simulation setup**

We start by exploring the first aspect of the proposed semiparametric inference procedures. In the simulations, $F_0$ and $F_1$ are the CDFs of $LN(0, 1)$ and $LN(0.5, 1)$, respectively, where $LN(a, b)$ denotes the lognormal distribution with mean $a$ and variance $b$, both with respect to the log scale. It is easy to show that $F_0$ and $F_1$ satisfy the DRM in (2.1) with $\boldsymbol{Q}(x) = (1, \log x)^\top$. The parameter of interest is the mean ratio $\boldsymbol{\psi} = \delta = \mu_1/\mu_0$ which was discussed in Example 2.1.

To examine the usefulness of auxiliary information, we construct another variable $Z$ using the following model:

$$Z = 1 + 0.5X + \epsilon \quad \text{and} \quad \epsilon \sim N(0, 1). \tag{2.18}$$

That is, given $X_{ij}$, $Z_{ij}$ is generated from (2.18), for $i = 0, 1, j = 1, \cdots, n_i$. Hence, the two-sample data consist of $\boldsymbol{T}_{ij} = (X_{ij}, Z_{ij})^\top$ for $i = 0, 1, j = 1, \cdots, n_i$. We treat $\mu_{z0} = E(Z|D = 0)$, the population mean of covariate $Z$ for the first group (i.e., the $D = 0$ group), as the known auxiliary information. Let the CDFs of $\boldsymbol{T}$ given $D = 0$ and $D = 1$

be $\boldsymbol{F}_0$ and $\boldsymbol{F}_1$, respectively. It can be checked that $\boldsymbol{F}_0$ and $\boldsymbol{F}_1$ satisfy the DRM with $\boldsymbol{Q}(x, z) = (1, \log x)^\top$.

To explore the effect of misspecified estimating equations for the auxiliary information, we introduce a bias by using $\kappa\mu_{z0}$ instead of the true value $\mu_{z0}$ for $E(Z|D = 0)$. We consider $\kappa = 0.90, 0.95, 1.00, 1.05, 1.10$. Note that $\kappa = 1.00$ corresponds to correctly specified auxiliary information. We incorporate the biased/unbiased auxiliary information into our problem by setting $\boldsymbol{\psi} = \delta$ and $\boldsymbol{g}(\boldsymbol{t}; \boldsymbol{\psi}, \boldsymbol{\theta}) = \big(\delta x - x \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\}, z - \kappa\mu_{z0}\big)^\top$ in (2.2).

**Performance of point estimators**

We compare three point estimators:

(i) EMP: $\bar{\delta} = \bar{\mu}_1/\bar{\mu}_0$, where $\bar{\mu}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}$ for $i = 0$ and 1;

(ii) DRM: $\tilde{\delta} = \tilde{\mu}_1/\tilde{\mu}_0$, where $\tilde{\mu}_i = \int x d\tilde{F}_i(x)$ for $i = 0$ and 1;

(iii) DRM-EE: $\hat{\delta} = \hat{\mu}_1/\hat{\mu}_0$, where $\hat{\mu}_i = \int x d\hat{F}_i(x)$ for $i = 0$ and 1.

Note that the asymptotic properties of $\tilde{\delta}$ and $\hat{\delta}$ are covered in Theorem 2.1. The performance of each estimator is evaluated by the relative bias (RB) and the mean squared error (MSE). Here, the RB in percentage is defined as

$$\text{RB}(\%) = \frac{1}{B} \sum_{b=1}^{B} \frac{a^{(b)} - a^*}{a^*} \times 100,$$

where $a^*$ is the true value of the parameter of interest, $a^{(b)}$ is the estimate of $a^*$ from the $b$th simulation run, and $B = 2,000$ is the number of simulation runs. Simulation results on the three point estimators are presented in Table 2.1.

We first compare the results reported in the third to fifth columns, i.e., EMP, DRM, and DRM-EE with correctly specified auxiliary information (DRM-EE with $\kappa = 1$). We see that the EMP estimator has the largest RBs and MSEs in all cases. The estimator of DRM-EE with $\kappa = 1$ has the best performance, followed by the DRM estimator. This suggests that using correctly specified auxiliary information improves the estimation efficiency, which agrees with Corollary 2.1 in Section 2.2. We also note that as the sample size increases, all three estimators have improved performance and the gaps between the three estimators become less pronounced, especially between DRM and DRM-EE.

Table 2.1: RB (%) and MSE (×100) of three point estimators of the mean ratio.

| $(n_0, n_1)$ | | EMP | DRM | DRM-EE $\kappa = 1$ | $\kappa = 0.9$ | $\kappa = 0.95$ | $\kappa = 1.05$ | $\kappa = 1.1$ |
|---|---|---|---|---|---|---|---|---|
| $(50, 50)$ | RB | 3.37 | 1.46 | 1.15 | 12.73 | 6.83 | -4.24 | -9.32 |
| | MSE | 20.03 | 12.50 | 9.61 | 16.59 | 12.07 | 9.00 | 9.96 |
| $(50, 150)$ | RB | 3.70 | 1.75 | 0.89 | 16.61 | 8.50 | -6.11 | -12.41 |
| | MSE | 12.91 | 8.07 | 4.67 | 13.94 | 7.46 | 4.92 | 7.50 |
| $(100, 100)$ | RB | 1.86 | 1.21 | 0.89 | 12.32 | 6.48 | -4.35 | -9.20 |
| | MSE | 9.35 | 6.17 | 5.08 | 10.46 | 6.78 | 5.11 | 6.56 |
| $(200, 200)$ | RB | 0.90 | 0.46 | 0.53 | 11.87 | 6.06 | -4.62 | -9.27 |
| | MSE | 4.88 | 3.15 | 2.56 | 7.03 | 3.84 | 2.92 | 4.60 |

The sensitivity of the DRM-EE estimator with respect to misspecified auxiliary information can be observed from the last four columns of Table 2.1. The DRM-EE estimator for $\kappa \neq 1$ are clearly not as good as the estimator for $\kappa = 1$. The absolute value of the RB increases as $\kappa$ moves further away from 1.

**Performance of confidence intervals**

We compare four CIs for $\delta$:

(i) EMP-NA: Wald-type CI for $\delta$ based on the asymptotic normality of $\log \bar{\delta}$;

(ii) EMP-EL: Owen (2001)'s ELR-based CI for $\delta$;

(iii) DRM: the ELR-based CI for $\delta$ in (2.14) without auxiliary information;

(iv) DRM-EE: the ELR-based CI for $\delta$ in (2.14) with auxiliary information.

The performance of a CI is evaluated in terms of coverage probability (CP) and average length (AL). The simulation results for the four CIs at the 95% nominal level are shown in Table 2.2.

As we can see in the third to sixth columns, EMP-NA and EMP-EL are comparable but are clearly inferior to DRM and DRM-EE ($\kappa = 1$) in terms of CP and AL. The CPs of the CIs for DRM and DRM-EE with $\kappa = 1$ are close to the nominal level for all sample size combinations. This suggests that the limiting distributions provide accurate approximations to the finite-sample distributions of the ELR statistics. The ALs of the CIs for DRM-EE with $\kappa = 1$ are always shorter than other CIs, a strong evidence that using correctly specified auxiliary information improves the performance of a CI. On the other

Table 2.2: CP (%) and AL of four CIs for the mean ratio at 95% nominal level.

| $(n_0, n_1)$ | | EMP-NA | EMP-EL | DRM | DRM-EE | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\kappa = 1$ | $\kappa = 0.9$ | $\kappa = 0.95$ | $\kappa = 1.05$ | $\kappa = 1.1$ |
| $(50, 50)$ | CP | 92.6 | 91.6 | 94.5 | 94.2 | 90.7 | 93.9 | 92.1 | 88.1 |
| | AL | 1.65 | 1.65 | 1.41 | 1.23 | 1.38 | 1.30 | 1.16 | 1.10 |
| $(50, 150)$ | CP | 92.9 | 92.2 | 95.6 | 94.3 | 78.1 | 91.4 | 88.5 | 75.9 |
| | AL | 1.33 | 1.31 | 1.15 | 0.84 | 1.00 | 0.92 | 0.77 | 0.71 |
| $(100, 100)$ | CP | 94.9 | 93.9 | 95.3 | 94.3 | 85.6 | 92.5 | 92.0 | 85.1 |
| | AL | 1.18 | 1.20 | 1.00 | 0.88 | 0.98 | 0.93 | 0.84 | 0.80 |
| $(200, 200)$ | CP | 93.8 | 93.3 | 94.6 | 94.7 | 75.3 | 89.0 | 90.4 | 78.4 |
| | AL | 0.84 | 0.86 | 0.70 | 0.62 | 0.69 | 0.66 | 0.60 | 0.58 |

hand, misspecified auxiliary information results in inaccurate CIs. As $\kappa$ moves further away from 1, the CP of the ELR-based CI shifts away from the nominal value.

### Power of the validity test

In this section, we explore the second aspect of the proposed semiparametric inference procedures on the power of the ELR test for the validity of the auxiliary information. The null hypothesis for the ELR test is $H_0 : E_0(z - \kappa\mu_{z0}) = 0$. According to Theorem 2.3 and Corollary 2.2, the ELR statistic has a $\chi_1^2$ limiting distribution under the null hypothesis. We consider misspecified auxiliary information with $\kappa = 0.90, 0.95, 1.05, 1.10$ as the alternatives. Table 2.3 gives the simulated power ($\kappa \neq 1$) and type I error rate ($\kappa = 1$) of the ELR test at the 5% significance level.

Table 2.3: Power and type I error rate of the ELR test (%) at 5% significance level.

| $(n_0, n_1)$ | $\kappa = 0.9$ | $\kappa = 0.95$ | $\kappa = 1$ | $\kappa = 1.05$ | $\kappa = 1.1$ |
|---|---|---|---|---|---|
| $(50, 50)$ | 21.43 | 8.76 | 5.36 | 9.41 | 20.48 |
| $(50, 150)$ | 27.33 | 10.08 | 5.37 | 10.13 | 22.97 |
| $(100, 100)$ | 36.44 | 11.26 | 5.51 | 13.61 | 32.48 |
| $(200, 200)$ | 62.98 | 20.66 | 5.15 | 19.16 | 55.23 |

We observe from Table 2.3 that the type I error rates of the ELR tests are close to the 5% nominal level in all cases, which suggests that the limiting distribution for the ELR test works very well. As $\kappa$ deviates from 1 and the sample size increases, the power of the test increases, as expected.

### 2.4.2 Simulation studies for inferences on quantiles

#### Simulation setup

The third aspect of the proposed semiparametric inference procedures is inference on population quantiles with auxiliary information. In the simulations, we consider two distributional settings:

(1) $f_0 \sim N(18, 4)$ and $f_1 \sim N(18, 9)$;

(2) $f_0 \sim Gam(6, 1.5)$ and $f_1 \sim Gam(8, 1.125)$.

Here $N(a, b)$ denotes the normal distribution with mean $a$ and variance $b$ and $Gam(a, b)$ is the gamma distribution with shape parameter $a$ and scale parameter $b$. We are interested in estimating and constructing CIs for the quantiles of $F_0$ and $F_1$ at the levels $\tau = 0.10, 0.25, 0.5, 0.75, 0.90$.

#### Performance of quantile estimators

We compare four quantile estimators:

(i) EMP: the quantile estimator based on the empirical CDFs;

(ii) EL: the quantile estimator based on the MELEs of the CDFs in Tsao and Wu (2006), in which a common mean is assumed;

(iii) DRM: the DRM based quantile estimator in (2.17);

(iv) DRM-EE: our proposed quantile estimator in (2.16) with the common-mean assumption or the estimating equation (2.6) in Example 2.3.

The DRM and DRM-EE methods are calculated with the correctly specified $\boldsymbol{q}(x)$, where $\boldsymbol{q}(x) = (x, x^2)^\top$ for the normal distributional setting and $\boldsymbol{q}(x) = (x, \log x)^\top$ for the gamma distributional setting. The performance of an estimator is evaluated by the RB and MSE. The general patterns of the simulation results for the four methods are similar in the two settings. Hence, Table 2.4 presented here is only for the normal setting; the results under gamma distributions are included in Section 2.6.4.

Table 2.4 shows that the RBs are negligibly small for all methods under all scenarios. The EMP estimator has the largest MSEs. The DRM-EE quantile estimators have the smallest MSEs due to its use of additional information, and the results agree with Theorem 2.6. We also notice that the EL and DRM quantile estimators are comparable.

Table 2.4: RB (%) and MSE (×100) for quantile estimators (normal distributions).

| | | | $N(18,4)$ | | | | $N(18,9)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $(n_0,n_1)$ | $\tau$ | | EMP | EL | DRM | DRM-EE | EMP | EL | DRM | DRM-EE |
| (50, 50) | 0.10 | RB | -0.58 | 0.08 | 0.25 | 0.19 | -1.07 | -0.10 | 0.17 | -0.07 |
| | | MSE | 23.87 | 19.88 | 18.85 | 16.32 | 59.74 | 44.17 | 46.26 | 37.35 |
| | 0.25 | RB | 0.04 | 0.02 | 0.15 | 0.14 | 0.01 | -0.06 | -0.14 | -0.25 |
| | | MSE | 14.73 | 12.25 | 12.23 | 9.57 | 33.32 | 22.42 | 29.22 | 18.11 |
| | 0.50 | RB | -0.21 | 0.03 | 0.04 | 0.03 | -0.43 | 0.03 | 0.00 | 0.03 |
| | | MSE | 12.47 | 9.93 | 10.06 | 7.76 | 29.21 | 16.25 | 25.08 | 11.10 |
| | 0.75 | RB | -0.01 | -0.01 | -0.08 | -0.07 | -0.05 | 0.02 | 0.03 | 0.14 |
| | | MSE | 13.92 | 11.81 | 11.97 | 9.64 | 34.86 | 21.55 | 29.68 | 16.95 |
| | 0.90 | RB | -0.62 | -0.08 | -0.21 | -0.18 | -0.87 | 0.08 | -0.08 | 0.10 |
| | | MSE | 23.36 | 21.36 | 19.51 | 17.66 | 53.89 | 43.03 | 46.50 | 37.61 |
| (50, 150) | 0.10 | RB | -0.60 | 0.01 | 0.26 | 0.17 | -0.28 | 0.09 | 0.17 | 0.13 |
| | | MSE | 23.91 | 18.16 | 16.36 | 11.49 | 17.62 | 14.72 | 16.05 | 13.34 |
| | 0.25 | RB | 0.04 | 0.02 | 0.14 | 0.12 | 0.06 | 0.03 | -0.01 | -0.03 |
| | | MSE | 14.81 | 10.08 | 11.22 | 6.64 | 11.00 | 8.67 | 10.20 | 7.88 |
| | 0.50 | RB | -0.21 | 0.07 | 0.04 | 0.04 | -0.10 | 0.05 | 0.05 | 0.06 |
| | | MSE | 12.39 | 7.69 | 9.09 | 4.59 | 8.97 | 6.92 | 8.15 | 5.84 |
| | 0.75 | RB | -0.01 | 0.02 | -0.10 | -0.05 | -0.06 | -0.04 | -0.01 | 0.00 |
| | | MSE | 13.90 | 10.24 | 10.87 | 6.49 | 10.49 | 8.18 | 9.89 | 7.71 |
| | 0.90 | RB | -0.61 | -0.03 | -0.20 | -0.12 | -0.30 | -0.02 | -0.04 | -0.02 |
| | | MSE | 23.32 | 19.87 | 17.26 | 12.94 | 17.04 | 14.93 | 16.25 | 14.40 |
| (100, 100) | 0.10 | RB | -0.35 | 0.03 | 0.10 | 0.09 | -0.34 | 0.15 | 0.23 | 0.11 |
| | | MSE | 11.82 | 10.05 | 9.13 | 7.86 | 25.71 | 19.44 | 22.01 | 16.62 |
| | 0.25 | RB | -0.17 | 0.03 | 0.04 | 0.04 | -0.18 | 0.02 | 0.03 | -0.06 |
| | | MSE | 7.42 | 6.20 | 6.33 | 5.04 | 15.56 | 9.84 | 13.54 | 8.01 |
| | 0.50 | RB | -0.11 | 0.03 | 0.01 | 0.03 | -0.15 | 0.03 | 0.07 | 0.05 |
| | | MSE | 6.07 | 4.81 | 5.21 | 3.88 | 13.53 | 7.87 | 11.53 | 5.41 |
| | 0.75 | RB | -0.17 | 0.01 | -0.05 | -0.02 | -0.30 | -0.05 | 0.01 | 0.02 |
| | | MSE | 7.37 | 6.20 | 6.10 | 5.02 | 15.95 | 9.94 | 13.60 | 7.94 |
| | 0.90 | RB | -0.35 | -0.02 | -0.11 | -0.08 | -0.45 | -0.05 | -0.05 | -0.02 |
| | | MSE | 11.82 | 10.83 | 9.40 | 8.24 | 25.37 | 19.69 | 22.77 | 17.23 |
| (200, 200) | 0.10 | RB | -0.12 | 0.04 | 0.13 | 0.10 | -0.29 | -0.05 | 0.01 | -0.02 |
| | | MSE | 5.77 | 5.01 | 4.50 | 3.91 | 13.65 | 10.89 | 11.81 | 8.94 |
| | 0.25 | RB | -0.06 | 0.02 | 0.05 | 0.04 | -0.12 | 0.02 | -0.04 | -0.04 |
| | | MSE | 3.58 | 3.00 | 3.03 | 2.41 | 8.37 | 5.04 | 7.30 | 4.18 |
| | 0.50 | RB | -0.04 | 0.03 | 0.02 | 0.01 | -0.15 | -0.03 | -0.02 | 0.00 |
| | | MSE | 3.02 | 2.40 | 2.57 | 1.99 | 7.07 | 3.99 | 6.04 | 2.80 |
| | 0.75 | RB | -0.10 | -0.03 | -0.03 | -0.04 | -0.16 | 0.00 | 0.00 | 0.03 |
| | | MSE | 3.60 | 3.04 | 3.06 | 2.49 | 8.39 | 5.05 | 7.26 | 4.03 |
| | 0.90 | RB | -0.18 | -0.02 | -0.05 | -0.04 | -0.18 | 0.06 | 0.01 | 0.06 |
| | | MSE | 5.90 | 5.24 | 4.68 | 4.10 | 12.78 | 10.16 | 11.75 | 8.75 |

### Performance of confidence intervals

We compare three CIs:

(i) EMP: Owen (2001)'s ELR-based CI for quantiles;

(ii) DRM: the ELR-based CI under the DRM without the common-mean assumption (Zhang et al., 2020);

(iii) DRM-EE: the proposed ELR-based CI.

The construction of CIs for the quantiles under the two-sample empirical likelihood method with the common-mean assumption has not been discussed in the literature, and hence is not included in the simulation. The CP and AL are used to compare CIs. We present the simulation results for the normal case in Table 2.5. The results for the gamma distributions display similar patterns and are included in Section 2.6.4.

The CIs for all the methods have satisfactory performance in terms of CP. However, the CIs using the DRM-EE method have the shortest AL. The results indicate that the limiting distribution of the ELR statistic in Corollary 2.3 works very well, and additional auxiliary information leads to shorter CIs.

## 2.5 Real Data Applications

The first dataset (Simpson et al., 1975) is from a randomized airborne pyrotechnic seeding experiment, which is designed to test whether seeding clouds with silver iodide increase rainfall. The measurements are the amount of rainfall (in acre-feet) from 52 isolated cumulus clouds, half of which were randomly chosen and massively injected with silver iodide smoke. The rest were untreated. We use $D = 0$ to indicate untreated clouds and $D = 1$ for seeded clouds. We estimate the mean ratio $\delta$ of the two populations and construct CIs for $\delta$.

To use our proposed method to analyze the dataset, we need to choose an appropriate $\boldsymbol{q}(x)$ in the DRM (2.1). Simpson et al. (1975) and Krishnamoorthy and Mathew (2003) argued that this dataset is highly skewed. This suggests that the two-sample data can be fitted by the DRM with $\boldsymbol{q}(x) = \log x$. The goodness-of-fit test of Qin and Zhang (1997) gives a $p$-value of 0.568, which indicates that the DRM with $\boldsymbol{q}(x) = \log x$ provides an adequate fit to the two-sample data. Since there is no auxiliary information available, we

Table 2.5: CP (%) and AL for 95% CIs of 100$\tau$%-quantiles (normal distributions).

| | | | $N(18,4)$ | | | $N(18,9)$ | | |
|---|---|---|---|---|---|---|---|---|
| $(n_0, n_1)$ | $\tau$ | | EMP | DRM | DRM-EE | EMP | DRM | DRM-EE |
| (50,50) | 0.10 | CP | 94.5 | 94.3 | 94.2 | 94.4 | 94.5 | 94.8 |
| | | AL | 1.96 | 1.74 | 1.61 | 2.94 | 2.89 | 2.48 |
| | 0.25 | CP | 95.9 | 95.1 | 95.2 | 95.0 | 94.8 | 94.2 |
| | | AL | 1.60 | 1.40 | 1.25 | 2.36 | 2.18 | 1.64 |
| | 0.50 | CP | 94.3 | 94.6 | 95.4 | 93.8 | 94.8 | 95.4 |
| | | AL | 1.32 | 1.28 | 1.11 | 1.98 | 1.98 | 1.36 |
| | 0.75 | CP | 95.2 | 94.3 | 94.8 | 95.2 | 94.8 | 95.1 |
| | | AL | 1.59 | 1.39 | 1.24 | 2.36 | 2.16 | 1.63 |
| | 0.90 | CP | 94.2 | 94.5 | 93.9 | 94.3 | 95.0 | 94.9 |
| | | AL | 1.97 | 1.74 | 1.62 | 2.97 | 2.92 | 2.50 |
| (50,150) | 0.10 | CP | 94.5 | 94.3 | 95.0 | 93.7 | 94.7 | 94.7 |
| | | AL | 1.96 | 1.62 | 1.38 | 1.63 | 1.63 | 1.49 |
| | 0.25 | CP | 95.9 | 95.1 | 95.2 | 95.8 | 95.4 | 95.3 |
| | | AL | 1.60 | 1.33 | 1.02 | 1.34 | 1.28 | 1.11 |
| | 0.50 | CP | 94.3 | 95.1 | 95.5 | 94.4 | 96.0 | 96.0 |
| | | AL | 1.32 | 1.20 | 0.86 | 1.16 | 1.16 | 0.97 |
| | 0.75 | CP | 95.2 | 94.5 | 94.8 | 95.3 | 95.8 | 96.1 |
| | | AL | 1.59 | 1.31 | 1.00 | 1.32 | 1.27 | 1.10 |
| | 0.90 | CP | 94.2 | 94.8 | 94.2 | 95.2 | 94.6 | 94.3 |
| | | AL | 1.97 | 1.62 | 1.39 | 1.65 | 1.63 | 1.50 |
| (100,100) | 0.10 | CP | 95.6 | 95.0 | 95.2 | 95.9 | 94.3 | 95.2 |
| | | AL | 1.42 | 1.20 | 1.12 | 2.12 | 1.95 | 1.67 |
| | 0.25 | CP | 95.7 | 94.4 | 95.4 | 94.9 | 95.3 | 95.1 |
| | | AL | 1.10 | 1.00 | 0.89 | 1.66 | 1.51 | 1.14 |
| | 0.50 | CP | 94.8 | 94.7 | 95.2 | 95.2 | 96.1 | 95.5 |
| | | AL | 0.96 | 0.90 | 0.79 | 1.45 | 1.38 | 0.94 |
| | 0.75 | CP | 95.2 | 94.7 | 95.5 | 95.3 | 95.8 | 95.5 |
| | | AL | 1.09 | 0.98 | 0.87 | 1.62 | 1.51 | 1.14 |
| | 0.90 | CP | 95.5 | 94.2 | 94.3 | 95.6 | 95.2 | 94.8 |
| | | AL | 1.43 | 1.21 | 1.13 | 2.15 | 1.96 | 1.66 |
| (200,200) | 0.10 | CP | 93.8 | 95.4 | 95.1 | 94.5 | 94.4 | 94.9 |
| | | AL | 0.93 | 0.84 | 0.79 | 1.39 | 1.36 | 1.16 |
| | 0.25 | CP | 95.8 | 95.7 | 95.3 | 95.0 | 95.0 | 94.0 |
| | | AL | 0.77 | 0.69 | 0.62 | 1.14 | 1.06 | 0.80 |
| | 0.50 | CP | 94.9 | 95.0 | 94.6 | 95.2 | 94.9 | 95.4 |
| | | AL | 0.68 | 0.63 | 0.55 | 1.03 | 0.96 | 0.66 |
| | 0.75 | CP | 94.9 | 95.5 | 95.2 | 95.0 | 95.2 | 95.4 |
| | | AL | 0.76 | 0.69 | 0.62 | 1.14 | 1.07 | 0.81 |
| | 0.90 | CP | 95.0 | 94.4 | 95.0 | 93.7 | 94.5 | 94.6 |
| | | AL | 0.94 | 0.85 | 0.79 | 1.41 | 1.37 | 1.18 |

analyze the data using DRM and the other methods discussed in Section 2.4.1. For the point estimates, the EMP method gives 2.685, while our proposed DRM based estimate is 2.369. As we have demonstrated in Section 2.4.1, DRM provides smaller MSEs and RBs than EMP, so we expect that the DRM estimate is more accurate. We consider the three CIs at the 95% nominal level, EMP-NA, EMP-EL, and DRM. Table 2.6 presents the lower bound (LB), the upper bound (UB), and the length of the CIs. The EMP-NA CI is significantly longer than the others, and DRM provides the shortest CI. This agrees with the simulation results in Section 2.4.1. The LBs of all three CIs are greater than 1, indicating that the seeded clouds slightly increase rainfall.

Table 2.6: Summary of 95% CIs for $\delta$ (cloud data).

| | LB | UB | Length |
|---|---|---|---|
| EMP-NA | 1.13 | 6.36 | 5.23 |
| EMP-EL | 1.41 | 5.24 | 3.83 |
| DRM | 1.21 | 4.89 | 3.68 |

The second dataset (Hawkins, 2002) is from a clinical study of cyclosporine measurements in blood samples of organ transplant recipients. In total, 56 assay pairs for cyclosporine are obtained by a standard approved method, high-performance liquid chromatography (HPLC), and an alternative radio-immunoassay (RIA) method. We would like to investigate whether the RIA assay is essentially equivalent to the HPLC assay. The results in Hawkins (2002) and Bebu and Mathew (2008) indicate that the measurements from the two methods can be modeled by lognormal distributions and have a common mean. Since the quantiles are important characteristics of the population, we consider inference on these quantities at $\tau = 0, 25, 0.50, 0.75$.

Our methods and theory are applicable to two independent samples, but in this dataset, two methods are used to measure the same blood sample, so the two measurements may be correlated. To demonstrate the value of auxiliary information, we randomly split the 56 blood samples into two equal groups. We use $D = 0$ to indicate the HPLC method for the first group and $D = 1$ to indicate the RIA method for the second group. This gives two independent samples, shown in Table 2.7. We set $\boldsymbol{q}(x)$ in the DRM (2.1) to $\boldsymbol{q}(x) = (\log x, \log^2 x)^\top$. For this choice, the goodness-of-fit test of Qin and Zhang (1997) gives a $p$-value of 0.839. An ELR test to check the validity of the common-mean assumption gives a $p$-value of 0.530. This preliminary analysis indicates that the DRM with the common-mean assumption is reasonable.

We use the methods of Section 2.4.2 to analyze the independent samples. Table 2.8 summarizes the point estimates and 95% CIs. Note that the empirical likelihood method does not specify how to construct CIs for quantiles with the common-mean assumption.

Table 2.7: Measurements from HPLC and RIA methods in two independent samples.

| HPLC ($D = 0$) | | | | | | | RIA ($D = 1$) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 77 | 87 | 93 | 109 | 109 | 129 | 130 | 38 | 98 | 108 | 109 | 111 | 118 | 125 |
| 153 | 156 | 159 | 185 | 198 | 203 | 227 | 130 | 144 | 149 | 162 | 165 | 169 | 172 |
| 244 | 245 | 271 | 280 | 285 | 318 | 336 | 204 | 218 | 234 | 235 | 293 | 294 | 303 |
| 339 | 340 | 440 | 498 | 521 | 556 | 578 | 311 | 341 | 376 | 404 | 406 | 477 | 679 |

We also provide the results of analyzing the original 56 pairs using the EMP method; these are recorded under "EMP–ALL" in Table 2.8 and serve as the benchmarks. Table 2.8 shows that the DRM-EE CIs are always shorter than the DRM and EMP CIs. This is in line with our simulation results. Although each independent sample is half the size of the original sample, the DRM-EE quantile estimates and CIs are similar to the EMP-ALL quantile estimates and CIs. This indicates that our method can combine information from two samples and effectively utilize available auxiliary information.

Table 2.8: Summary of point estimates and 95% CIs for quantiles (cyclosporine data).

| | | HPLC ($D = 0$) | | | | RIA ($D = 1$) | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\tau$ | | Estimate | LB | UB | Length | Estimate | LB | UB | Length |
| 0.25 | EMP-ALL | 127 | 109 | 159 | 50 | 141 | 118 | 162 | 50 |
| | EMP | 130 | 93 | 198 | 105 | 125 | 108 | 165 | 105 |
| | EL | 130 | – | – | – | 130 | – | – | – |
| | DRM | 144 | 109 | 185 | 76 | 129 | 108 | 162 | 54 |
| | DRM-EE | 130 | 109 | 165 | 56 | 130 | 109 | 162 | 53 |
| 0.5 | EMP-ALL | 206 | 159 | 271 | 112 | 196 | 162 | 287 | 112 |
| | EMP | 227 | 156 | 318 | 162 | 172 | 144 | 294 | 162 |
| | EL | 227 | – | – | – | 204 | – | – | – |
| | DRM | 234 | 162 | 303 | 141 | 198 | 149 | 280 | 131 |
| | DRM-EE | 218 | 162 | 280 | 118 | 204 | 162 | 280 | 118 |
| 0.75 | EMP-ALL | 336 | 271 | 402 | 131 | 311 | 287 | 408 | 131 |
| | EMP | 336 | 240 | 432 | 192 | 303 | 218 | 388 | 192 |
| | EL | 336 | – | – | – | 311 | – | – | – |
| | DRM | 339 | 280 | 477 | 197 | 311 | 235 | 406 | 171 |
| | DRM-EE | 318 | 280 | 404 | 124 | 336 | 280 | 406 | 126 |

## 2.6 Technical Details and Additional Simulation Results

### 2.6.1 More examples of summary quantities

In this section, we provide some more examples to demonstrate that the estimating equations $E_0\{\boldsymbol{g}(X;\boldsymbol{\psi},\boldsymbol{\theta})\} = \mathbf{0}$ can define many important summary quantities.

**Example 2.4.** *(Means and variances) Let $\mu_i$ and $\sigma_i^2$ be the mean and variance of $F_i$ for $i = 0, 1$. Further, let $\boldsymbol{\psi} = (\mu_0, \mu_1, \sigma_0^2, \sigma_1^2)^\top$ and*

$$\boldsymbol{g}(x;\boldsymbol{\psi},\boldsymbol{\theta}) = \begin{pmatrix} x - \mu_0 \\ x\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} - \mu_1 \\ x^2 - \mu_0^2 - \sigma_0^2 \\ x^2 \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} - \mu_1^2 - \sigma_1^2 \end{pmatrix}.$$

*Then these means and variances can be defined through $E_0\{\boldsymbol{g}(X;\boldsymbol{\psi},\boldsymbol{\theta})\} = \mathbf{0}$. The general uncentered and centered moments can be defined similarly.*

*Applying the results in Theorem 2.2, we can construct an empirical likelihood ratio (ELR) statistic for testing $H_0 : \sigma_0^2 = \sigma_1^2$, which to our best knowledge is new for such a testing problem.*

**Example 2.5.** *(Generalized entropy class of inequality measures) Suppose the $X_{ij}$'s are positive random variables. Let*

$$GE_i^{(\xi)} = \begin{cases} \frac{1}{\xi^2-\xi}\left\{\int_0^\infty \left(\frac{x}{\mu_i}\right)^\xi dF_i(x) - 1\right\}, & \text{if } \xi \neq 0, 1, \\ -\int_0^\infty \log\left(\frac{x}{\mu_i}\right) dF_i(x), & \text{if } \xi = 0, \\ \int_0^\infty \frac{x}{\mu_i} \log\left(\frac{x}{\mu_i}\right) dF_i(x), & \text{if } \xi = 1 \end{cases}$$

*be the generalized entropy class of inequality measures of the $i$th population, $i = 0, 1$. We assume that $GE_i^{(\xi)}$ exists. In our setup, $(GE_0^{(\xi)}, GE_1^{(\xi)})^\top$ together with $(\mu_0, \mu_1)$ can also be defined through the estimating equations. For illustration, we consider $\xi = 1$.*

*Let* $\boldsymbol{\psi} = (\mu_0, \mu_1, GE_0^{(1)}, GE_0^{(1)})^\top$ *and*

$$\boldsymbol{g}(x; \boldsymbol{\psi}, \boldsymbol{\theta}) = \begin{pmatrix} x - \mu_0 \\ x\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} - \mu_1 \\ x\log(x/\mu_0) - \mu_0 GE_0^{(1)} \\ x\log(x/\mu_1)\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} - \mu_1 GE_1^{(1)} \end{pmatrix}.$$

*Then* $(GE_0^{(\xi)}, GE_1^{(\xi)})^\top$ *together with* $(\mu_0, \mu_1)$ *can be defined through* $E_0\{\boldsymbol{g}(X; \boldsymbol{\psi}, \boldsymbol{\theta})\} = \mathbf{0}$*. For other values of* $\xi$*, we can define the corresponding estimating equations similarly.*

*Applying the results in Theorem 2.2, we can also construct an ELR statistic for testing* $H_0 : GE_0^{(\xi)} = GE_1^{(\xi)}$*. Again, to our best knowledge this ELR statistic is new for such testing problems.*

**Example 2.6.** *(Cumulative distribution functions) Suppose we are interested in* $\zeta_0 = F_0(x_0)$ *and* $\zeta_1 = F_1(x_1)$*, where* $x_0$ *and* $x_1$ *are fixed points. Let* $\boldsymbol{\psi} = (\zeta_0, \zeta_1)^\top$ *and*

$$\boldsymbol{g}(x; \boldsymbol{\psi}, \boldsymbol{\theta}) = \begin{pmatrix} I(x \le x_0) - \zeta_0 \\ \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} I(x \le x_1) - \zeta_1 \end{pmatrix}.$$

*Then* $(\zeta_0, \zeta_1)^\top$ *can be defined through* $E_0\{\boldsymbol{g}(X; \boldsymbol{\psi}, \boldsymbol{\theta})\} = \mathbf{0}$*.*

*Applying the results in Theorem 2.2, we can also construct an ELR-based confidence interval (CI) for* $\zeta_0$ *or* $\zeta_1$ *or an ELR-based confidence region for* $(\zeta_0, \zeta_1)^\top$*.*

**Example 2.7.** *(Quantiles) Suppose we are interested in* $\xi_{0,\tau_0} = \inf\{x : F_0(x) \ge \tau_0\}$ *and* $\xi_{1,\tau_1} = \inf\{x : F_1(x) \ge \tau_1\}$*, where* $\tau_0, \tau_1 \in (0, 1)$*. Let* $\boldsymbol{\psi} = (\zeta_0, \zeta_1)^\top$ *and*

$$\boldsymbol{g}(x; \boldsymbol{\psi}, \boldsymbol{\theta}) = \begin{pmatrix} I(x \le \xi_{0,\tau_0}) - \tau_0 \\ \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} I(x \le \xi_{1,\tau_1}) - \tau_1 \end{pmatrix}.$$

*Then* $(\xi_{0,\tau_0}, \xi_{1,\tau_1})^\top$ *can be defined through* $E_0\{\boldsymbol{g}(X; \boldsymbol{\psi}, \boldsymbol{\theta})\} = \mathbf{0}$*.*

*Applying the result of Corollary 2.2 or 2.3, we can also construct an ELR-based CI for* $\xi_{0,\tau_0}$ *or* $\xi_{1,\tau_1}$ *or an ELR-based confidence region for* $(\xi_{0,\tau_0}, \xi_{1,\tau_1})^\top$*.*

### 2.6.2 Summary-level information from external case-control studies

Let $\{(Y_i, D_i) : i = 1, \ldots, n_E\}$ be the data from an external study, where $D_i = 0$ or 1 indicates that the individual is from a disease-free or diseased group. We model the relationship between $D$ and $Y$ through a logistic regression model, which may not be the true model:

$$h(Y;\alpha_Y,\boldsymbol{\beta}_Y)=P(D=1|Y)=\frac{\exp(\alpha_Y+\boldsymbol{\beta}_Y^\top Y)}{1+\exp(\alpha_Y+\boldsymbol{\beta}_Y^\top Y)}. \tag{2.19}$$

Let

$$\boldsymbol{a}(\alpha_Y,\boldsymbol{\beta}_Y)=\frac{1}{n_E}\sum_{i=1}^{n_E}\{D_i-h(Y_i;\alpha_Y,\boldsymbol{\beta}_Y)\}(1,Y^\top)^\top,$$

which are the score functions based on the logistic regression model in (2.19). Further, let $(\alpha_Y^*,\boldsymbol{\beta}_Y^*)$ be the solution to $E\{\boldsymbol{a}(\alpha_Y,\boldsymbol{\beta}_Y)\}=\mathbf{0}$. That is,

$$E\{\boldsymbol{a}(\alpha_Y^*,\boldsymbol{\beta}_Y^*)\}=\mathbf{0}.$$

Note that $(\alpha_Y^*,\boldsymbol{\beta}_Y^*)$ may not be known exactly. We can solve the score equations $\boldsymbol{a}(\alpha_Y,\boldsymbol{\beta}_Y)=\mathbf{0}$ to obtain the estimator $(\hat\alpha_Y,\hat{\boldsymbol{\beta}}_Y)$. That is, $\boldsymbol{a}(\hat\alpha_Y,\hat{\boldsymbol{\beta}}_Y)=\mathbf{0}$. Assume that we have access to the estimator $(\hat\alpha_Y,\hat{\boldsymbol{\beta}}_Y)$ but not necessarily to the individual-level data $\{(Y_i,D_i):i=1,\ldots,n_E\}$.

When the total sample size $n=n_0+n_1$ for the internal study satisfies $n/n_E\to 0$, we can use $(\hat\alpha_Y,\hat{\boldsymbol{\beta}}_Y)$ for $(\alpha_Y^*,\boldsymbol{\beta}_Y^*)$. This will cause a negligible error for inference for the internal study. In the following, we assume that $(\alpha_Y^*,\boldsymbol{\beta}_Y^*)$ is known and we denote $h(y)=h(y;\alpha_Y^*,\boldsymbol{\beta}_Y^*)$.

Next, we discuss how to summarize the information from $E\{\boldsymbol{a}(\alpha_Y^*,\boldsymbol{\beta}_Y^*)\}=\mathbf{0}$ into unbiased estimating equations with respect to $F_0$. When the external study is a prospective case-control study, by defining the unknown overall disease prevalence $\pi=P(D=1)$, we have

$$\begin{aligned}&E\{\boldsymbol{a}(\alpha_Y^*,\boldsymbol{\beta}_Y^*)\}\\=\;&E\left[\{D-h(Y)\}(1,Y^\top)^\top\right] &(2.20)\\=\;&E_0\Big([-(1-\pi)h(Y)+\pi\exp\{\boldsymbol{\theta}^\top\boldsymbol{Q}(X)\}\{1-h(Y)\}](1,Y^\top)^\top\Big), &(2.21)\end{aligned}$$

where we have used the law of total expectation and the DRM (2.1) in the last step.

When the external study is a retrospective case-control study, we have

$$\begin{aligned}&E\{\boldsymbol{a}(\alpha_Y^*,\boldsymbol{\beta}_Y^*)\}\\=\;&-(1-\pi_E)E_0\{h(Y)(1,Y^\top)^\top\}+\pi_E E_1[\{1-h(Y)\}(1,Y^\top)^\top], &(2.22)\end{aligned}$$

where $E_1$ represents the expectation operators with respect to $F_1$, and $\pi_E$ is the proportion of diseased individuals in the external case-control study. Note that $\pi_E$ is a known and fixed value.

Using the DRM (2.1), we further get

$$\begin{aligned} & E\{\boldsymbol{a}(\alpha_Y^*, \boldsymbol{\beta}_Y^*)\} \\ = \; & E_0\Big([-(1-\pi_E)h(Y) + \pi_E \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X)\}\{1-h(Y)\}](1, Y^\top)^\top\Big). \end{aligned} \tag{2.23}$$

Summarizing (2.21) and (2.23), we have that if the external study is a prospective case-control study, then $E_0\{\boldsymbol{g}(X; \boldsymbol{\psi}, \boldsymbol{\theta})\} = \mathbf{0}$, where

$$\boldsymbol{g}(x; \boldsymbol{\psi}, \boldsymbol{\theta}) = [-(1-\pi)h(y) + \pi \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\}\{1-h(y)\}](1, y^\top)^\top$$

with $\boldsymbol{\psi} = \pi$; if the external study is a retrospective case-control study, then $E_0\{\boldsymbol{g}(X; \boldsymbol{\theta})\} = \mathbf{0}$, where

$$\boldsymbol{g}(x; \boldsymbol{\theta}) = [-(1-\pi_E)h(y) + \pi_E \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\}\{1-h(y)\}](1, y^\top)^\top.$$

Similarly, we summarize the information from $E\{\boldsymbol{a}(\alpha_Y^*, \boldsymbol{\beta}_Y^*)\} = \mathbf{0}$ into unbiased estimating equations with respect to the joint distribution of $(D, Y)$, which is the setup in Chatterjee et al. (2016). Note that when the external study is a retrospective case-control study, Equation (2.22) can be equivalently written as

$$\begin{aligned} & E\{\boldsymbol{a}(\alpha_Y^*, \boldsymbol{\beta}_Y^*)\} \\ = \; & E\left[\frac{1-\pi_E}{1-\pi}(1-D)\{D-h(Y)\}(1, Y^\top)^\top + \frac{\pi_E}{\pi} D\{D-h(Y)\}(1, Y^\top)^\top\right]. \end{aligned} \tag{2.24}$$

Summarizing (2.20) and (2.24), we have that if the external study is a prospective case-control study, then $E\{\boldsymbol{u}(D, Y)\} = \mathbf{0}$, where

$$\boldsymbol{u}(D, Y) = \{D - h(Y)\}(1, Y^\top)^\top;$$

if the external study is a retrospective case-control study, then $E\{\boldsymbol{u}(D, Y; \pi)\} = \mathbf{0}$, where

$$\boldsymbol{u}(D, Y; \pi) = \frac{1-\pi_E}{1-\pi}(1-D)\{D-h(Y)\}(1, Y^\top)^\top + \frac{\pi_E}{\pi} D\{D-h(Y)\}(1, Y^\top)^\top.$$

Note that the method and theory in Chatterjee et al. (2016) are applicable when there is no unknown parameter in the functions $\boldsymbol{u}(\cdot)$. Hence, their general results do not apply when the external study is a retrospective case-control study.

### 2.6.3 Proofs

**Regularity conditions**

The asymptotic results in this chapter are established under the following regularity conditions. We use $||\cdot||$ to denote the Euclidean norm, i.e., $||\cdot||^2$ is the sum of squares of the elements.

C1. The total sample size $n = n_0 + n_1 \to \infty$ and $n_1/n \to \lambda^*$ for some constant $\lambda^* \in (0,1)$.

C2. The two CDFs $F_0$ and $F_1$ satisfy the DRM (2.1) with a true parameter value $\boldsymbol{\theta}^*$, and $\int \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} dF_0(x) < \infty$ in a neighborhood of the true value $\boldsymbol{\theta}^*$.

C3. $\int \boldsymbol{Q}(x)^\top \boldsymbol{Q}(x) dF_0(x)$ exists and is positive definite.

C4. $E_0\{\boldsymbol{g}(X;\boldsymbol{\psi}^*,\boldsymbol{\theta}^*)\} = \boldsymbol{0}$, $E_0\{\partial \boldsymbol{g}(X;\boldsymbol{\psi}^*,\boldsymbol{\theta}^*)/\partial\boldsymbol{\eta}\}$ has rank $p$, and $\int \boldsymbol{G}(X)\boldsymbol{G}(X)^\top dF_0(x)$ exists and is positive definite, where $\boldsymbol{G}(x)$ is defined before Theorem 2.1.

C5. $\boldsymbol{G}(x;\boldsymbol{\eta})$ is twice differentiable with respect to $\boldsymbol{\eta}$, and $||\boldsymbol{G}(x,\boldsymbol{\eta})||^3$, $||\partial\boldsymbol{G}(x,\boldsymbol{\eta})/\partial\boldsymbol{\eta}||^2$, and $||\partial\boldsymbol{G}(x,\boldsymbol{\eta})/\{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\tau\}||$ are bounded by some integrable function $R(x)$ with respect to both $F_0$ and $F_1$ in the neighborhood of $\boldsymbol{\eta}^*$.

Conditions C1–C3 ensure that the quadratic approximation of the dual likelihood $\ell_{nd}$ in (2.13) is applicable. Condition C1 indicates that both $n_0$ and $n_1$ go to infinity at the same rate. For simplicity, and convenience of presentation, we write $\lambda^* = n_1/n$ and assume that it is a constant. This does not affect our technical development. Condition C2 guarantees the existence of finite moments of $\boldsymbol{Q}(x)$ in a neighborhood of $\boldsymbol{\theta}^*$. Condition C3 is an identifiability condition, and it ensures that the components of $\boldsymbol{Q}(x)$ are linearly independent under both $F_i$'s, and hence the elements of $\boldsymbol{Q}(x)$ except the first cannot be constant functions. Conditions C3 and C4 together ensure that $\boldsymbol{U}$ and $\boldsymbol{V}$ in Theorem 2.1 have full rank, guaranteeing that $\boldsymbol{J}$ is invertible. Conditions C1–C5 guarantee that quadratic approximations of the profile empirical log-likelihood $\ell_n(\boldsymbol{\psi},\boldsymbol{\theta})$ are applicable.

**Some preliminary results**

Recall that the profile empirical log-likelihood of $(\boldsymbol{\psi}, \boldsymbol{\theta})$ is

$$\ell_n(\boldsymbol{\psi}, \boldsymbol{\theta}) = -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \log\left\{1 + \lambda\left[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1\right] + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})\right\} + \sum_{j=1}^{n_1} \boldsymbol{\theta}^\top \boldsymbol{Q}(X_{1j}),$$

where the Lagrange multipliers satisfy

$$\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1}{1 + \lambda\left[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1\right] + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})} = 0,$$
$$\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})}{1 + \lambda\left[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1\right] + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})} = \mathbf{0}.$$

Then $\ell_n(\boldsymbol{\psi}, \boldsymbol{\theta})$ can be rewritten as

$$\ell_n(\boldsymbol{\psi}, \boldsymbol{\theta}) = \inf_{\lambda, \boldsymbol{\nu}} l_n(\boldsymbol{\psi}, \boldsymbol{\theta}, \lambda, \boldsymbol{\nu}),$$

where

$$\begin{aligned} & l_n(\boldsymbol{\psi}, \boldsymbol{\theta}, \lambda, \boldsymbol{\nu}) \\ = \; & -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \log\left\{1 + \lambda\left[\exp\left\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\right\} - 1\right] + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})\right\} + \sum_{j=1}^{n_1}\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{1j})\}. \end{aligned}$$

Equivalently, $\ell_n(\boldsymbol{\psi}, \boldsymbol{\theta}) = l_n(\boldsymbol{\psi}, \boldsymbol{\theta}, \lambda, \boldsymbol{\nu})$ with $\lambda$ and $\boldsymbol{\nu}$ being the solution to

$$\frac{\partial l_n(\boldsymbol{\psi}, \boldsymbol{\theta}, \lambda, \boldsymbol{\nu})}{\partial \lambda} = 0 \quad \text{and} \quad \frac{\partial l_n(\boldsymbol{\psi}, \boldsymbol{\theta}, \lambda, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} = \mathbf{0}.$$

With the above preparation, it can be verified that the MELE $(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})$ of $(\boldsymbol{\psi}, \boldsymbol{\theta})$ and the corresponding Lagrange multipliers $(\hat{\lambda}, \hat{\boldsymbol{\nu}})$ satisfy

$$\frac{\partial l_n(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}, \hat{\lambda}, \hat{\boldsymbol{\nu}})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \quad \frac{\partial l_n(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}, \hat{\lambda}, \hat{\boldsymbol{\nu}})}{\partial \boldsymbol{\beta}} = \mathbf{0}, \quad \frac{\partial l_n(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}, \hat{\lambda}, \hat{\boldsymbol{\nu}})}{\partial \lambda} = 0, \quad \frac{\partial l_n(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}, \hat{\lambda}, \hat{\boldsymbol{\nu}})}{\partial \boldsymbol{\nu}} = \mathbf{0}.$$

To investigate the asymptotic properties of $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\theta}}$, we need their approximations. We first find the first and second derivatives of $l_n(\boldsymbol{\psi}, \boldsymbol{\theta}, \lambda, \boldsymbol{\nu})$.

Recall that $\boldsymbol{\eta} = (\boldsymbol{\psi}^\top, \boldsymbol{\theta}^\top)^\top$ and $\boldsymbol{u} = (\lambda, \boldsymbol{\nu}^\top)^\top$. The MELE and true value of $\boldsymbol{\eta}$ are $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\psi}}^\top, \hat{\boldsymbol{\theta}}^\top)^\top$ and $\boldsymbol{\eta}^* = (\boldsymbol{\psi}^{*\top}, \boldsymbol{\theta}^{*\top})^\top$. Let $\boldsymbol{\gamma} = (\boldsymbol{\eta}^\top, \boldsymbol{u}^\top)^\top$. We further define

$$\hat{\boldsymbol{u}} = (\hat{\lambda}, \hat{\boldsymbol{\nu}}^\top)^\top, \quad \boldsymbol{u}^* = (\lambda^*, \mathbf{0}_{1\times r})^\top, \quad \hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\eta}}^\top, \hat{\boldsymbol{u}}^\top)^\top, \quad \boldsymbol{\gamma}^* = (\boldsymbol{\eta}^{*\top}, \boldsymbol{u}^{*\top})^\top.$$

In the following, we use $l_n(\boldsymbol{\gamma})$ and $\boldsymbol{g}(x;\boldsymbol{\eta})$ to denote $l_n(\boldsymbol{\psi}, \boldsymbol{\theta}, \lambda, \boldsymbol{\nu})$ and $\boldsymbol{g}(x;\boldsymbol{\psi}, \boldsymbol{\theta})$.

- *First and second derivatives of* $l_n(\boldsymbol{\gamma})$

After some straightforward algebraic manipulations, the first derivatives of $l_n(\boldsymbol{\gamma})$ are found to be:

$$\begin{aligned}
\frac{\partial l_n(\boldsymbol{\gamma})}{\partial \boldsymbol{\psi}} &= -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\{\partial \boldsymbol{g}(X_{ij};\boldsymbol{\eta})/\partial \boldsymbol{\psi}\}^\top \boldsymbol{\nu}}{1 + \lambda\{\omega(X_{ij};\boldsymbol{\theta}) - 1\} + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij};\boldsymbol{\eta})}, \\
\frac{\partial l_n(\boldsymbol{\gamma})}{\partial \boldsymbol{\theta}} &= -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\lambda \omega(X_{ij};\boldsymbol{\theta})\boldsymbol{Q}(X_{ij}) + \{\partial \boldsymbol{g}(X_{ij};\boldsymbol{\eta})/\partial \boldsymbol{\theta}\}^\top \boldsymbol{\nu}}{1 + \lambda\{\omega(X_{ij};\boldsymbol{\theta}) - 1\} + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij};\boldsymbol{\eta})} + \sum_{j=1}^{n_1} \boldsymbol{Q}(X_{1j}), \\
\frac{\partial l_n(\boldsymbol{\gamma})}{\partial \boldsymbol{u}} &= -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\boldsymbol{G}(X_{ij};\boldsymbol{\eta})}{1 + \lambda\{\omega(X_{ij};\boldsymbol{\theta}) - 1\} + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij};\boldsymbol{\eta})}.
\end{aligned}$$

Then the first derivatives at the true values $\boldsymbol{\eta}^*$ and $\boldsymbol{u}^*$ are

$$\boldsymbol{S}_n = \frac{\partial l_n(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{\gamma}} = \begin{pmatrix} \frac{\partial l_n(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{\psi}} \\ \frac{\partial l_n(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{\theta}} \\ \frac{\partial l_n(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{S}_{n\boldsymbol{\theta}} \\ \boldsymbol{S}_{n\boldsymbol{u}} \end{pmatrix},$$

where

$$\boldsymbol{S}_{n\boldsymbol{\theta}} = \sum_{j=1}^{n_1} \boldsymbol{Q}(X_{1j}) - \sum_{i=0}^{1}\sum_{j=1}^{n_i} h_1(X_{ij})\boldsymbol{Q}(X_{ij}), \quad \boldsymbol{S}_{n\boldsymbol{u}} = -\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\boldsymbol{G}(X_{ij})}{h(X_{ij})}.$$

Similarly, we calculate the second derivatives of $l_n(\boldsymbol{\gamma})$. Evaluating them at $\boldsymbol{\gamma}^*$ gives:

$$\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{u}^\top} \\ \mathbf{0} & \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} & \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{\beta} \partial \boldsymbol{u}^\top} \\ \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{u} \partial \boldsymbol{\theta}^\top} & \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{u} \partial \boldsymbol{\beta}^\top} & \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{u} \partial \boldsymbol{u}^\top} \end{pmatrix}, \tag{2.25}$$

where $h_0(x) = (1-\lambda^*)/h(x) = 1 - h_1(x)$ and

$$
\begin{aligned}
\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\psi}\partial\boldsymbol{u}^\top} &= \left(\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{u}\partial\boldsymbol{\psi}^\top}\right)^\top = -\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\{\partial\boldsymbol{G}(X_{ij};\boldsymbol{\eta}^*)/\partial\boldsymbol{\psi}\}^\top}{h(X_{ij})};\\
\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} &= -\sum_{i=0}^{1}\sum_{j=1}^{n_i} h_0(X_{ij})h_1(X_{ij})\boldsymbol{Q}(X_{ij})\boldsymbol{Q}(X_{ij})^\top;\\
\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{u}^\top} &= \left(\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{u}\partial\boldsymbol{\theta}^\top}\right)^\top\\
&= \sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{h_1(X_{ij})\boldsymbol{Q}(X_{ij})\boldsymbol{G}(X_{ij})^\top}{h(X_{ij})} - \sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\{\partial\boldsymbol{G}(X_{ij};\boldsymbol{\eta}^*)/\partial\boldsymbol{\theta}\}^\top}{h(X_{ij})};\\
\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{u}\partial\boldsymbol{u}^\top} &= \sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\boldsymbol{G}(X_{ij})\boldsymbol{G}(X_{ij})^\top}{h(X_{ij})^2}.
\end{aligned}
$$

- *Some useful lemmas*

We first review a lemma from the supplementary material of Qin et al. (2015), which helps to ease the calculation in our proofs. In the following, we assume that the DRM (2.1) is satisfied as required in Condition C2.

**Lemma 2.1.** *Suppose that $\mathcal{S}$ is an arbitrary vector-valued function. Let $E_0(\cdot)$ represent the expectation operator with respect to $F_0$ and $X$ refer to a random variable from $F_0$. Then we have for $j = 1, \cdots, n_1$,*

$$
E\{\mathcal{S}(X_{1j})\} = E_0\{\omega(X)\mathcal{S}(X)\} \quad \text{and} \quad E\left\{\sum_{i=0}^{1}\sum_{j=1}^{n_i}\mathcal{S}(X_{ij})\right\} = nE_0\{\mathcal{S}(X)h(X)\}.
$$

*Proof.* Under the DRM with true parameter $\boldsymbol{\theta}^*$, we have

$$
E\{\mathcal{S}(X_{1j})\} = \int \mathcal{S}(x)dF_1(x) = \int \mathcal{S}(x)\omega(x)dF_0(x) = E_0\{\omega(X)\mathcal{S}(X)\}.
$$

Using the fact that $\lambda^* = n_1/n$ and the definition of the function $h(\cdot)$, we further have

$$
\begin{aligned}
E\left\{\sum_{i=0}^{1}\sum_{j=1}^{n_i}\mathcal{S}(X_{ij})\right\} &= n_0 E_0\left\{\mathcal{S}(X)\right\} + n_1 E_0\left\{\omega(X)\mathcal{S}(X)\right\} \\
&= n\left[(1-\lambda^*)E_0\left\{\mathcal{S}(X)\right\} + \lambda^* E_0\left\{\omega(X)\mathcal{S}(X)\right\}\right] \\
&= nE_0\left[\{(1-\lambda^*) + \lambda^*\omega(X)\}\mathcal{S}(X)\right] \\
&= nE_0\left\{\omega(X)\mathcal{S}(X)\right\}.
\end{aligned}
$$

This completes the proof. □

Recall that

$$
\begin{aligned}
&\boldsymbol{A}_{\boldsymbol{\theta\theta}} = (1-\lambda^*)E_0\left\{h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}(X)^\top\right\}, \\
&\boldsymbol{A}_{\boldsymbol{\theta u}} = \boldsymbol{A}_{\boldsymbol{u\theta}}^\top = E_0\left\{\frac{\partial \boldsymbol{G}(X;\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}}\right\}^\top - E_0\left\{h_1(X)\boldsymbol{Q}(X)\boldsymbol{G}(X)^\top\right\}, \\
&\boldsymbol{A}_{\boldsymbol{\psi u}} = \boldsymbol{A}_{\boldsymbol{u\psi}}^\top = E_0\left\{\frac{\partial \boldsymbol{G}(X;\boldsymbol{\eta}^*)}{\partial\boldsymbol{\psi}}\right\}^\top, \quad \boldsymbol{A}_{\boldsymbol{uu}} = E_0\left\{\frac{\boldsymbol{G}(X)\boldsymbol{G}(X)^\top}{h(X)}\right\}.
\end{aligned}
$$

Applying Lemma 2.1, after some algebra, we have the following Lemma.

**Lemma 2.2.** *(a) With the form of $\partial^2 l_n(\boldsymbol{\gamma}^*)/(\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top)$ defined in (2.25), we have*

$$
-\frac{1}{n}E\left\{\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right\} = \boldsymbol{A} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\boldsymbol{\psi u}} \\ \boldsymbol{0} & \boldsymbol{A}_{\boldsymbol{\theta\theta}} & \boldsymbol{A}_{\boldsymbol{\theta u}} \\ \boldsymbol{A}_{\boldsymbol{u\psi}} & \boldsymbol{A}_{\boldsymbol{u\theta}} & -\boldsymbol{A}_{\boldsymbol{uu}} \end{pmatrix}.
$$

*(b) Let $\boldsymbol{S}_n^* = (\boldsymbol{S}_{n\boldsymbol{\theta}}^\top, \boldsymbol{S}_{n\boldsymbol{u}}^\top)^\top$. Then as $n \to \infty$,*

$$
n^{-1/2}\boldsymbol{S}_n^* \to N(\boldsymbol{0}, \boldsymbol{\Gamma})
$$

*in distribution with*

$$
\boldsymbol{e}_{\boldsymbol{\theta}} = \begin{pmatrix} 1 \\ \boldsymbol{0}_{d\times 1} \end{pmatrix}, \quad \boldsymbol{e}_{\boldsymbol{u}} = \begin{pmatrix} 1 \\ \boldsymbol{0}_{r\times 1} \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}} \\ -\lambda^*(1-\lambda^*)\boldsymbol{A}_{\boldsymbol{uu}}\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix},
$$

$$
\text{and } \boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta\theta}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{\boldsymbol{uu}} \end{pmatrix} - \frac{1}{\lambda^*(1-\lambda^*)}\boldsymbol{C}\boldsymbol{C}^\top.
$$

*Proof.* For (a): Note that Conditions C3 and C4 ensure that $\boldsymbol{A}$ is well defined. The results then follow by applying Lemma 2.1 to each term of $E\left\{\partial^2 l_n(\boldsymbol{\gamma}^*)/(\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top)\right\}$. We use $E\left\{\partial^2 l_n(\boldsymbol{\gamma}^*)/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top)\right\}$ as an illustration; for the other entries, the idea is similar and we omit the details.

With Lemma 2.1 and the fact that $h_0(x)h(x) = 1-\lambda^*$, we have

$$
\begin{aligned}
-\frac{1}{n}E\left\{\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\} &= \frac{1}{n}E\left\{\sum_{i=0}^{1}\sum_{j=1}^{n_i} h_0(X_{ij})h_1(X_{ij})\boldsymbol{Q}(X_{ij})\boldsymbol{Q}(X_{ij})^\top\right\} \\
&= (1-\lambda^*)E_0\left\{h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}(X)^\top\right\} \\
&= \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}.
\end{aligned}
$$

For (b): Conditions C2–C4 ensure that $E(\boldsymbol{S}_n^*)$ and $Var(\boldsymbol{S}_n^*)$ are well defined. We first use the results in Lemma 2.1 to show that $E(\boldsymbol{S}_n^*) = \boldsymbol{0}$. For $E(\boldsymbol{S}_{n\boldsymbol{\theta}})$,

$$
\begin{aligned}
E(\boldsymbol{S}_{n\boldsymbol{\theta}}) &= n_1 E\{\boldsymbol{Q}(X_{11})\} - nE_0\{h(X)h_1(X)\boldsymbol{Q}(X)\} \\
&= n_1 E_0\{\omega(X)\boldsymbol{Q}(X)\} - nE_0\{\lambda^*\omega(X)\boldsymbol{Q}(X)\} \\
&= \boldsymbol{0}.
\end{aligned}
$$

The last step follows from the fact that $\lambda^* = n_1/n$.

The unbiasedness of the estimating equations leads to

$$
E(\boldsymbol{S}_{n\boldsymbol{u}}) = -nE_0\{\boldsymbol{G}(\boldsymbol{X};\boldsymbol{\eta}^*)\} = \boldsymbol{0}.
$$

Hence, we have $E(\boldsymbol{S}_n^*) = \boldsymbol{0}$.

Since $\boldsymbol{S}_n^*$ is a summation of independent random vectors, by the central limit theorem,

$$
n^{-1/2}\boldsymbol{S}_n^* \to N(\boldsymbol{0}, \boldsymbol{\Gamma})
$$

for some $\boldsymbol{\Gamma}$. Next, we show that $\boldsymbol{\Gamma}$ has the form claimed in the lemma.

We start with the variances of $n^{-1/2}\boldsymbol{S}_{n\boldsymbol{\theta}}$ and $n^{-1/2}\boldsymbol{S}_{n\boldsymbol{u}}$. Note that

$$
\boldsymbol{S}_{n\boldsymbol{\theta}} = \sum_{j=1}^{n_1} h_0(X_{1j})\boldsymbol{Q}(X_{1j}) - \sum_{j=1}^{n_0} h_1(X_{0j})\boldsymbol{Q}(X_{0j}).
$$

With the help of Lemma 2.1, we have

$$
\begin{aligned}
Var(n^{-1/2}\boldsymbol{S}_{n\boldsymbol{\theta}}) &= \frac{1}{n}Var\left(\sum_{j=1}^{n_1} h_0(X_{1j})\boldsymbol{Q}(X_{1j}) - \sum_{j=1}^{n_0} h_1(X_{0j})\boldsymbol{Q}(X_{0j})\right)\\
&= \lambda^* E_0\left\{h_0(X)^2\omega(X)\boldsymbol{Q}(X)\boldsymbol{Q}(X)^\top\right\}\\
&\quad +(1-\lambda^*)E_0\left\{h_1(X)^2\boldsymbol{Q}(X)\boldsymbol{Q}(X)^\top\right\}\\
&\quad -\lambda^* E_0\left\{h_0(X)\omega(X)\boldsymbol{Q}(X)\right\}E_0\left\{h_0(X)\omega(X)\boldsymbol{Q}(X)^\top\right\}\\
&\quad -(1-\lambda^*)E_0\left\{h_1(X)\boldsymbol{Q}(X)\right\}E_0\left\{h_1(X)\boldsymbol{Q}(X)^\top\right\}.
\end{aligned}
$$

Using the definitions of functions $h_1(\cdot)$ and $h_0(\cdot)$ and the fact that $\lambda^* = n_1/n$, we further have

$$
\begin{aligned}
Var(n^{-1/2}\boldsymbol{S}_{n\boldsymbol{\theta}}) &= (1-\lambda^*)E_0\left\{h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}(X)^\top\right\}\\
&\quad -\frac{1-\lambda^*}{\lambda^*}E_0\left\{h_1(X)\boldsymbol{Q}(X)\right\}E_0\left\{h_1(X)\boldsymbol{Q}(X)^\top\right\}\\
&= \boldsymbol{A}_{\boldsymbol{\theta\theta}} - \{\lambda^*(1-\lambda^*)\}^{-1}\boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}}\left(\boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}}\right)^\top.
\end{aligned}
$$

Similarly, we calculate the variance of $n^{-1/2}\boldsymbol{S}_{n\boldsymbol{u}}$ as

$$
\begin{aligned}
&Var(n^{-1/2}\boldsymbol{S}_{n\boldsymbol{u}})\\
=\ & \frac{1}{n}Var\left\{-\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\boldsymbol{G}(X_{ij})}{h(X_{ij})}\right\}\\
=\ & \frac{1}{n}\sum_{i=0}^{1}\sum_{j=1}^{n_i}E_0\left\{\frac{\boldsymbol{G}(X_{ij})\boldsymbol{G}(X_{ij})^\top}{h(X_{ij})^2}\right\} - \frac{1}{n}\sum_{j=1}^{n_0}E_0\left\{\frac{\boldsymbol{G}(X_{0j})}{h(X_{0j})}\right\}E_0\left\{\frac{\boldsymbol{G}(X_{0j})^\top}{h(X_{0j})}\right\}\\
&-\frac{1}{n}\sum_{j=1}^{n_1}E_0\left\{\frac{\omega(X_{1j})\boldsymbol{G}(X_{1j})}{h(X_{1j})}\right\}E_0\left\{\frac{\omega(X_{1j})\boldsymbol{G}(X_{1j})^\top}{h(X_{1j})}\right\}\\
=\ & \boldsymbol{A}_{\boldsymbol{uu}} - (1-\lambda^*)E_0\left\{\frac{\boldsymbol{G}(X)}{h(X)}\right\}E_0\left\{\frac{\boldsymbol{G}(X)^\top}{h(X)}\right\}\\
&-\lambda^* E_0\left\{\frac{\omega(X)\boldsymbol{G}(X)}{h(X)}\right\}E_0\left\{\frac{\omega(X)\boldsymbol{G}(X)^\top}{h(X)}\right\}.
\end{aligned}
$$

It can easily be verified that

$$(1-\lambda^*)E_0\left\{\frac{\boldsymbol{G}(X)}{h(X)}\right\}+\lambda^*E_0\left\{\frac{\omega(X)\boldsymbol{G}(X)}{h(X)}\right\}=E_0\left\{\boldsymbol{G}(X)\right\}=\boldsymbol{0},$$

which implies that

$$E_0\left\{\frac{\{\omega(X)-1\}\boldsymbol{G}(X)}{h(X)}\right\}=-\frac{1}{\lambda^*}E_0\left\{\frac{\boldsymbol{G}(X)}{h(X)}\right\}=\boldsymbol{A}_{\boldsymbol{uu}}\boldsymbol{e}_{\boldsymbol{u}}.$$

Therefore,

$$Var(n^{-1/2}\boldsymbol{S}_{n\boldsymbol{u}})=\boldsymbol{A}_{\boldsymbol{uu}}-\lambda^*(1-\lambda^*)\boldsymbol{A}_{\boldsymbol{uu}}\boldsymbol{e}_{\boldsymbol{u}}(\boldsymbol{A}_{\boldsymbol{uu}}\boldsymbol{e}_{\boldsymbol{u}})^\top.$$

Lastly, we consider the covariance between $n^{-1/2}\boldsymbol{S}_{n\boldsymbol{\theta}}$ and $n^{-1/2}\boldsymbol{S}_{n\boldsymbol{u}}$:

$$\begin{aligned}
& Cov(n^{-1/2}\boldsymbol{S}_{n\boldsymbol{\theta}}, n^{-1/2}\boldsymbol{S}_{n\boldsymbol{u}}) \\
= & -\frac{1}{n}Cov\left(\sum_{j=1}^{n_1}h_0(X_{1j})\boldsymbol{Q}(X_{1j})-\sum_{j=1}^{n_0}h_1(X_{0j})\boldsymbol{Q}(X_{0j}),\quad \sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\boldsymbol{G}(X_{ij})^\top}{h(X_{ij})}\right) \\
= & -\frac{1}{n}\sum_{j=1}^{n_1}Cov\left(h_0(X_{1j})\boldsymbol{Q}(X_{1j}), \frac{\boldsymbol{G}(X_{1j})^\top}{h(X_{1j})}\right)+\frac{1}{n}\sum_{j=1}^{n_0}Cov\left(h_1(X_{0j})\boldsymbol{Q}(X_{0j}), \frac{\boldsymbol{G}(X_{0j})^\top}{h(X_{0j})}\right) \\
= & \lambda^*E_0\left\{\omega(X)h_0(X)\boldsymbol{Q}(X)\right\}E_0\left\{\frac{\omega(X)\boldsymbol{G}(X)^\top}{h(X)}\right\} \\
& -(1-\lambda^*)E_0\left\{h_1(X)\boldsymbol{Q}(X)\right\}E_0\left\{\frac{\boldsymbol{G}(X)^\top}{h(X)}\right\} \\
= & (1-\lambda^*)E_0\left\{h_1(X)\boldsymbol{Q}(X)\right\}E_0\left\{\frac{\{\omega(X)-1\}\boldsymbol{G}(X)^\top}{h(X)}\right\} \\
= & \boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}}(\boldsymbol{A}_{\boldsymbol{uu}}\boldsymbol{e}_{\boldsymbol{u}})^\top.
\end{aligned}$$

Then $\boldsymbol{\Gamma}=Var(n^{-1/2}\boldsymbol{S}_n^*)$ has the form claimed in the lemma. This completes the proof. □

**Proof of Theorem 2.1**

Recall that $\hat{\boldsymbol{\gamma}}=(\hat{\boldsymbol{\eta}}^\top,\hat{\boldsymbol{u}}^\top)^\top$ is the MELE of $\boldsymbol{\gamma}$. Using an argument similar to that in Qin and Lawless (1994) and Qin et al. (2015), we have that $\hat{\boldsymbol{\eta}}=\boldsymbol{\eta}^*+O_p(n^{-1/2})$ and $\hat{\boldsymbol{u}}=\boldsymbol{u}^*+O_p(n^{-1/2})$. To develop the asymptotic approximation of $\hat{\boldsymbol{\eta}}$, we apply the first-order Taylor expansion to $\partial l_n(\hat{\boldsymbol{\gamma}})/\partial\boldsymbol{\gamma}$ at the true value $\boldsymbol{\gamma}^*$. This, together with Condition

C5, gives

$$\mathbf{0} = \boldsymbol{S}_n + \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*) + o_p(n^{1/2}).$$

With the law of large numbers and Lemma 2.2, we have

$$\frac{1}{n}\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top} = \frac{1}{n}E\left\{\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right\} + o_p(1) = -\boldsymbol{A} + o_p(1). \tag{2.26}$$

Hence, we can write

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{pmatrix}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + \begin{pmatrix} \boldsymbol{A}_{\psi\boldsymbol{u}} \\ \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{u}} \end{pmatrix}(\hat{\boldsymbol{u}} - \boldsymbol{u}_0) = \frac{1}{n}\begin{pmatrix} \mathbf{0} \\ \boldsymbol{S}_{n\boldsymbol{\theta}} \end{pmatrix} + o_p(n^{-\frac{1}{2}}); \tag{2.27}$$

$$\begin{pmatrix} \boldsymbol{A}_{\boldsymbol{u}\psi} & \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{\theta}} \end{pmatrix}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) - \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}(\hat{\boldsymbol{u}} - \boldsymbol{u}_0) = \frac{1}{n}\boldsymbol{S}_{n\boldsymbol{u}} + o_p(n^{-\frac{1}{2}}). \tag{2.28}$$

Recall that

$$\boldsymbol{U} = \begin{pmatrix} \mathbf{0} & \boldsymbol{A}_{\psi\boldsymbol{u}} \\ \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{u}} \end{pmatrix}, \quad \boldsymbol{V} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{J} = \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top. \tag{2.29}$$

Conditions C3 and C4 ensure that $\boldsymbol{U}$, $\boldsymbol{V}$, and $\boldsymbol{J}$ have full rank. Then (2.27) and (2.28) together imply that

$$n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) = \boldsymbol{J}^{-1}\boldsymbol{U}\boldsymbol{V}^{-1}(n^{-1/2}\boldsymbol{S}_n^*) + o_p(1).$$

Applying Lemma 2.2 and Slusky's theorem, we have as $n \to \infty$

$$n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \to N(\mathbf{0}, \boldsymbol{\Sigma})$$

in distribution with $\boldsymbol{\Sigma} = \boldsymbol{J}^{-1}\boldsymbol{U}\boldsymbol{V}^{-1}Var(n^{-1/2}\boldsymbol{S}_n^*)\boldsymbol{V}^{-1}\boldsymbol{U}^\top\boldsymbol{J}^{-1}$.

Recall that

$$Var(n^{-1/2}\boldsymbol{S}_n^*) = \boldsymbol{\Gamma} = \boldsymbol{V} - \frac{1}{\lambda^*(1-\lambda^*)}\boldsymbol{C}\boldsymbol{C}^\top \quad \text{and} \quad \boldsymbol{C} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}\boldsymbol{e}_{\boldsymbol{\theta}} \\ -\lambda^*(1-\lambda^*)\boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix}.$$

Since

$$\boldsymbol{A}_{\psi\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}} = \mathbf{0} \quad \text{and} \quad \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}} = \frac{1}{\lambda^*}E_0\left\{h_1(X)\boldsymbol{Q}(X)\right\} = \frac{1}{\lambda^*(1-\lambda^*)}\boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}\boldsymbol{e}_{\boldsymbol{\theta}},$$

we have

$$\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{C} = \boldsymbol{U}\boldsymbol{V}^{-1}\begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}} \\ -\lambda^*(1-\lambda^*)\boldsymbol{A}_{\boldsymbol{uu}}\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} -\lambda^*(1-\lambda^*)\boldsymbol{A}_{\psi\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}} \\ \boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}} - \lambda^*(1-\lambda^*)\boldsymbol{A}_{\boldsymbol{\theta u}}\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix} = \mathbf{0}.$$

This leads to $\boldsymbol{\Sigma} = \boldsymbol{J}^{-1}$ and completes the proof.

**Proof of Corollary 2.1**

*Part (a).* The results in Theorem 2.1 imply that

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \to N\left(\mathbf{0}, \boldsymbol{J}_{\boldsymbol{\theta}}\right)$$

in distribution, where

$$\boldsymbol{J}_{\boldsymbol{\theta}} = \left\{\boldsymbol{A}_{\boldsymbol{\theta\theta}} + \boldsymbol{A}_{\boldsymbol{\theta u}}\boldsymbol{A}_{\boldsymbol{uu}}^{-1}\boldsymbol{A}_{\boldsymbol{u\theta}} - \boldsymbol{A}_{\boldsymbol{\theta u}}\boldsymbol{A}_{\boldsymbol{uu}}^{-1}\boldsymbol{A}_{\boldsymbol{u}\psi}\left(\boldsymbol{A}_{\psi\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{uu}}^{-1}\boldsymbol{A}_{\boldsymbol{u}\psi}\right)^{-1}\boldsymbol{A}_{\psi\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{uu}}^{-1}\boldsymbol{A}_{\boldsymbol{u\theta}}\right\}^{-1}.$$

From the definitions of $\boldsymbol{A}_{\boldsymbol{u}\psi}$ and $\boldsymbol{A}_{\boldsymbol{uu}}$, we have

$$\boldsymbol{A}_{\boldsymbol{u}\psi} = \begin{pmatrix} 0 \\ E_0\left\{\frac{\partial \boldsymbol{g}(\boldsymbol{X};\boldsymbol{\eta}^*)}{\partial \psi}\right\} \end{pmatrix}$$

and

$$\boldsymbol{A}_{\boldsymbol{uu}} = \begin{pmatrix} E_0\left\{\frac{\{\omega(X)-1\}^2}{h(X)}\right\} & E_0\left\{\frac{\{\omega(X)-1\}\boldsymbol{g}(\boldsymbol{X};\boldsymbol{\eta}^*)}{h(X)}\right\} \\ E_0\left\{\frac{\{\omega(X)-1\}\boldsymbol{g}(\boldsymbol{X};\boldsymbol{\eta}^*)^\top}{h(X)}\right\} & E_0\left\{\frac{\boldsymbol{g}(\boldsymbol{X};\boldsymbol{\eta}^*)\boldsymbol{g}(\boldsymbol{X};\boldsymbol{\eta}^*)^\top}{h(X)}\right\} \end{pmatrix}.$$

We write

$$\boldsymbol{A}_{\boldsymbol{uu}}^{-1} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{uu}}^{11} & \boldsymbol{A}_{\boldsymbol{uu}}^{12} \\ \boldsymbol{A}_{\boldsymbol{uu}}^{21} & \boldsymbol{A}_{\boldsymbol{uu}}^{22} \end{pmatrix}.$$

When $r = p$, we have

$$\begin{aligned} \left(\boldsymbol{A}_{\psi\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{uu}}^{-1}\boldsymbol{A}_{\boldsymbol{u}\psi}\right)^{-1} &= \left[E_0\left\{\frac{\partial \boldsymbol{g}(\boldsymbol{X};\boldsymbol{\eta}^*)}{\partial \boldsymbol{\psi}}\right\}^\top \boldsymbol{A}_{\boldsymbol{uu}}^{22} E_0\left\{\frac{\partial \boldsymbol{g}(\boldsymbol{X};\boldsymbol{\eta}^*)}{\partial \boldsymbol{\psi}}\right\}\right]^{-1} \\ &= \left[E_0\left\{\frac{\partial \boldsymbol{g}(\boldsymbol{X};\boldsymbol{\eta}^*)}{\partial \boldsymbol{\psi}}\right\}^\top\right]^{-1}\left(\boldsymbol{A}_{\boldsymbol{uu}}^{22}\right)^{-1}\left[E_0\left\{\frac{\partial \boldsymbol{g}(\boldsymbol{X};\boldsymbol{\eta}^*)}{\partial \boldsymbol{\psi}}\right\}\right]^{-1}. \end{aligned}$$

This leads to

$$\begin{aligned}
& \boldsymbol{A}_{\boldsymbol{uu}}^{-1}\boldsymbol{A}_{\boldsymbol{u}\psi}\left(\boldsymbol{A}_{\psi\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{uu}}^{-1}\boldsymbol{A}_{\boldsymbol{u}\psi}\right)^{-1}\boldsymbol{A}_{\psi\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{uu}}^{-1} \\
= & \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{uu}}^{12}\left(\boldsymbol{A}_{\boldsymbol{uu}}^{22}\right)^{-1}\boldsymbol{A}_{\boldsymbol{uu}}^{21} & \boldsymbol{A}_{\boldsymbol{uu}}^{12} \\ \boldsymbol{A}_{\boldsymbol{uu}}^{21} & \boldsymbol{A}_{\boldsymbol{uu}}^{22} \end{pmatrix} \\
= & \boldsymbol{A}_{\boldsymbol{uu}}^{-1} - \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{uu}}^{11} - \boldsymbol{A}_{\boldsymbol{uu}}^{12}\left(\boldsymbol{A}_{\boldsymbol{uu}}^{22}\right)^{-1}\boldsymbol{A}_{\boldsymbol{uu}}^{21} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}.
\end{aligned}$$

It can be verified that $\boldsymbol{A}_{\boldsymbol{\theta u}}\boldsymbol{e}_{\boldsymbol{u}} = \{\lambda^*(1-\lambda^*)\}^{-1}\boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}}$ and

$$\left\{\boldsymbol{A}_{\boldsymbol{uu}}^{11} - \boldsymbol{A}_{\boldsymbol{uu}}^{12}\left(\boldsymbol{A}_{\boldsymbol{uu}}^{22}\right)^{-1}\boldsymbol{A}_{\boldsymbol{uu}}^{21}\right\}^{-1} = E_0\left\{\frac{\{\omega(X)-1\}^2}{h(X)}\right\} = \frac{1}{\lambda^*(1-\lambda^*)}\left\{1 - \frac{\boldsymbol{e}_{\boldsymbol{\theta}}^\top\boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}}}{\lambda^*(1-\lambda^*)}\right\}.$$

By the Woodbury matrix identity, the variance matrix $\boldsymbol{J}_{\boldsymbol{\theta}}$ can be simplified as

$$\begin{aligned}
\boldsymbol{J}_{\boldsymbol{\theta}} &= \left\{\boldsymbol{A}_{\boldsymbol{\theta\theta}} + \left\{\frac{\boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}}}{\lambda^*(1-\lambda^*)}\right\}\left[E_0\left\{\frac{\{\omega(X)-1\}^2}{h(X)}\right\}\right]^{-1}\left\{\frac{\boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}}}{\lambda^*(1-\lambda^*)}\right\}^\top\right\}^{-1} \\
&= \boldsymbol{A}_{\boldsymbol{\theta\theta}}^{-1} - \frac{\boldsymbol{e}_{\boldsymbol{\theta}}\boldsymbol{e}_{\boldsymbol{\theta}}^\top}{\lambda^*(1-\lambda^*)}.
\end{aligned}$$

This is the same as the asymptotic variance of $n^{1/2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ shown in Lemma 1 of Qin and Zhang (1997) under Conditions C1–C3.

*Part (b).* For $r > p$, let $\boldsymbol{U}_m, \boldsymbol{V}_m, \boldsymbol{J}_m$ denote the corresponding $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{J}$ matrices obtained by using only the first $m$ estimating equations of $\boldsymbol{g}(\boldsymbol{x};\boldsymbol{\eta})$. With Theorem 2.1, to complete the proof of this part it suffices to show that

$$\boldsymbol{J}_m \geq \boldsymbol{J}_{m-1}.$$

From the definition of the matrix $\boldsymbol{U}$, we notice that $\boldsymbol{U}_m$ has one more column than $\boldsymbol{U}_{m-1}$, and we denote this extra column $u_m$. Then we have $\boldsymbol{U}_m = (\boldsymbol{U}_{m-1}, u_m)$. Following the proof of Corollary 1 of Qin and Lawless (1994), we have

$$\boldsymbol{V}_m^{-1} \geq \begin{pmatrix} \boldsymbol{V}_{m-1}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & 0 \end{pmatrix}. \tag{2.30}$$

Therefore,

$$\boldsymbol{J}_m = \boldsymbol{U}_m \boldsymbol{V}_m^{-1} \boldsymbol{U}_m^\top \geq (\boldsymbol{U}_{m-1}, u_m) \begin{pmatrix} \boldsymbol{V}_{m-1}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & 0 \end{pmatrix} (\boldsymbol{U}_{m-1}, u_m)^\top = \boldsymbol{J}_{m-1}, \tag{2.31}$$

as required. This completes the proof.

**Proof of Theorem 2.2**

Recall that the null hypothesis forms a constraint

$$\mathcal{C}_3 = \{\boldsymbol{\eta} : \boldsymbol{H}(\boldsymbol{\eta}) = \boldsymbol{0}\},$$

and the ELR statistic for testing $H_0 : \boldsymbol{H}(\boldsymbol{\eta}) = 0$ is defined as

$$\begin{aligned} R_n &= 2\left\{\sup_{\boldsymbol{\psi},\boldsymbol{\theta}} \ell_n(\boldsymbol{\psi}, \boldsymbol{\theta}) - \sup_{\boldsymbol{\eta}\in\mathcal{C}_3} \ell_n(\boldsymbol{\psi}, \boldsymbol{\theta})\right\} \\ &= 2\left\{\ell_n(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}) - \ell_n(\check{\boldsymbol{\psi}}, \check{\boldsymbol{\theta}})\right\}, \end{aligned}$$

where

$$(\check{\boldsymbol{\psi}}, \check{\boldsymbol{\theta}}) = \arg\max_{\boldsymbol{\eta}\in\mathcal{C}_3} \ell_n(\boldsymbol{\psi}, \boldsymbol{\theta}).$$

In the following steps, we find the approximations of $\ell_n(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})$ and $\ell_n(\check{\boldsymbol{\psi}}, \check{\boldsymbol{\theta}})$.

We first derive the approximation of $l_n(\boldsymbol{\gamma})$ when $\boldsymbol{\gamma}$ is in the $n^{-1/2}$ neighborhood of its true value $\boldsymbol{\gamma}^*$. Applying the second-order Taylor expansion to $l_n(\boldsymbol{\gamma})$, and using (2.26) and Condition C5, we have

$$\begin{aligned} l_n(\boldsymbol{\gamma}) &= l_n(\boldsymbol{\gamma}^*) + \boldsymbol{S}_n^\top(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*) - \frac{n}{2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)^\top \boldsymbol{A}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*) + o_p(1) \\ &= l_n(\boldsymbol{\gamma}^*) + \begin{pmatrix} \boldsymbol{0} & \boldsymbol{S}_{n\boldsymbol{\theta}}^\top \end{pmatrix} (\boldsymbol{\eta} - \boldsymbol{\eta}^*) + \boldsymbol{S}_{n\boldsymbol{u}}^\top(\boldsymbol{u} - \boldsymbol{u}_0) \\ &\quad - \frac{n}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}^*)^\top \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{pmatrix} (\boldsymbol{\eta} - \boldsymbol{\eta}^*) - n(\boldsymbol{\eta} - \boldsymbol{\eta}^*)^\top \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\psi}\boldsymbol{u}} \\ \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{u}} \end{pmatrix} (\boldsymbol{u} - \boldsymbol{u}^*) \\ &\quad + \frac{n}{2}(\boldsymbol{u} - \boldsymbol{u}^*)^\top \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}(\boldsymbol{u} - \boldsymbol{u}^*) + o_p(1). \end{aligned}$$

Setting the derivative of $l_n(\boldsymbol{\gamma})$ with respect to $\boldsymbol{u}$ equal to zero gives

$$\boldsymbol{u} - \boldsymbol{u}^* = \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}^{-1} \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{u}\psi} & \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{\theta}} \end{pmatrix} (\boldsymbol{\eta} - \boldsymbol{\eta}^*) - \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}^{-1}\left(\frac{1}{n}\boldsymbol{S}_{n\boldsymbol{u}}\right) + o_p(n^{-\frac{1}{2}}).$$

Substituting the approximation of $\boldsymbol{u}-\boldsymbol{u}^*$ into $l_n(\boldsymbol{\gamma})$ leads to an approximation of $\ell_n(\boldsymbol{\psi},\boldsymbol{\theta})$:

$$\begin{aligned}\ell_n(\boldsymbol{\psi},\boldsymbol{\theta}) &= l_n(\boldsymbol{\gamma}^*)+(\boldsymbol{\eta}-\boldsymbol{\eta}^*)^\top\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{S}_n^*-\frac{n}{2}(\boldsymbol{\eta}-\boldsymbol{\eta}^*)^\top\boldsymbol{J}(\boldsymbol{\eta}-\boldsymbol{\eta}^*)\\ &\quad -\frac{1}{2n}\boldsymbol{S}_{n\boldsymbol{u}}^\top\boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}^{-1}\boldsymbol{S}_{n\boldsymbol{u}}+o_p(1).\end{aligned} \tag{2.32}$$

With the approximation of $\hat{\boldsymbol{\eta}}$ in (2.32), we then have

$$\ell_n(\hat{\boldsymbol{\psi}},\hat{\boldsymbol{\theta}}) = l_n(\boldsymbol{\gamma}^*)+\frac{1}{2n}\boldsymbol{S}_n^{*\top}\boldsymbol{V}^{-1}\boldsymbol{U}^\top\boldsymbol{J}^{-1}\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{S}_n^*-\frac{1}{2n}\boldsymbol{S}_{n\boldsymbol{u}}^\top\boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}^{-1}\boldsymbol{S}_{n\boldsymbol{u}}+o_p(1).$$

Next, we find an approximation for $\check{\boldsymbol{\eta}}=(\check{\boldsymbol{\psi}}^\top,\check{\boldsymbol{\theta}}^\top)^\top$. We first define

$$\ell_n^*(\boldsymbol{\psi},\boldsymbol{\theta},\boldsymbol{v})=\ell_n(\boldsymbol{\psi},\boldsymbol{\theta})+n\boldsymbol{v}^\top\boldsymbol{H}(\boldsymbol{\eta}),$$

where $\boldsymbol{v}$ is the Lagrange multiplier. Then $\check{\boldsymbol{\eta}}$ and the corresponding Lagrange multiplier $\check{\boldsymbol{v}}$ satisfy

$$\frac{\partial\ell_n^*(\check{\boldsymbol{\psi}},\check{\boldsymbol{\theta}},\check{\boldsymbol{v}})}{\partial\boldsymbol{\psi}}=\mathbf{0},\quad \frac{\partial\ell_n^*(\check{\boldsymbol{\psi}},\check{\boldsymbol{\theta}},\check{\boldsymbol{v}})}{\partial\boldsymbol{\theta}}=\mathbf{0},\quad \frac{\partial\ell_n^*(\check{\boldsymbol{\psi}},\check{\boldsymbol{\theta}},\check{\boldsymbol{v}})}{\partial\boldsymbol{v}}=\mathbf{0}. \tag{2.33}$$

It is easy to verify that $\check{\boldsymbol{\gamma}}=\boldsymbol{\gamma}^*+O_p(n^{-1/2})$ and $\check{\boldsymbol{v}}=O_p(n^{-1/2})$ (Qin and Lawless, 1995; Qin et al., 2015).

Let $\boldsymbol{h}^*=\partial\boldsymbol{H}(\boldsymbol{\eta}^*)/\partial\boldsymbol{\eta}$. When $\boldsymbol{\eta}$ is in the $n^{-1/2}$ neighborhood of the true value $\boldsymbol{\eta}^*$, we approximate $\boldsymbol{H}(\boldsymbol{\eta})$ with $\boldsymbol{H}(\boldsymbol{\eta})=\boldsymbol{h}^*(\boldsymbol{\eta}-\boldsymbol{\eta}^*)+o_p(n^{-1/2})$. Together with the approximation of $\ell_n(\boldsymbol{\psi},\boldsymbol{\theta})$ in (2.32), we approximate $\ell_n^*(\boldsymbol{\psi},\boldsymbol{\theta},\boldsymbol{v})$ at an $n^{-1/2}$ neighbor of $(\boldsymbol{\psi}_0^\top,\boldsymbol{\theta}_0^\top,\mathbf{0}_{1\times q})^\top$ with

$$\begin{aligned}\ell_n^*(\boldsymbol{\psi},\boldsymbol{\theta},\boldsymbol{v}) &= l_n(\boldsymbol{\gamma}^*)+(\boldsymbol{\eta}-\boldsymbol{\eta}^*)^\top\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{S}_n^*-\frac{n}{2}(\boldsymbol{\eta}-\boldsymbol{\eta}^*)^\top\boldsymbol{J}(\boldsymbol{\eta}-\boldsymbol{\eta}^*)\\ &\quad +n\boldsymbol{v}^\top\boldsymbol{h}^*(\boldsymbol{\eta}-\boldsymbol{\eta}^*)-\frac{1}{2n}\boldsymbol{S}_{n\boldsymbol{u}}^\top\boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}^{-1}\boldsymbol{S}_{n\boldsymbol{u}}+o_p(1).\end{aligned}$$

Applying the first-order Taylor expansion to (2.33), we have

$$\begin{pmatrix}\boldsymbol{J} & -\boldsymbol{h}^{*\top}\\ -\boldsymbol{h}^* & \mathbf{0}\end{pmatrix}\begin{pmatrix}\check{\boldsymbol{\eta}}-\boldsymbol{\eta}^*\\ \check{\boldsymbol{v}}\end{pmatrix}=\frac{1}{n}\begin{pmatrix}\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{S}_n^*\\ \mathbf{0}\end{pmatrix}+o_p(n^{-\frac{1}{2}}).$$

Hence,

$$
\begin{aligned}
& n^{1/2}(\check{\boldsymbol{\eta}}-\boldsymbol{\eta}^*) \\
= \ & (\boldsymbol{I},\mathbf{0})\begin{pmatrix} \boldsymbol{J} & -\boldsymbol{h}^{*\top} \\ -\boldsymbol{h}^{*\top} & \mathbf{0} \end{pmatrix}^{-1}\begin{pmatrix} n^{-1/2}\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{S}_n \\ \mathbf{0} \end{pmatrix} + o_p(1) \\
= \ & \{\boldsymbol{J}^{-1}-\boldsymbol{J}^{-1}\boldsymbol{h}^{*\top}(\boldsymbol{h}^*\boldsymbol{J}^{-1}\boldsymbol{h}^{*\top})^{-1}\boldsymbol{h}^*\boldsymbol{J}^{-1}\}\boldsymbol{U}\boldsymbol{V}^{-1}(n^{-1/2}\boldsymbol{S}_n^*)+o_p(1), \qquad (2.34)
\end{aligned}
$$

where $\boldsymbol{I}$ is the identity matrix with dimension $p+d+1$.

Substituting the expression of $\check{\boldsymbol{\eta}}$ in (2.34) into (2.32) gives

$$
\begin{aligned}
\ell_n(\check{\boldsymbol{\psi}},\check{\boldsymbol{\theta}}) = \ & l_n(\boldsymbol{\gamma}^*)+\frac{1}{2n}\boldsymbol{S}_n^{*\top}\boldsymbol{V}^{-1}\boldsymbol{U}^\top\{\boldsymbol{J}^{-1}-\boldsymbol{J}^{-1}\boldsymbol{h}^{*\top}(\boldsymbol{h}^*\boldsymbol{J}^{-1}\boldsymbol{h}^{*\top})^{-1}\boldsymbol{h}^*\boldsymbol{J}^{-1}\}\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{S}_n^* \\
& -\frac{1}{2n}\boldsymbol{S}_{n\boldsymbol{u}}^\top\boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}^{-1}\boldsymbol{S}_{n\boldsymbol{u}}+o_p(1).
\end{aligned}
$$

Hence, the ELR statistic $R_n$ can be written as

$$
R_n = \frac{1}{n}\boldsymbol{S}_n^{*\top}\boldsymbol{V}^{-1}\boldsymbol{U}^\top\boldsymbol{J}^{-1}\boldsymbol{h}^{*\top}(\boldsymbol{h}^*\boldsymbol{J}^{-1}\boldsymbol{h}^{*\top})^{-1}\boldsymbol{h}^*\boldsymbol{J}^{-1}\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{S}_n^*+o_p(1).
$$

We find that $\boldsymbol{J}^{-1/2}\boldsymbol{h}^{*\top}(\boldsymbol{h}^*\boldsymbol{J}^{-1}\boldsymbol{h}^{*\top})^{-1}\boldsymbol{h}^*\boldsymbol{J}^{-1/2}$ is an idempotent matrix with rank $q$. Further, as $n\to\infty$,

$$
\boldsymbol{J}^{-1/2}\boldsymbol{U}\boldsymbol{V}^{-1}(n^{-1/2}\boldsymbol{S}_n^*)\to N(0,\boldsymbol{I})
$$

in distribution. Therefore, the limiting distribution of $R_n$ is $\chi^2_q$ under $H_0$.

**Proofs of Theorem 2.3 and Corollary 2.2**

We start with the proof of Theorem 2.3. Recall that the ELR statistic for testing the validity of the estimating equations is defined as

$$
W_n = 2\left\{\ell_{nd}(\tilde{\boldsymbol{\theta}})-\ell_n(\hat{\boldsymbol{\psi}},\hat{\boldsymbol{\theta}})\right\}.
$$

We first find an approximation of $\ell_{nd}(\tilde{\boldsymbol{\theta}})$. Applying the second-order Taylor expansion to $\ell_{nd}(\tilde{\boldsymbol{\theta}})$ at the true value $\boldsymbol{\theta}^*$, we have

$$
\ell_{nd}(\tilde{\boldsymbol{\theta}})=\ell_{nd}(\boldsymbol{\theta}^*)+(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)^\top\frac{\partial\ell_{nd}(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}}+\frac{1}{2}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)^\top\frac{\partial^2\ell_{nd}(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)+o_p(1).
$$

The fact that $\boldsymbol{\nu}^*=\mathbf{0}$ implies $\ell_{nd}(\boldsymbol{\theta}^*)=l_n(\boldsymbol{\gamma}^*)$. According to Qin and Zhang (1997), it is easy to verify that

$$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = \frac{1}{n}\boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}\frac{\partial \ell_{nd}(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}} + o_p(n^{-1/2}), \quad \frac{\partial \ell_{nd}(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}} = \boldsymbol{S}_{n\boldsymbol{\theta}}, \text{ and } \frac{1}{n}\frac{\partial^2 \ell_{nd}(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} = -\boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}} + o_p(1).$$

Then

$$\ell_{nd}(\tilde{\boldsymbol{\theta}}) = l_n(\boldsymbol{\gamma}^*) + \frac{1}{2n}\boldsymbol{S}_{n\boldsymbol{\theta}}^\top \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}\boldsymbol{S}_{n\boldsymbol{\theta}} + o_p(1).$$

Hence, the ELR statistic can be written as

$$\begin{aligned} W_n &= 2\left\{\ell_{nd}(\tilde{\boldsymbol{\theta}}) - \ell_n(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})\right\} \\ &= \frac{1}{n}\boldsymbol{S}_{n\boldsymbol{\theta}}^\top \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}\boldsymbol{S}_{n\boldsymbol{\theta}} + \frac{1}{n}\boldsymbol{S}_{n\boldsymbol{u}}^\top \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}^{-1}\boldsymbol{S}_{n\boldsymbol{u}} - \frac{1}{n}\boldsymbol{S}_n^{*\top}\boldsymbol{V}^{-1}\boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{S}_n^* \\ &= \frac{1}{n}\boldsymbol{S}_n^{*\top}\boldsymbol{V}^{-1}(\boldsymbol{V} - \boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U})\boldsymbol{V}^{-1}\boldsymbol{S}_n^* + o_p(1). \end{aligned} \qquad (2.35)$$

Since $\boldsymbol{V}$ is a positive-definite matrix, we define an inner product on the vector space $\mathbb{R}^{2+d+r}$ as $< \boldsymbol{a}, \boldsymbol{b} >_{\boldsymbol{V}^{-1}} = \boldsymbol{a}^\top \boldsymbol{V}^{-1}\boldsymbol{b}$ for any vector $\boldsymbol{a}, \boldsymbol{b}$ in the vector space. Recall that

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta}\boldsymbol{\theta}}\boldsymbol{e}_{\boldsymbol{\theta}} \\ -\lambda^*(1-\lambda^*)\boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix}.$$

The vector $\boldsymbol{C}$ and each row in $\boldsymbol{U}$ are linearly independent in the inner product space because $\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{C} = \boldsymbol{0}$. Let $\mathcal{V}$ be the inner product space spanned by the vector $\boldsymbol{C}$ and each row in $\boldsymbol{U}$. Then there exists an orthogonal complement $\mathcal{B}$ of the subspace $\mathcal{V}$ with the dimension $r - p$. Let the columns of $\boldsymbol{C}^*$ be the basis of the orthogonal complement $\mathcal{B}$. Then $\boldsymbol{C}^*$ satisfies $\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}(\boldsymbol{C}, \boldsymbol{U}^\top) = \boldsymbol{0}$. Define $\mathcal{M}^\top = (\boldsymbol{C}^*, \boldsymbol{C}, \boldsymbol{U}^\top)$, which satisfies

$$\mathcal{M}\boldsymbol{V}^{-1}\mathcal{M}^\top = \begin{pmatrix} \boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}\boldsymbol{C}^* & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}^\top \boldsymbol{V}^{-1}\boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J} \end{pmatrix}.$$

With the above construction, $\mathcal{M}$ is a full rank matrix and can be inverted. We can write the inverse of $\mathcal{M}\boldsymbol{V}^{-1}\mathcal{M}^\top$ as

$$(\mathcal{M}^\top)^{-1}\boldsymbol{V}\mathcal{M}^{-1} = \begin{pmatrix} (\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}\boldsymbol{C}^*)^{-1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & (\boldsymbol{C}^\top \boldsymbol{V}^{-1}\boldsymbol{C})^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J}^{-1} \end{pmatrix}.$$

Then

$$\begin{aligned}\boldsymbol{V} &= \mathcal{M}^\top(\mathcal{M}^\top)^{-1}\boldsymbol{V}\mathcal{M}^{-1}\mathcal{M}\\ &= \boldsymbol{C}^*(\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}\boldsymbol{C}^*)^{-1}\boldsymbol{C}^{*\top} + \boldsymbol{C}(\boldsymbol{C}^\top\boldsymbol{V}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}^\top + \boldsymbol{U}^\top\boldsymbol{J}^{-1}\boldsymbol{U}.\end{aligned}$$

Note that

$$\begin{aligned}\boldsymbol{C}^\top\boldsymbol{V}^{-1}\boldsymbol{S}_n^* &= \boldsymbol{e}_{\boldsymbol{\theta}}^\top\boldsymbol{S}_{n\boldsymbol{\theta}} - \lambda^*(1-\lambda^*)\boldsymbol{e}_{\boldsymbol{u}}^\top\boldsymbol{S}_{n\boldsymbol{u}}\\ &= n_1 - \sum_{i=0}^{1}\sum_{j=1}^{n_i} h_1(X_{ij}) + \lambda^*(1-\lambda^*)\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\omega(X_{ij})-1}{h(X_{ij})}\\ &= 0.\end{aligned}$$

This helps to simplify $W_n$ as

$$W_n = \frac{1}{n}\boldsymbol{S}_n^{*\top}\boldsymbol{V}^{-1}\boldsymbol{C}^*(\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}\boldsymbol{C}^*)^{-1}\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}\boldsymbol{S}_n^* + o_p(1).$$

According to Lemma 2.2, we have

$$Var\left(n^{-1/2}\boldsymbol{S}_n^*\right) = \boldsymbol{V} - \frac{1}{\lambda^*(1-\lambda^*)}\boldsymbol{C}\boldsymbol{C}^\top.$$

Together with $\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}\boldsymbol{C} = \boldsymbol{0}$ and the fact that $\boldsymbol{V}^{-1/2}\{\boldsymbol{C}^*(\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}\boldsymbol{C}^*)^{-1}\boldsymbol{C}^{*\top}\}\boldsymbol{V}^{-1/2}$ is idempotent with rank $r-p$, we have

$$\begin{aligned}&\{\boldsymbol{V}^{-1/2}\boldsymbol{C}^*(\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}\boldsymbol{C}^*)^{-1}\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}\}Var\left(n^{-1/2}\boldsymbol{S}_n^*\right)\{\boldsymbol{V}^{-1}\boldsymbol{C}^*(\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}\boldsymbol{C}^*)^{-1}\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1/2}\}\\ &= \boldsymbol{V}^{-1/2}\boldsymbol{C}^*(\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1/}\boldsymbol{C}^*)^{-1}\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1/2}.\end{aligned}$$

Therefore, $W_n$ asymptotically follows $\chi^2_{r-p}$ under $H_0$ as $n\to\infty$.

We now prove Corollary 2.2. Let $\boldsymbol{S}_{n1}^*$ be the first $d+r-m+2$ elements of $\boldsymbol{S}_n^{*\top}$, $\boldsymbol{U}_1$ be the first $r-m$ columns of $\boldsymbol{U}$, $\boldsymbol{V}_1$ be the upper $(d+r-m+2)\times(d+r-m+2)$ matrix of $\boldsymbol{V}$, and $\boldsymbol{J}_1 = \boldsymbol{U}_1\boldsymbol{V}_1^{-1}\boldsymbol{U}_1^\top$. Further, let $\ell_{n1}(\boldsymbol{\psi},\boldsymbol{\theta})$ be the profile empirical log-likelihood of $(\boldsymbol{\psi},\boldsymbol{\theta})$ using only $\boldsymbol{g}_1(\boldsymbol{x};\boldsymbol{\eta})$ and

$$(\hat{\boldsymbol{\psi}}^*, \hat{\boldsymbol{\theta}}^*) = \arg\max_{\boldsymbol{\psi},\boldsymbol{\theta}} \ell_{n1}(\boldsymbol{\psi},\boldsymbol{\theta}).$$

Following the techniques used to obtain (2.35), we have

$$2\left\{\ell_{nd}(\tilde{\boldsymbol{\theta}}) - \ell_{n1}(\hat{\boldsymbol{\psi}}^*, \hat{\boldsymbol{\theta}}^*)\right\} = \boldsymbol{S}_{n1}^{*\top}\boldsymbol{V}_1^{-1}(\boldsymbol{V}_1 - \boldsymbol{U}_1^\top\boldsymbol{J}_1^{-1}\boldsymbol{U}_1)\boldsymbol{V}_1^{-1}\boldsymbol{S}_{n1}^* + o_p(1).$$

Then, the ELR statistic $W_n^*$ has the following approximation:

$$\begin{aligned}
W_n^* &= 2\{\ell_{n1}(\hat{\boldsymbol{\psi}}^*, \hat{\boldsymbol{\theta}}^*) - \ell_n(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})\} \\
&= 2\left\{\ell_{nd}(\tilde{\boldsymbol{\theta}}) - \ell_n(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})\right\} - 2\left\{\ell_{nd}(\tilde{\boldsymbol{\theta}}) - \ell_{n1}(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\beta}}^*)\right\} \\
&= \frac{1}{n}\left[\boldsymbol{S}_n^{*\top}\boldsymbol{V}^{-1}(\boldsymbol{V} - \boldsymbol{U}^\top\boldsymbol{J}^{-1}\boldsymbol{U})\boldsymbol{V}^{-1}\boldsymbol{S}_n^* - \boldsymbol{S}_{n1}^{*\top}\boldsymbol{V}_1^{-1}(\boldsymbol{V}_1 - \boldsymbol{U}_1^\top\boldsymbol{J}_1^{-1}\boldsymbol{U}_1)\boldsymbol{V}_1^{-1}\boldsymbol{S}_{n1}^*\right] + o_p(1).
\end{aligned}$$

With the technique used to prove Corollary 2.1, we have

$$\boldsymbol{V}^{-1}(\boldsymbol{V} - \boldsymbol{U}^\top\boldsymbol{J}^{-1}\boldsymbol{U})\boldsymbol{V}^{-1} \geq \begin{pmatrix} \boldsymbol{V}_1^{-1}\{\boldsymbol{V}_1 - \boldsymbol{U}_1^\top\boldsymbol{J}_1^{-1}\boldsymbol{U}_1\}\boldsymbol{V}_1^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}.$$

Then

$$\frac{1}{n}\left[\boldsymbol{S}_n^{*\top}\boldsymbol{V}^{-1}(\boldsymbol{V} - \boldsymbol{U}^\top\boldsymbol{J}^{-1}\boldsymbol{U})\boldsymbol{V}^{-1}\boldsymbol{S}_n^* - \boldsymbol{S}_{n1}^{*\top}\boldsymbol{V}_1^{-1}(\boldsymbol{V}_1 - \boldsymbol{U}_1^\top\boldsymbol{J}_1^{-1}\boldsymbol{U}_1)\boldsymbol{V}_1^{-1}\boldsymbol{S}_{n1}^*\right] \geq 0.$$

Recall that as $n \to \infty$,

$$\frac{1}{n}\boldsymbol{S}_n^{*\top}\boldsymbol{V}^{-1}(\boldsymbol{V} - \boldsymbol{U}^\top\boldsymbol{J}^{-1}\boldsymbol{U})\boldsymbol{V}^{-1}\boldsymbol{S}_n^* \to \chi^2_{r-p}$$

in distribution. We can similarly prove that as $n \to \infty$,

$$\frac{1}{n}\boldsymbol{S}_{n1}^{*\top}\boldsymbol{V}_1^{-1}(\boldsymbol{V}_1 - \boldsymbol{U}_1^\top\boldsymbol{J}_1^{-1}\boldsymbol{U}_1)\boldsymbol{V}_1^{-1}\boldsymbol{S}_{n1}^* \to \chi^2_{r-m-p}$$

in distribution.

By the arguments in Qin and Lawless (1994), we conclude that $W_n^* \to \chi^2_{(r-p)-(r-m-p)} = \chi^2_m$ in distribution as $n \to \infty$.

**Proof of Theorem 2.4**

For (a): We start with some preparation. For any $x$ in the support of $F_0$, let

$$F_0(x,\boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{I(X_{ij}\le x)}{1+\lambda\left\{\omega(X_{ij};\boldsymbol{\theta})-1\right\}+\boldsymbol{\nu}^\top\boldsymbol{g}(X_{ij};\boldsymbol{\psi},\boldsymbol{\theta})},$$
$$F_1(x,\boldsymbol{\gamma}) = \frac{1}{n}\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\omega(X_{ij};\boldsymbol{\theta})I(X_{ij}\le x)}{1+\lambda\left\{\omega(X_{ij};\boldsymbol{\theta})-1\right\}+\boldsymbol{\nu}^\top\boldsymbol{g}(X_{ij};\boldsymbol{\psi},\boldsymbol{\theta})}.$$

Then

$$\hat{F}_0(x) = F_0(x,\hat{\boldsymbol{\gamma}}),\quad F_0(x,\boldsymbol{\gamma}^*) = \frac{1}{n}\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{I(X_{ij}\le x)}{h(X_{ij})},$$
$$\hat{F}_1(x) = F_1(x,\hat{\boldsymbol{\gamma}}),\quad F_1(x,\boldsymbol{\gamma}^*) = \frac{1}{n}\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\omega(X_{ij})I(X_{ij}\le x)}{h(X_{ij})}.$$

Next, we explore the properties of the first derivatives of $F_0(x,\boldsymbol{\gamma})$ and $F_1(x,\boldsymbol{\gamma})$ at the true value $\boldsymbol{\gamma}^*$. Define

$$\frac{\partial F_0(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}} = \begin{pmatrix}\frac{\partial F_0(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\psi}}\\ \frac{\partial F_0(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\theta}}\\ \frac{\partial F_0(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{u}}\end{pmatrix},\quad \frac{\partial F_1(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}} = \begin{pmatrix}\frac{\partial F_1(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\psi}}\\ \frac{\partial F_1(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\theta}}\\ \frac{\partial F_1(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{u}}\end{pmatrix},$$

where

$$\frac{\partial F_0(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\psi}} = \frac{\partial F_1(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\psi}} = \boldsymbol{0},$$
$$\frac{\partial F_0(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\theta}} = -\frac{1}{n}\sum_{i=0}^{1}\sum_{j=1}^{n_i}h_1(X_{ij})h(X_{ij})\boldsymbol{Q}(X_{ij})I(X_{ij}\le x),$$
$$\frac{\partial F_0(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{u}} = -\frac{1}{n}\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\boldsymbol{G}(X_{ij})}{\{h(X_{ij})\}^2}I(X_{ij}\le x),$$
$$\frac{\partial F_1(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\theta}} = \frac{1}{n}\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\omega(X_{ij})}{h(X_{ij})}h_0(X_{ij})\boldsymbol{Q}(X_{ij})I(X_{ij}\le x),$$

$$\frac{\partial F_1(x,\boldsymbol{\gamma}^*)}{\partial \boldsymbol{u}} = -\frac{1}{n}\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\omega(X_{ij})}{\{h(X_{ij})\}^2}\boldsymbol{G}(X_{ij})I(X_{ij}\le x).$$

Applying Lemma 2.1, we have the following results for $E\left\{\frac{\partial F_0(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\right\}$ and $E\left\{\frac{\partial F_1(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\right\}$.

**Lemma 2.3.** *With the form of $\partial F_0(x,\boldsymbol{\gamma}^*)/\partial\boldsymbol{\gamma}$ and $\partial F_1(x,\boldsymbol{\gamma}^*)/\partial\boldsymbol{\gamma}$ defined above, we have*

$$-E\left\{\frac{\partial F_0(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\right\} = \boldsymbol{B}_0(x) = \begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{B}_{0\boldsymbol{\theta}}(x)\\ \boldsymbol{B}_{0\boldsymbol{u}}(x)\end{pmatrix} = \begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{B}_0^*(x)\end{pmatrix},$$
$$-E\left\{\frac{\partial F_1(x,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\right\} = \boldsymbol{B}_1(x) = \begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{B}_{1\boldsymbol{\theta}}(x)\\ \boldsymbol{B}_{1\boldsymbol{u}}(x)\end{pmatrix} = \begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{B}_1^*(x)\end{pmatrix},$$

*where*

$$\boldsymbol{B}_{0\boldsymbol{\theta}}(x) = E_0\left\{h_1(X)\boldsymbol{Q}(X)I(X\le x)\right\},\quad \boldsymbol{B}_{0\boldsymbol{u}}(x) = E_0\left\{\frac{\boldsymbol{G}(X)}{h(X)}I(X\le x)\right\},$$
$$\boldsymbol{B}_{1\boldsymbol{\theta}}(x) = \frac{\lambda^*-1}{\lambda^*}E_0\left\{h_1(X)\boldsymbol{Q}(X)I(X\le x)\right\},\quad \boldsymbol{B}_{1\boldsymbol{u}}(x) = E_0\left\{\frac{\omega(X)\boldsymbol{G}(X)}{h(X)}I(X\le x)\right\}.$$

We now move to the joint asymptotic normality of $\hat{F}_l(x)$ and $\hat{F}_s(y)$. We first find an approximation for $\hat{F}_l(x)$ for $l=0$ and 1. Applying the first-order Taylor expansion to $\hat{F}_l(x)$ and using the results in Lemma 2.3, we have

$$\begin{aligned}\hat{F}_l(x) &= F_l(x,\boldsymbol{\gamma}^*) - \boldsymbol{B}_l^*(x)^\top(\hat{\boldsymbol{\gamma}}^*-\boldsymbol{\gamma}^*) + o_p(n^{-1/2})\\ &= F_l(x,\boldsymbol{\gamma}^*) - (\boldsymbol{0},\boldsymbol{B}_{l\boldsymbol{\theta}}(x)^\top)(\hat{\boldsymbol{\eta}}^*-\boldsymbol{\eta}^*) - \boldsymbol{B}_{0\boldsymbol{u}}(x)^\top(\hat{\boldsymbol{u}}-\boldsymbol{u}^*) + o_p(n^{-1/2}).\end{aligned}$$

Using the relationship in (2.28) and the definitions of the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in (2.29), we have

$$\begin{aligned}\hat{F}_l(x) &= F_l(x,\boldsymbol{\gamma}^*) - \boldsymbol{B}_l^*(x)^\top\boldsymbol{V}^{-1}\boldsymbol{U}^\top(\hat{\boldsymbol{\eta}}^*-\boldsymbol{\eta}^*) + \frac{1}{n}\boldsymbol{B}_{l\boldsymbol{u}}(x)^\top\boldsymbol{A}_{\boldsymbol{uu}}^{-1}\boldsymbol{S}_{n\boldsymbol{u}} + o_p(n^{-1/2})\\ &= F_l(x,\boldsymbol{\gamma}^*) - \boldsymbol{B}_l^*(x)^\top\left\{\boldsymbol{V}^{-1}\boldsymbol{U}^\top(\hat{\boldsymbol{\eta}}^*-\boldsymbol{\eta}^*) - \begin{pmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_{\boldsymbol{uu}}^{-1}\end{pmatrix}\left(\frac{1}{n}\boldsymbol{S}_n^*\right)\right\} + o_p(n^{-1/2}).\end{aligned}$$

Recall that $\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}^* = \boldsymbol{J}^{-1}\boldsymbol{U}\boldsymbol{V}^{-1}(n^{-1}\boldsymbol{S}_n^*) + o_p(n^{-1/2})$. The approximation of $\hat{F}_l(x)$ is then given by

$$\hat{F}_l(x) = F_l(x, \boldsymbol{\gamma}^*) - \frac{1}{n}\boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{S}_n^* + o_p(n^{-1/2})$$

with

$$\boldsymbol{W} = \boldsymbol{V}^{-1}\boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U}\boldsymbol{V}^{-1} - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_{\boldsymbol{uu}}^{-1} \end{pmatrix}.$$

Note that $F_l(x) = E_0\{F_l(x, \boldsymbol{\gamma}^*)\}$. Then

$$n^{1/2}\{\hat{F}_l(x) - F_l(x)\} = n^{1/2}\{F_l(x, \boldsymbol{\gamma}^*) - F_l(x)\} - n^{-1/2}\boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{S}_n^* + o_p(1).$$

The two leading terms are summations of independent random variables and both have mean zero. Hence, as $n \to \infty$,

$$\sqrt{n}\begin{pmatrix} \hat{F}_l(x) - F_l(x) \\ \hat{F}_s(y) - F_s(y) \end{pmatrix} \to N\Big(\mathbf{0}, \boldsymbol{\Sigma}_{ls}(x, y)\Big),$$

where

$$\boldsymbol{\Sigma}_{ls}(x, y) = \begin{pmatrix} \sigma_{ll}(x, x) & \sigma_{ls}(x, y) \\ \sigma_{sl}(y, x) & \sigma_{ss}(y, y) \end{pmatrix}.$$

To complete the proof of (a), we need to argue that $\boldsymbol{\Sigma}_{ls}(x, y)$ has the form claimed in the lemma. According to the expression of $\hat{F}_l(x) - F_l(x)$, we have

$$\begin{aligned}
\sigma_{ll}(x, x) &= nVar\{F_l(x, \boldsymbol{\gamma}^*)\} + n^{-1}Var(\boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{S}_n^*) \\
&\quad -2Cov\left\{F_l(x, \boldsymbol{\gamma}^*), \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{S}_n^*\right\}; \\
\sigma_{ss}(y, y) &= nVar\{F_s(y, \boldsymbol{\gamma}^*)\} + n^{-1}Var(\boldsymbol{B}_s^*(y)^\top \boldsymbol{W}\boldsymbol{S}_n^*) \\
&\quad -2Cov\left\{F_s(y, \boldsymbol{\gamma}^*), \boldsymbol{B}_s^*(y)^\top \boldsymbol{W}\boldsymbol{S}_n^*\right\}; \\
\sigma_{ls}(x, y) &= nCov\{F_l(x, \boldsymbol{\gamma}^*), F_s(y, \boldsymbol{\gamma}^*)\} - Cov\left\{F_l(x, \boldsymbol{\gamma}^*), \boldsymbol{B}_s^*(y)^\top \boldsymbol{W}\boldsymbol{S}_n^*\right\} \\
&\quad -Cov\left\{F_s(y, \boldsymbol{\gamma}^*), \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{S}_n^*\right\} + \boldsymbol{B}_l^*(x)^\top\{n^{-1}Var(\boldsymbol{W}\boldsymbol{S}_n^*)\}\boldsymbol{B}_s^*(y); \\
\sigma_{sl}(y, x) &= \sigma_{ls}(x, y).
\end{aligned}$$

Next, we calculate the covariances and variances appearing above. We start with the covariance and variance related to $F_l(x, \boldsymbol{\gamma}^*)$ and $F_s(y, \boldsymbol{\gamma}^*)$. Let $x \wedge y = \min\{x, y\}$. Using

Lemma 2.1, we have

$$\begin{aligned}
& nCov\left\{F_0(x,\boldsymbol{\gamma}^*),\ F_0(y,\boldsymbol{\gamma}^*)\right\} \\
= \ & (1-\lambda^*)Cov\left\{\frac{I(X_{01}\le x)}{h(X_{01})},\ \frac{I(X_{01}\le y)}{h(X_{01})}\right\} + \lambda^* Cov\left\{\frac{I(X_{11}\le x)}{h(X_{11})},\ \frac{I(X_{11}\le y)}{h(X_{11})}\right\} \\
= \ & E_0\left\{\frac{I(X\le x\wedge y)}{h(X)}\right\} - (1-\lambda^*)E_0\left\{\frac{I(X\le x)}{h(X)}\right\}E_0\left\{\frac{I(X\le y)}{h(X)}\right\} \\
& -\lambda^* E_0\left\{\frac{\omega(X)I(X\le x)}{h(X)}\right\}E_0\left\{\frac{\omega(X)I(X\le y)}{h(X)}\right\}.
\end{aligned}$$

After some algebra, we have that for any $x$ in the support of $F_0$,

$$\begin{aligned}
\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}} &= E_0\left\{\frac{\omega(X)I(X\le x)}{h(X)}\right\} - E_0\left\{\frac{I(X\le x)}{h(X)}\right\}, \\
F_0(x) &= E_0\left\{\frac{I(X\le x)}{h(X)}\right\} + \lambda^*\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}.
\end{aligned}$$

Then the covariance $nCov\left\{F_0(x,\boldsymbol{\gamma}^*),\ F_0(y,\boldsymbol{\gamma}^*)\right\}$ is simplified as

$$\begin{aligned}
& nCov\left\{F_0(x,\boldsymbol{\gamma}^*),\ F_0(y,\boldsymbol{\gamma}^*)\right\} \\
= \ & E_0\left\{\frac{I(X\le x\wedge y)}{h(X)}\right\} - \lambda^*\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{0\boldsymbol{u}}(y) - \lambda^*\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}} E_0\left\{\frac{I(X\le y)}{h(X)}\right\} \\
& -\lambda^* E_0\left\{\frac{I(X\le x)}{h(X)}\right\}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{0\boldsymbol{u}}(y) - E_0\left\{\frac{I(X\le x)}{h(X)}\right\}E_0\left\{\frac{I(X\le y)}{h(X)}\right\} \\
= \ & E_0\left\{\frac{I(X\le x\wedge y)}{h(X)}\right\} - \lambda^*\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{0\boldsymbol{u}}(y) - \lambda^*\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\left[F_0(y) - \lambda^*\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{0\boldsymbol{u}}(y)\right] \\
& -E_0\left\{\frac{I(X\le x)}{h(X)}\right\}F_0(y) \\
= \ & E_0\left\{\frac{I(X\le x\wedge y)}{h(X)}\right\} - F_0(x)F_0(y) - \lambda^*(1-\lambda^*)\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{0\boldsymbol{u}}(y).
\end{aligned}$$

The covariances $nCov\left\{F_0(x,\boldsymbol{\gamma}^*),\ F_0(y,\boldsymbol{\gamma}^*)\right\}$ and $nCov\left\{F_0(x,\boldsymbol{\gamma}^*),\ F_1(y,\boldsymbol{\gamma}^*)\right\}$ can be found in a similar manner. For $nCov\left\{F_1(x,\boldsymbol{\gamma}^*),\ F_1(y,\boldsymbol{\gamma}^*)\right\}$, we have

$$
\begin{aligned}
& nCov\left\{F_1(x,\boldsymbol{\gamma}^*),\ F_1(y,\boldsymbol{\gamma}^*)\right\} \\
= \;& E_0\left\{\frac{\omega^2(X)I(X\le x\wedge y)}{h(X)}\right\} - (1-\lambda^*)E_0\left\{\frac{\omega(X)I(X\le x)}{h(X)}\right\}E_0\left\{\frac{\omega(X)I(X\le y)}{h(X)}\right\} \\
& -\lambda^* E_0\left\{\frac{\omega^2(X)I(X\le x)}{h(X)}\right\}E_0\left\{\frac{\omega^2(X)I(X\le y)}{h(X)}\right\} \\
= \;& E_0\left\{\frac{\omega^2(X)I(X\le x\wedge y)}{h(X)}\right\} - F_1(x)F_1(y) - \lambda^*(1-\lambda^*)\boldsymbol{B}_{1\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{1\boldsymbol{u}}(y)
\end{aligned}
$$

and

$$
\begin{aligned}
& nCov\left\{F_0(x,\boldsymbol{\gamma}^*),\ F_1(y,\boldsymbol{\gamma}^*)\right\} \\
= \;& E_0\left\{\frac{\omega(X)I(X\le x\wedge y)}{h(X)}\right\} - (1-\lambda^*)E_0\left\{\frac{I(X\le x)}{h(X)}\right\}E_0\left\{\frac{\omega(X)I(X\le y)}{h(X)}\right\} \\
& -\lambda^* E_0\left\{\frac{\omega(X)I(X\le x)}{h(X)}\right\}E_0\left\{\frac{\omega^2(X)I(X\le y)}{h(X)}\right\} \\
= \;& E_0\left\{\frac{\omega(X)I(X\le x\wedge y)}{h(X)}\right\} - F_0(x)F_1(y) - \lambda^*(1-\lambda^*)\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{1\boldsymbol{u}}(y).
\end{aligned}
$$

In summary, for any $l,s\in\{0,1\}$, we get

$$
\begin{aligned}
nCov\left\{F_l(x,\boldsymbol{\gamma}^*),\ F_s(y,\boldsymbol{\gamma}^*)\right\} \;=\;& E_0\left\{\frac{\omega^{l+s}(X)I(X\le x\wedge y)}{h(X)}\right\} - F_l(x)F_s(y) \\
& -\lambda^*(1-\lambda^*)\boldsymbol{B}_{l\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{s\boldsymbol{u}}(y). \qquad (2.36)
\end{aligned}
$$

Next, we consider the cross-terms with $\boldsymbol{S}_n^*$. We present the calculation of covariance between $F_0(x,\boldsymbol{\gamma}^*)$ and $\boldsymbol{S}_n^*$ as an illustration. Using Lemma 2.1, we get

$$
\begin{aligned}
& Cov\left\{F_0(x,\boldsymbol{\gamma}^*),\boldsymbol{S}_{n\boldsymbol{\theta}}\right\} \\
= \;& \frac{1}{n}Cov\left\{\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{I(X_{ij}\le x)}{h(X_{ij})},\ \ \sum_{j=1}^{n_1}h_0(X_{1j})\boldsymbol{Q}(X_{1j})^\top - \sum_{j=1}^{n_0}h_1(X_{0j})\boldsymbol{Q}(X_{0j})^\top\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \lambda^* Cov\left\{\frac{I(X_{11} \le x)}{h(X_{11})},\ h_0(X_{11})\boldsymbol{Q}(X_{11})^\top\right\} \\
&\quad -(1-\lambda^*)Cov\left\{\frac{I(X_{01} \le x)}{h(X_{01})},\ h_1(X_{01})\boldsymbol{Q}(X_{01})^\top\right\} \\
&= \left[E_0\{h_0(X)I(X \le x)\} - \frac{1-\lambda^*}{\lambda^*}E_0\{h_1(X)I(X \le x)\}\right] E_0\{h_1(X)\boldsymbol{Q}(X)^\top\}.
\end{aligned}
$$

It can be checked that

$$
\begin{aligned}
&E_0\{h_1(X)\boldsymbol{Q}(X)\} = \frac{1}{1-\lambda^*}\boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}}, \\
&E_0\{h_0(X)I(X \le x)\} - \frac{1-\lambda^*}{\lambda^*}E_0\{h_1(X)I(X \le x)\} = -(1-\lambda^*)\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}.
\end{aligned}
$$

Then we have

$$
Cov\{F_0(x,\boldsymbol{\gamma}^*), \boldsymbol{S}_{n\boldsymbol{\theta}}\} = -\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}(\boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}})^\top.
$$

Similarly,

$$
\begin{aligned}
&Cov\{F_0(x,\boldsymbol{\gamma}^*), \boldsymbol{S}_{n\boldsymbol{u}}\} \\
&= -\frac{1}{n}Cov\left\{\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{I(X_{ij} \le x)}{h(X_{ij})},\quad \sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{\boldsymbol{G}(X_{ij})^\top}{h(X_{ij})}\right\} \\
&= -\lambda^* Cov\left\{\frac{I(X_{11} \le x)}{h(X_{11})},\ \frac{\boldsymbol{G}(X_{11})^\top}{h(X_{11})}\right\} - (1-\lambda^*)Cov\left\{\frac{I(X_{01} \le x)}{h(X_{01})},\ \frac{\boldsymbol{G}(X_{01})^\top}{h(X_{01})}\right\} \\
&= -E_0\left\{\frac{I(X \le x)\boldsymbol{G}(X)^\top}{h(X)}\right\} \\
&\quad +\frac{1}{1-\lambda^*}\left[E_0\{h_0(X)I(X \le x)\} - \frac{1-\lambda^*}{\lambda^*}E_0\{h_1(X)I(X \le x)\}\right] E_0\{h_0(X)\boldsymbol{G}(X)^\top\} \\
&= -E_0\left\{\frac{I(X \le x)\boldsymbol{G}(X)^\top}{h(X)}\right\} - \boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}} \cdot E_0\{h_0(X)\boldsymbol{G}(X)^\top\} \\
&= -\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top + \lambda^*(1-\lambda^*)\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}(\boldsymbol{A}_{\boldsymbol{uu}}\boldsymbol{e}_{\boldsymbol{u}})^\top,
\end{aligned}
$$

where in the last step we used the facts that

$$
\boldsymbol{B}_{0\boldsymbol{u}}(x) = E_0\left\{\frac{I(X \le x)\boldsymbol{G}(X)}{h(X)}\right\} \quad \text{and} \quad E_0\{h_0(X)\boldsymbol{G}(X)\} = -\lambda^*(1-\lambda^*)\boldsymbol{A}_{\boldsymbol{uu}}\boldsymbol{e}_{\boldsymbol{u}}.
$$

Recall that

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}} \\ -\lambda^*(1-\lambda^*)\boldsymbol{A}_{\boldsymbol{uu}}\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix}.$$

Hence,

$$Cov\left\{F_0(x,\boldsymbol{\gamma}^*), \boldsymbol{S}_n^*\right\} = -\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{B}_{0\boldsymbol{u}}(x) \end{pmatrix}^\top - \boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{C}^\top.$$

The covariance between $F_1(x,\boldsymbol{\gamma}^*)$ and $\boldsymbol{S}_n^*$ can be found in a similar manner; the details are omitted. We conclude that for any $x$ in the support of $F_0$,

$$Cov\left\{F_l(x,\boldsymbol{\gamma}^*), \boldsymbol{S}_n^*\right\} = -\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{B}_{l\boldsymbol{u}}(x) \end{pmatrix}^\top - \boldsymbol{B}_{l\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{C}^\top, \quad l \in \{0,1\}.$$

We now return to the form of $\boldsymbol{\Sigma}(x,y)$. Recall that

$$n^{-1}Var(\boldsymbol{S}_n) = \boldsymbol{\Gamma} = \boldsymbol{V} - \frac{1}{\lambda^*(1-\lambda^*)}\boldsymbol{C}\boldsymbol{C}^\top \text{ and } \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{C} = \boldsymbol{0}.$$

This leads to

$$\begin{aligned} \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{\Gamma} &= \boldsymbol{B}_l^*(x)^\top \boldsymbol{V}^{-1}\boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U} - \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{B}_{l\boldsymbol{u}}(x) \end{pmatrix}^\top - \boldsymbol{B}_{l\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{C}^\top \\ &= \boldsymbol{B}_l^*(x)^\top \boldsymbol{V}^{-1}\boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U} + Cov\left\{F_l(x,\boldsymbol{\gamma}^*), \boldsymbol{S}_n^*\right\}. \end{aligned}$$

Consequently, for $l = 0,1$, the summation of the last two terms in $\sigma_{ll}(x,x)$ is

$$\begin{aligned} & n^{-1}Var(\boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{S}_n^*) - 2Cov\left\{F_l(x,\boldsymbol{\gamma}^*), \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{S}_n^*\right\} \\ = & \left[\boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{\Gamma} - 2Cov\left\{F_l(x,\boldsymbol{\gamma}^*), \boldsymbol{S}_n^*\right\}\right] \boldsymbol{W}\boldsymbol{B}_l^*(x) \\ = & \left[\boldsymbol{B}_l^*(x)^\top \boldsymbol{V}^{-1}\boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U} + \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{B}_{l\boldsymbol{u}}(x) \end{pmatrix}^\top + \boldsymbol{B}_{l\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{C}^\top\right] \boldsymbol{W}\boldsymbol{B}_l^*(x) \\ = & \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{B}_l^*(x) + \lambda^*(1-\lambda^*)\boldsymbol{B}_{l\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{l\boldsymbol{u}}(x). \end{aligned} \tag{2.37}$$

Combining (2.36) and (2.37) leads to

$$\sigma_{ll}(x,x) = E_0\left\{\frac{\omega^{2l}(X)I(X \le x)}{h(X)}\right\} - F_l(x)^2 + \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{B}_l^*(x). \tag{2.38}$$

Using similar steps to derive (2.37), we find that the summation of the last three terms in $\sigma_{ls}(x,y)$ is

$$\begin{aligned}
& \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{\Gamma}\boldsymbol{W}\boldsymbol{B}_s^*(y) - Cov\left\{F_l(x,\boldsymbol{\gamma}^*), \boldsymbol{S}_n^*\right\}\boldsymbol{W}\boldsymbol{B}_s^*(y) - \boldsymbol{B}_l^*(x)^\top \boldsymbol{W} Cov\left\{\boldsymbol{S}_n^*, F_s(y,\boldsymbol{\gamma}^*)\right\} \\
= \; & \boldsymbol{B}_l^*(x)^\top \boldsymbol{V}^{-1}\boldsymbol{U}\boldsymbol{J}^{-1}\boldsymbol{U}^\top \boldsymbol{W}\boldsymbol{B}_s^*(y) - \boldsymbol{B}_l^*(x)^\top \boldsymbol{W} Cov\left\{\boldsymbol{S}_n^*, F_s(y,\boldsymbol{\gamma}^*)\right\} \\
= \; & \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{B}_s^*(y) + \lambda^*(1-\lambda^*)\boldsymbol{B}_{l\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{s\boldsymbol{u}}(y). \qquad (2.39)
\end{aligned}$$

Combining (2.36) and (2.39) gives

$$\sigma_{ls}(x,y) = E_0\left\{\frac{\omega^{l+s}(X) I(X \le x \wedge y)}{h(X)}\right\} - F_l(x)F_s(y) + \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{B}_s^*(y). \qquad (2.40)$$

Summarizing (2.38) and (2.40), we conclude that for any $i,j \in \{l,s\}$

$$\sigma_{ij}(x,y) = E_0\left\{\frac{\omega^{i+j}(X) I(X \le x \wedge y)}{h(X)}\right\} - F_i(x)F_j(y) + \boldsymbol{B}_i^*(x)^\top \boldsymbol{W}\boldsymbol{B}_j^*(y), \qquad (2.41)$$

which is as claimed in the lemma. This completes the proof of (a).

For (b): We prove that the claim in (b) is correct for $l=0$ and $s=1$. The proofs for the other cases are similar and are omitted.

We first simplify the matrix $\boldsymbol{W}$. Let $\mathcal{M}_q^\top = (\boldsymbol{C}, \boldsymbol{U}^\top)$. Then $\mathcal{M}_q$ is full rank and therefore invertible. Note that

$$\boldsymbol{V} = \mathcal{M}_q^\top (\mathcal{M}_q^\top)^{-1}\boldsymbol{V}\mathcal{M}_q^{-1}\mathcal{M}_q = \mathcal{M}_q^\top (\mathcal{M}_q\boldsymbol{V}^{-1}\mathcal{M}_q^\top)^{-1}\mathcal{M}_q.$$

Recall that $\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{C} = \boldsymbol{0}$ and $\boldsymbol{J} = \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top$. Then

$$\mathcal{M}_q\boldsymbol{V}^{-1}\mathcal{M}_q^\top = \begin{pmatrix} \boldsymbol{C}^\top\boldsymbol{V}^{-1}\boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J} \end{pmatrix}$$

and

$$\boldsymbol{V} = \boldsymbol{C}(\boldsymbol{C}^\top\boldsymbol{V}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}^\top + \boldsymbol{U}^\top\boldsymbol{J}^{-1}\boldsymbol{U}.$$

Note that

$$\begin{aligned}
\boldsymbol{C}^\top \boldsymbol{V}^{-1}\boldsymbol{C} &= \boldsymbol{e}_{\boldsymbol{\theta}}^\top \boldsymbol{A}_{\boldsymbol{\theta\theta}}\boldsymbol{e}_{\boldsymbol{\theta}} + \{\lambda^*(1-\lambda^*)\}^2 \boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{A}_{\boldsymbol{uu}} e_{\boldsymbol{u}} \\
&= (1-\lambda^*)E_0\{h_1(X)\} + \{\lambda^*(1-\lambda^*)\}^2 E_0\left[\frac{\{\omega(X)-1\}^2}{h(X)}\right] \\
&= \lambda^*(1-\lambda^*),
\end{aligned}$$

where we use the fact that

$$\lambda^* E_0\left[\frac{\{\omega(X)-1\}^2}{h(X)}\right] + E_0\left\{\frac{\omega(X)-1}{h(X)}\right\} = 0$$

in the last step. The matrix $\boldsymbol{V}$ is expressed as

$$\boldsymbol{V} = \{\lambda^*(1-\lambda^*)\}^{-1}\boldsymbol{C}\boldsymbol{C}^\top + \boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U}.$$

This expression helps us to simplify $\boldsymbol{W}$ as

$$\begin{aligned}
\boldsymbol{W} &= \boldsymbol{V}^{-1}\boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U}\boldsymbol{V}^{-1} - \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{\boldsymbol{uu}}^{-1} \end{pmatrix} \\
&= \boldsymbol{V}^{-1}\{\boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U} - \boldsymbol{V}\}\boldsymbol{V}^{-1} + \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta\theta}}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta\theta}}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \{\lambda^*(1-\lambda^*)\}^{-1}\boldsymbol{V}^{-1}\boldsymbol{C}\boldsymbol{C}^\top \boldsymbol{V}^{-1} \\
&= \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta\theta}}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \{\lambda^*(1-\lambda^*)\}^{-1} \begin{pmatrix} \boldsymbol{e}_{\boldsymbol{\theta}} \\ -\lambda^*(1-\lambda^*)\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix} \begin{pmatrix} \boldsymbol{e}_{\boldsymbol{\theta}} \\ -\lambda^*(1-\lambda^*)\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix}^\top .
\end{aligned}$$

Substituting $\boldsymbol{W}$ into (2.41) and using the fact that

$$\begin{aligned}
\boldsymbol{B}_0^*(x)^\top \begin{pmatrix} \boldsymbol{e}_{\boldsymbol{\theta}} \\ -\lambda^*(1-\lambda^*)\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix} &= \lambda^* F_0(x), \\
\boldsymbol{B}_1^*(x)^\top \begin{pmatrix} \boldsymbol{e}_{\boldsymbol{\theta}} \\ -\lambda^*(1-\lambda^*)\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix} &= -(1-\lambda^*) F_1(x),
\end{aligned}$$

we find that for any $i, j \in \{l, s\}$

$$\sigma_{ij}(x,y) = E_0\left\{\frac{\omega^{i+j}(X) I(X \le x \wedge y)}{h(X)}\right\} + \boldsymbol{B}_{i\boldsymbol{\theta}}(x)^\top \boldsymbol{A}_{\boldsymbol{\theta\theta}}^{-1}\boldsymbol{B}_{j\boldsymbol{\theta}}(y) - \delta_{ij}F_i(x)F_j(y),$$

where

$$\delta_{ij} = \begin{cases} (1-\lambda^*)^{-1}, & i=j=0 \\ (\lambda^*)^{-1}, & i=j=1 \\ 0, & i \neq j. \end{cases}$$

This form is the same as that in Chen and Liu (2013) for the two-sample case, which completes the proof of (b).

For (c): Recall that $\boldsymbol{U}_m, \boldsymbol{V}_m$, and $\boldsymbol{J}_m$ denote the corresponding $\boldsymbol{U}, \boldsymbol{V}$, and $\boldsymbol{J}$ matrices obtained by using only the first $m$ estimating equations of $\boldsymbol{g}(\boldsymbol{x};\boldsymbol{\eta})$. We further define $\boldsymbol{\Sigma}_{ls}^{(m)}(x,y) = \{\sigma_{ij}^{(m)}(x,y)\}_{i,j\in\{l,s\}}$ and $\boldsymbol{B}_l^{*(m)}(x)$ to denote the corresponding matrix $\boldsymbol{\Sigma}_{ls}(x,y)$ and vector $\boldsymbol{B}_l(x)$ obtained by using the first $m$ estimating equations.

From the definitions of these matrices and vectors, we notice the following relationships:

$$\boldsymbol{U}_m = (\boldsymbol{U}_{m-1}, u_m); \quad \boldsymbol{V}_m = \begin{pmatrix} \boldsymbol{V}_{m-1} & \vartheta_{m-1,m} \\ \vartheta_{m,m-1} & \vartheta_{m,m} \end{pmatrix}; \quad \boldsymbol{B}_l^{*(m)}(x) = \begin{pmatrix} \boldsymbol{B}_l^{*(m-1)}(x) \\ b_{lm}(x) \end{pmatrix},$$

where $u_m$, $\vartheta_{m-1,m}$, $\vartheta_{m,m}$, and $b_{lm}(x)$ are the extra terms coming from the $m$th dimension of the estimating equations.

With the fact that

$$\boldsymbol{W} = \boldsymbol{V}^{-1}(\boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U} - \boldsymbol{V})\boldsymbol{V}^{-1} + \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta\theta}}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix},$$

the entry in the covariance matrix $\boldsymbol{\Sigma}_{ls}^{(m)}(x,y)$ can be written as

$$\begin{aligned} \sigma_{ij}^{(m)}(x,y) &= E_0\left\{\frac{\omega^{i+j}(X)I(X\leq x\wedge y)}{h(X)}\right\} - F_i(x)F_j(y) + \boldsymbol{B}_i^{*(m)}(x)^\top \boldsymbol{W}\boldsymbol{B}_j^{*(m)}(y) \\ &= E_0\left\{\frac{\omega^{i+j}(X)I(X\leq x\wedge y)}{h(X)}\right\} - F_i(x)F_j(y) + \boldsymbol{B}_{i\boldsymbol{\theta}}(x)^\top \boldsymbol{A}_{\boldsymbol{\theta\theta}}^{-1}\boldsymbol{B}_{j\boldsymbol{\theta}}(y) \\ &\quad -\boldsymbol{B}_i^{*(m)}(x)^\top \boldsymbol{V}_m^{-1}(\boldsymbol{V}_m - \boldsymbol{U}_m^\top \boldsymbol{J}_m^{-1}\boldsymbol{U}_m)\boldsymbol{V}_m^{-1}\boldsymbol{B}_j^{*(m)}(x) \end{aligned}$$

for any $i,j\in\{l,s\}$.

Therefore,

$$
\begin{aligned}
&\boldsymbol{\Sigma}_{ls}^{(m-1)}(x,y) - \boldsymbol{\Sigma}_{ls}^{(m)}(x,y) \\
= &\begin{pmatrix} \boldsymbol{B}_l^{*(m)}(x) \\ \boldsymbol{B}_s^{*(m)}(y) \end{pmatrix}^\top (\boldsymbol{V}_m - \boldsymbol{U}_m^\top \boldsymbol{J}_m^{-1} \boldsymbol{U}_m) \boldsymbol{V}_m^{-1} \begin{pmatrix} \boldsymbol{B}_l^{*(m)}(x) \\ \boldsymbol{B}_s^{*(m)}(y) \end{pmatrix} \\
&- \begin{pmatrix} \boldsymbol{B}_l^{*(m-1)}(x) \\ \boldsymbol{B}_s^{*(m-1)}(y) \end{pmatrix}^\top (\boldsymbol{V}_{m-1} - \boldsymbol{U}_{m-1}^\top \boldsymbol{J}_{m-1}^{-1} \boldsymbol{U}_{m-1}) \boldsymbol{V}_{m-1}^{-1} \begin{pmatrix} \boldsymbol{B}_l^{*(m-1)}(x) \\ \boldsymbol{B}_s^{*(m-1)}(y) \end{pmatrix}.
\end{aligned}
$$

Using the results in (2.30) and (2.31), we have

$$
\begin{aligned}
&\boldsymbol{V}_m^{-1}\{\boldsymbol{V}_m - \boldsymbol{U}_m^\top \boldsymbol{J}_m^{-1} \boldsymbol{U}_m\} \boldsymbol{V}_m^{-1} \\
\geq &\begin{pmatrix} \boldsymbol{V}_{m-1}^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \left\{ \begin{pmatrix} \boldsymbol{V}_{m-1} & \vartheta_{m-1,m} \\ \vartheta_{m,m-1} & \vartheta_{m,m} \end{pmatrix} \right. \\
&\qquad \left. - \begin{pmatrix} \boldsymbol{U}_{m-1}^\top \\ u_m^\top \end{pmatrix} \boldsymbol{J}_{m-1}^{-1} (\boldsymbol{U}_{m-1}, u_m) \right\} \begin{pmatrix} \boldsymbol{V}_{m-1}^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \\
\geq &\begin{pmatrix} \boldsymbol{V}_{m-1}^{-1}\{\boldsymbol{V}_{m-1} - \boldsymbol{U}_{m-1}^\top \boldsymbol{J}_{m-1}^{-1} \boldsymbol{U}_{m-1}\} \boldsymbol{V}_{m-1}^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix}.
\end{aligned}
$$

This implies that

$$
\begin{aligned}
&\boldsymbol{\Sigma}_{ls}^{(m-1)}(x,y) - \boldsymbol{\Sigma}_{ls}^{(m)}(x,y) \\
\geq &\begin{pmatrix} \boldsymbol{B}_l^{*(m)}(x) \\ \boldsymbol{B}_s^{*(m)}(y) \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{V}_{m-1}^{-1}\{\boldsymbol{V}_{m-1} - \boldsymbol{U}_{m-1}^\top \boldsymbol{J}_{m-1}^{-1} \boldsymbol{U}_{m-1}\} \boldsymbol{V}_{m-1}^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{B}_l^{*(m)}(x) \\ \boldsymbol{B}_s^{*(m)}(y) \end{pmatrix} \\
&- \begin{pmatrix} \boldsymbol{B}_l^{*(m-1)}(x) \\ \boldsymbol{B}_s^{*(m-1)}(y) \end{pmatrix}^\top (\boldsymbol{V}_{m-1} - \boldsymbol{U}_{m-1}^\top \boldsymbol{J}_{m-1}^{-1} \boldsymbol{U}_{m-1}) \boldsymbol{V}_{m-1}^{-1} \begin{pmatrix} \boldsymbol{B}_l^{*(m-1)}(x) \\ \boldsymbol{B}_s^{*(m-1)}(y) \end{pmatrix} \\
= &\mathbf{0}.
\end{aligned}
$$

This completes the proof of (c).

**Proof of Theorem 2.5**

We first introduce two lemmas that will be helpful in the proof of Theorem 2.5. The following lemma establishes the convergence rate of $\hat{\xi}_{i,\tau}$.

**Lemma 2.4.** *Assume the conditions of Theorem 5 are satisfied. For each fixed $\tau \in (0,1)$ and $i = 0, 1$, we have*

$$\hat{\xi}_{i,\tau} - \xi_{i,\tau} = O_p(n^{-1/2}).$$

*Proof.* We concentrate on the case $i = 0$; the case $i = 1$ can be proved similarly. Let $\Delta_n = \sup_x |\hat{F}_0(x) - F_0(x)|$. It suffices to show that (Chen and Liu, 2013; Chen et al., 2021)

$$\Delta_n = O_p(n^{-1/2}). \tag{2.42}$$

Define

$$\check{F}_0(x) = \frac{1}{n}\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{I(X_{ij} \le x)}{1 + \lambda^* \left[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} - 1\right]}.$$

Then

$$\Delta_n = \sup_x |\hat{F}_0(x) - F_0(x)| \le \sup_x |\hat{F}_0(x) - \check{F}_0(x)| + \sup_x |\check{F}_0(x) - F_i(x)| = \Delta_{n1} + \Delta_{n2},$$

where

$$\Delta_{n1} = \sup_x |\hat{F}_0(x) - \check{F}_0(x)|$$

and

$$\Delta_{n2} = \sup_x |\check{F}_0(x) - F_0(x)|.$$

Following the proof of Theorem 3.1 in Chen and Liu (2013) and Lemma 1 in Chen et al. (2021), we can verify that

$$\Delta_{n2} = O_p(n^{-1/2}).$$

With this result, the claim (2.42) is proved if $\Delta_{n1} = O_p(n^{-1/2})$.

As preparation, we argue that

$$(n\hat{p}_{ij})^{-1} = 1 + \hat{\lambda}[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} - 1] + \hat{\boldsymbol{\nu}}^\top \boldsymbol{g}(X_{ij}; \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}) \ge 1 - \lambda^* + o_p(1) \tag{2.43}$$

or equivalently $\hat{p}_{ij} \le n^{-1}\{1 - \lambda^* + o_p(1)\}^{-1} = O_p(1/n)$. Note that

$$(n\hat{p}_{ij})^{-1} \ge 1 - \hat{\lambda} + \hat{\boldsymbol{\nu}}^\top \boldsymbol{g}(X_{ij}; \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}) \ge 1 - \hat{\lambda} - \|\hat{\boldsymbol{\nu}}\| \max_{ij} \|\boldsymbol{g}(X_{ij}; \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})\|.$$

By Condition C5,

$$\max_{ij} \|\boldsymbol{g}(X_{ij}; \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})\| \le \max_{ij} R^{1/3}(X_{ij}) = o_p(n^{1/2}),$$

which, together with $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^* = O_p(n^{-1/2})$, implies that (2.43) is valid.

We now return to argue that $\Delta_{n1} = O_p(n^{-1/2})$. After some algebra, we have

$$\begin{aligned} & \hat{F}_0(x) - \check{F}_0(x) \\ = & \sum_{i=0}^{1}\sum_{j=1}^{n_i} \hat{p}_{ij} \frac{(\lambda^* - \hat{\lambda})\left[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} - 1\right] - \hat{\boldsymbol{\nu}}^\top \boldsymbol{g}(X_{ij}; \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})}{1 + \lambda^*\left[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} - 1\right]} I(X_{ij} \le x). \end{aligned}$$

Using (2.43), we have

$$\begin{aligned} |\hat{F}_0(x) - \check{F}_0(x)| \le & \; O_p(1/n) \sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{|\hat{\lambda} - \lambda^*| \left[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} + 1\right]}{1 + \lambda^*\left[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} - 1\right]} I(X_{ij} \le x) \\ & + O_p(1/n) \sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{|\hat{\boldsymbol{\nu}}^\top \boldsymbol{g}(X_{ij}; \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})|}{1 + \lambda^*\left[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} - 1\right]} I(X_{ij} \le x) \\ \le & \; O_p(1/n) \sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{|\hat{\lambda} - \lambda^*|}{\lambda^*(1 - \lambda^*)} I(X_{ij} \le x) \\ & + O_p(1/n) \sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{|\hat{\boldsymbol{\nu}}^\top \boldsymbol{g}(X_{ij}; \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})|}{1 - \lambda^*} I(X_{ij} \le x). \end{aligned} \tag{2.44}$$

By Condition C5,

$$\Delta_{n1} = \sup_x |\hat{F}_0(x) - \check{F}_0(x)| \le O_p(1)|\hat{\lambda} - \lambda^*| + O_p(1)\frac{1}{n}\sum_{i=0}^{1}\sum_{j=1}^{n_i} \left\{\|\hat{\boldsymbol{\nu}}\| R^{1/3}(X_{ij})\right\},$$

which, together with $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^* = O_p(n^{-1/2})$, implies that

$$\Delta_{n1} = O_p(n^{-1/2}).$$

This completes the proof. □

**Lemma 2.5.** *Under the regularity conditions, for any $c > 0$ and $i = 0, 1$, we have*

$$\sup_{x:\ |x - \xi_{i,\tau}| < cn^{-1/2}} |\{\hat{F}_i(x) - \hat{F}_i(\xi_{i,\tau})\} - \{F_i(x) - F_i(\xi_{i,\tau})\}| = O_p(n^{-3/4}(\log(n))^{1/2}).$$

*Proof.* We prove this lemma for $i = 0$; the case $i = 1$ is equivalent. Without loss of generality we assume $x \geq \xi_{0,\tau}$. Note that

$$\begin{aligned} & |\{\hat{F}_0(x) - \hat{F}_0(\xi_{0,\tau})\} - \{F_0(x) - F_0(\xi_{0,\tau})\}| \\ \leq & |\{\hat{F}_0(x) - \hat{F}_0(\xi_{0,\tau})\} - \{\check{F}_0(x) - \check{F}_0(\xi_{0,\tau})\}| \\ & + |\{\check{F}_0(x) - \check{F}_0(\xi_{0,\tau})\} - \{F_0(x) - F_0(\xi_{0,\tau})\}|. \end{aligned} \tag{2.45}$$

Following the proof of Lemma A.2 in Chen and Liu (2013), we can verify that

$$\sup_{x:\ 0 \leq x - \xi_{0,\tau} < cn^{-1/2}} |\{\check{F}_0(x) - \check{F}_0(\xi_{0,\tau})\} - \{F_0(x) - F_0(\xi_{0,\tau})\}| = O_p(n^{-3/4}(\log(n))^{1/2}).$$

Consequently, we need to show only that the first term in (2.45) has a higher order than $n^{-3/4}(\log(n))^{1/2}$ uniformly in $0 \leq x - \xi_{0,\tau} < cn^{-1/2}$.

With the technique used to obtain (2.44), we have

$$\begin{aligned} & |\{\hat{F}_0(x) - \hat{F}_0(\xi_{0,\tau})\} - \{\check{F}_0(x) - \check{F}_0(\xi_{0,\tau})\}| \\ \leq & O_p(1/n) \sum_{i=0}^{1} \sum_{j=1}^{n_i} \frac{|\hat{\lambda} - \lambda^*|}{\lambda^*(1-\lambda^*)} I(\xi_{0,\tau} < X_{ij} \leq x) \\ & + O_p(1/n) \sum_{i=0}^{1} \sum_{j=1}^{n_i} \frac{\|\hat{\boldsymbol{\nu}}\| R^{1/3}(X_{ij})}{1-\lambda^*} I(\xi_{0,\tau} < X_{ij} \leq x) \\ = & O_p(n^{-1/2}) \frac{1}{n} \sum_{i=0}^{1} \sum_{j=1}^{n_i} \{1 + R^{1/3}(X_{ij})\} I(\xi_{0,\tau} < X_{ij} \leq x). \end{aligned}$$

By Condition C5,

$$E_0\{1 + R^{1/3}(X)\} < \infty \ \text{ and } \ E_1\{1 + R^{1/3}(X)\} < \infty,$$

then uniformly in $x$

$$E_0[\{1 + R^{1/3}(X)\} I(\xi_{0,\tau} < X_{ij} \leq x)] = O_p(n^{-1/2})$$

and

$$E_1[\{1 + R^{1/3}(X)\} I(\xi_{0,\tau} < X_{ij} \leq x)] = O_p(n^{-1/2}).$$

Therefore,

$$\sup_{x:\ 0\le x-\xi_{0,\tau}<cn^{-1/2}} |\{\hat{F}_0(x)-\hat{F}_0(\xi_{0,\tau})\}-\{\check{F}_0(x)-\check{F}_0(\xi_{0,\tau})\}| = O_p(n^{-1}).$$

This completes the proof. □

We are now ready to prove Theorem 2.5. By Lemma 2.4, for $i=0,1$,

$$\begin{aligned} F_i(\hat{\xi}_{i,\tau})-F_i(\xi_{i,\tau}) &= f_i(\xi_{i,\tau})(\hat{\xi}_{i,\tau}-\xi_{i,\tau})+O_p((\hat{\xi}_{i,\tau}-\xi_{i,\tau})^2) \\ &= f_i(\xi_{i,\tau})(\hat{\xi}_{i,\tau}-\xi_{i,\tau})+O_p(n^{-1}). \end{aligned} \tag{2.46}$$

Note that $\hat{F}_i(\hat{\xi}_{i,\tau})=\tau+O_p(n^{-1})$. Replacing $x$ by $\hat{\xi}_{i,\tau}$ in Lemma 2.5 and using (2.46) yields

$$\tau-\hat{F}_i(\xi_{i,\tau}) = f_i(\xi_{i,\tau})(\hat{\xi}_{i,\tau}-\xi_{i,\tau})+O_p(n^{-3/4}(\log(n))^{1/2}).$$

This completes the proof.

**Proof of Theorem 2.6**

The results in (a) and (b) are direct consequences of Theorems 4 and 5.

For (c): We note that

$$\mathbf{\Omega}_{ls} = \begin{pmatrix} 1/f_l(\xi_{l,\tau_l}) & 0 \\ 0 & 1/f_s(\xi_{s,\tau_s}) \end{pmatrix} \mathbf{\Sigma}_{ls}(\xi_{l,\tau_l},\xi_{s,\tau_s}) \begin{pmatrix} 1/f_l(\xi_{l,\tau_l}) & 0 \\ 0 & 1/f_s(\xi_{s,\tau_s}) \end{pmatrix}.$$

Then Theorem 2.4(c) implies the results in (c). This completes the proof.

### 2.6.4 Additional simulation results

Table 2.9 presents the four quantile estimates under gamma distributions. Table 2.10 presents the three CIs for quantiles under gamma distributions. The general summary statements are similar to those for normal distributions, and hence are omitted.

Table 2.9: RB (%) and MSE (×100) for quantile estimators (gamma distributions).

| | | | $Gam(8, 1.125)$ | | | | $Gam(6, 1, 5)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $(n_0, n_1)$ | $\tau$ | | EMP | EL | DRM | DRM-EE | EMP | EL | DRM | DRM-EE |
| (50, 50) | 0.10 | RB | -2.25 | -0.05 | 0.25 | 0.16 | -1.40 | 0.71 | 1.26 | 0.65 |
| | | MSE | 29.71 | 25.04 | 23.26 | 20.29 | 31.70 | 26.96 | 26.66 | 22.88 |
| | 0.25 | RB | 0.01 | -0.04 | 0.08 | 0.03 | 0.75 | 0.30 | 0.47 | -0.06 |
| | | MSE | 25.02 | 19.93 | 21.38 | 16.39 | 32.91 | 24.71 | 27.78 | 20.32 |
| | 0.50 | RB | -1.03 | -0.04 | -0.15 | -0.02 | -0.74 | -0.07 | 0.28 | -0.08 |
| | | MSE | 30.99 | 23.20 | 25.91 | 17.32 | 40.46 | 25.74 | 35.52 | 19.68 |
| | 0.75 | RB | -0.13 | -0.02 | -0.33 | -0.13 | -0.02 | -0.20 | 0.15 | 0.12 |
| | | MSE | 48.41 | 35.85 | 42.11 | 28.23 | 65.70 | 43.10 | 57.48 | 33.81 |
| | 0.90 | RB | -1.85 | 0.15 | -0.47 | -0.20 | -1.93 | 0.01 | 0.12 | 0.14 |
| | | MSE | 99.19 | 86.91 | 83.12 | 62.28 | 133.79 | 110.01 | 120.28 | 86.79 |
| (50, 150) | 0.10 | RB | -2.25 | 0.05 | 0.41 | 0.32 | -0.36 | 0.36 | 0.42 | 0.33 |
| | | MSE | 29.98 | 23.32 | 20.31 | 15.18 | 10.40 | 9.74 | 9.86 | 9.10 |
| | 0.25 | RB | -0.02 | 0.01 | 0.03 | -0.02 | 0.19 | 0.09 | 0.12 | -0.03 |
| | | MSE | 25.11 | 17.45 | 19.28 | 11.05 | 10.58 | 9.27 | 9.89 | 8.61 |
| | 0.50 | RB | -1.03 | 0.02 | -0.18 | -0.01 | -0.21 | 0.01 | 0.12 | -0.03 |
| | | MSE | 31.26 | 17.31 | 22.92 | 9.55 | 14.17 | 11.46 | 12.98 | 10.15 |
| | 0.75 | RB | -0.15 | 0.04 | -0.45 | -0.16 | -0.06 | -0.18 | 0.02 | -0.06 |
| | | MSE | 48.19 | 27.80 | 36.99 | 15.98 | 21.18 | 17.52 | 19.94 | 15.74 |
| | 0.90 | RB | -1.83 | 0.42 | -0.56 | -0.09 | -0.62 | -0.05 | 0.11 | 0.03 |
| | | MSE | 99.26 | 74.83 | 74.58 | 43.00 | 44.60 | 40.83 | 40.68 | 36.31 |
| (100, 100) | 0.10 | RB | -1.03 | 0.07 | 0.41 | 0.32 | -0.92 | 0.25 | 0.35 | 0.15 |
| | | MSE | 14.47 | 13.00 | 11.19 | 9.91 | 16.95 | 14.43 | 14.18 | 11.95 |
| | 0.25 | RB | -0.54 | 0.06 | 0.06 | 0.03 | -0.52 | -0.02 | 0.10 | -0.12 |
| | | MSE | 12.76 | 10.64 | 10.81 | 8.35 | 15.41 | 11.85 | 13.73 | 9.82 |
| | 0.50 | RB | -0.48 | 0.03 | -0.03 | -0.02 | -0.41 | 0.02 | 0.14 | -0.03 |
| | | MSE | 15.70 | 11.67 | 12.92 | 8.89 | 20.57 | 13.41 | 17.58 | 9.84 |
| | 0.75 | RB | -0.61 | -0.04 | -0.19 | -0.14 | -0.71 | -0.17 | 0.04 | -0.06 |
| | | MSE | 24.94 | 18.73 | 19.98 | 13.73 | 32.29 | 20.67 | 27.94 | 16.02 |
| | 0.90 | RB | -0.94 | 0.05 | -0.20 | -0.09 | -1.11 | 0.03 | 0.01 | 0.09 |
| | | MSE | 48.17 | 42.30 | 41.07 | 31.30 | 70.72 | 54.02 | 57.47 | 40.26 |
| (200, 200) | 0.10 | RB | -0.44 | 0.04 | 0.24 | 0.16 | -0.50 | 0.15 | 0.15 | 0.07 |
| | | MSE | 7.03 | 6.34 | 5.54 | 4.80 | 8.17 | 7.06 | 6.80 | 5.81 |
| | 0.25 | RB | -0.29 | 0.01 | 0.08 | 0.05 | -0.31 | -0.04 | -0.01 | -0.10 |
| | | MSE | 6.53 | 5.24 | 5.19 | 3.92 | 7.59 | 5.89 | 6.52 | 4.79 |
| | 0.50 | RB | -0.23 | 0.02 | -0.03 | -0.03 | -0.31 | -0.11 | -0.02 | -0.07 |
| | | MSE | 7.83 | 5.84 | 6.15 | 4.25 | 9.90 | 6.03 | 8.39 | 4.76 |
| | 0.75 | RB | -0.38 | -0.12 | -0.11 | -0.10 | -0.29 | 0.05 | 0.02 | 0.03 |
| | | MSE | 11.98 | 9.21 | 10.19 | 7.24 | 17.41 | 11.09 | 14.98 | 8.33 |
| | 0.90 | RB | -0.48 | 0.00 | -0.09 | -0.07 | -0.42 | 0.09 | 0.08 | 0.13 |
| | | MSE | 23.81 | 20.31 | 19.73 | 15.34 | 36.06 | 26.76 | 31.15 | 20.87 |

Table 2.10: CP (%) and AL for three 95% CIs of $100\tau\%$-quantile (gamma distributions).

| | | | $Gam(8, 1.125)$ | | | $Gam(6, 1.5)$ | | |
|---|---|---|---|---|---|---|---|---|
| $(n_0, n_1)$ | $\tau$ | | EMP | DRM | DRM-EE | EMP | DRM | DRM-EE |
| (50,50) | 0.10 | CP | 94.7 | 95.1 | 95.5 | 93.7 | 94.5 | 94.9 |
| | | AL | 2.10 | 1.89 | 1.77 | 2.24 | 2.10 | 1.93 |
| | 0.25 | CP | 94.9 | 94.7 | 94.5 | 95.4 | 94.5 | 94.8 |
| | | AL | 2.03 | 1.82 | 1.60 | 2.25 | 2.04 | 1.73 |
| | 0.50 | CP | 93.2 | 94.4 | 94.3 | 94.2 | 95.1 | 94.9 |
| | | AL | 2.06 | 1.99 | 1.62 | 2.33 | 2.31 | 1.74 |
| | 0.75 | CP | 94.2 | 94.2 | 94.0 | 95.8 | 93.7 | 93.7 |
| | | AL | 2.86 | 2.55 | 2.10 | 3.46 | 3.04 | 2.29 |
| | 0.90 | CP | 94.8 | 94.7 | 94.9 | 94.7 | 94.3 | 94.9 |
| | | AL | 4.17 | 3.73 | 3.27 | 5.03 | 4.68 | 3.80 |
| (50,150) | 0.10 | CP | 94.7 | 95.2 | 95.4 | 94.1 | 95.2 | 95.5 |
| | | AL | 2.10 | 1.77 | 1.56 | 1.29 | 1.24 | 1.20 |
| | 0.25 | CP | 94.9 | 94.8 | 94.7 | 94.8 | 94.5 | 94.5 |
| | | AL | 2.03 | 1.72 | 1.33 | 1.28 | 1.22 | 1.14 |
| | 0.50 | CP | 93.2 | 94.7 | 94.7 | 94.2 | 94.1 | 94.0 |
| | | AL | 2.06 | 1.86 | 1.20 | 1.37 | 1.37 | 1.22 |
| | 0.75 | CP | 94.2 | 94.4 | 94.9 | 95.9 | 95.6 | 95.4 |
| | | AL | 2.86 | 2.41 | 1.58 | 1.88 | 1.79 | 1.60 |
| | 0.90 | CP | 94.8 | 94.8 | 95.0 | 94.4 | 95.4 | 95.1 |
| | | AL | 4.17 | 3.44 | 2.60 | 2.72 | 2.68 | 2.47 |
| (100,100) | 0.10 | CP | 95.0 | 95.2 | 94.5 | 95.0 | 94.0 | 93.7 |
| | | AL | 1.53 | 1.33 | 1.24 | 1.66 | 1.47 | 1.35 |
| | 0.25 | CP | 94.8 | 94.4 | 93.9 | 95.2 | 95.1 | 95.1 |
| | | AL | 1.42 | 1.28 | 1.12 | 1.58 | 1.44 | 1.22 |
| | 0.50 | CP | 93.8 | 94.9 | 94.2 | 94.3 | 94.3 | 94.3 |
| | | AL | 1.48 | 1.39 | 1.14 | 1.73 | 1.61 | 1.22 |
| | 0.75 | CP | 95.0 | 94.5 | 95.4 | 95.2 | 95.5 | 94.9 |
| | | AL | 1.99 | 1.78 | 1.46 | 2.33 | 2.11 | 1.60 |
| | 0.90 | CP | 96.2 | 95.5 | 94.8 | 94.9 | 94.9 | 95.3 |
| | | AL | 3.01 | 2.57 | 2.25 | 3.58 | 3.15 | 2.60 |
| (200,200) | 0.10 | CP | 93.8 | 95.2 | 94.7 | 93.8 | 95.2 | 95.4 |
| | | AL | 1.02 | 0.94 | 0.87 | 1.10 | 1.03 | 0.95 |
| | 0.25 | CP | 95.4 | 95.4 | 95.2 | 94.2 | 95.1 | 94.8 |
| | | AL | 0.99 | 0.90 | 0.79 | 1.10 | 1.01 | 0.85 |
| | 0.50 | CP | 94.4 | 95.0 | 94.8 | 94.2 | 94.8 | 94.8 |
| | | AL | 1.05 | 0.98 | 0.81 | 1.21 | 1.13 | 0.85 |
| | 0.75 | CP | 95.1 | 95.0 | 95.0 | 95.5 | 94.8 | 94.9 |
| | | AL | 1.37 | 1.26 | 1.04 | 1.61 | 1.49 | 1.13 |
| | 0.90 | CP | 93.7 | 94.9 | 94.7 | 94.6 | 94.1 | 95.0 |
| | | AL | 1.94 | 1.78 | 1.55 | 2.30 | 2.17 | 1.80 |

# Chapter 3

# Empirical Likelihood Inference on the Youden Index and the Optimal Cutoff Point under Density Ratio Models

## 3.1 Introduction

ROC curves are widely used statistical tools in medical research to evaluate the discriminatory effectiveness of a biomarker for distinguishing diseased individuals from healthy ones. When the sampling distribution of the biomarker is continuous, the ROC curve plots the proportion of true positive (sensitivity) versus proportion of false positive (one minus specificity) across all possible choices of threshold values, called cutoff points, of the biomarker. We refer to Zhou and McClish (2002), Pepe (2003), Krzanowski and Hand (2009), Zou et al. (2011), Chen et al. (2016), and references therein for comprehensive reviews and recent developments in ROC analysis.

The Youden index, first proposed by Youden (1950), is one of popular summary statistics of the ROC curve. It is defined as the maximum of the sum of sensitivity and specificity minus one when the relative seriousness of false positive and false negative are treated equally. The Youden index ranges from 0 to 1 with 1 indicating a complete separation of distributions of biomarkers in healthy and diseased populations and 0 indicating a complete overlap. It has the advantage of providing a criterion to choose the "optimal" cutoff point, which maximizes the sum of sensitivity and specificity minus one. See Fluss et al. (2005) for more discussions on the advantages of the Youden index.

In medical researches, larger values of a biomarker are generally associated with diseases. Therefore, an individual is classified as diseased when the biomarker of the individual is greater than a given cutoff point. Let $F_0$ and $F_1$ denote the CDFs of the healthy population and the diseased population, respectively. The sensitivity and the specificity are respectively equal to $1 - F_1(x)$ and $F_0(x)$ for the given cutoff point $x$. Therefore, the Youden index can be equivalently expressed as

$$J = \max_x\{F_0(x) - F_1(x)\} = F_0(c) - F_1(c), \tag{3.1}$$

where $c$ is the corresponding optimal cutoff point.

In the literature, there are two types of methods, namely, the parametric methods and the nonparametric methods, for estimating the Youden index $J$ and the corresponding optimal cutoff point $c$. For parametric methods, the original biomarkers or the biomarkers after the Box-Cox transformation (Box and Cox, 1964) in the healthy and diseased groups are assumed to follow the same parametric distribution family (Fluss et al., 2005; Bantis et al., 2019). Nonparametric methods employ techniques such as the empirical CDF (ECDF) method or the kernel method to obtain the estimators of $F_0$ and $F_1$, which are then used to obtain the point estimators of $J$ and $c$. More details about the ECDF-based and kernel-based methods, and their modified versions can be found in Hsieh and Turnbull (1996), Zhou and Qin (2012), and Shan (2015). Recently, Bantis et al. (2019) employed hazard constrained natural spline (HCNS) as an alternative nonparametric approach to estimate $J$ and $c$. The delta and bootstrap methods (Schisterman and Perkins, 2007; Yin and Tian, 2014; Bantis et al., 2019) and the empirical likelihood methods (Wang et al., 2017b) are used to construct CIs for $J$ and $c$.

In applications, the measurement of a biomarker may have a fixed and finite lower limit of detection (LLOD). For instance, the quantitation of human immunodeficiency virus RNA in human plasma has a LLOD 500 copies/ml with the Amplicor Monitor assay or has a LLOD 50 copies/ml with the Ultrasensitive assay (Gulick et al., 2000). More examples of LLODs can be found in Ruopp et al. (2008), Bantis et al. (2017), and references therein. Ruopp et al. (2008) adapted the parametric method, the ECDF method, and the ROC-generalized linear model (ROC-GLM) method (Pepe, 2000; Alonzo and Pepe, 2002; Pepe, 2003) to obtain point estimates and construct CIs for $J$ and $c$ in those situations.

Generally speaking, the parametric likelihood based estimators of $(J, c)$ are quite efficient, but may not be robust to model misspecifications (Fluss et al., 2005). Nonparametric methods are free from model assumptions on $F_0$ and $F_1$, but the resulting estimators of

$(J, c)$, especially the estimator of $c$, may be inefficient. When there is no LLOD, Hsieh and Turnbull (1996) showed that the convergence rates of the ECDF-based and the kernel-based estimators of $c$ are slower than $n^{-1/2}$, where $n$ is the total sample size.

In this chapter, we develop a semiparametric method that enables efficient estimation of both $J$ and $c$ without making risky parametric assumptions on $F_0$ and $F_1$. In medical researches, the two populations under consideration usually share certain common characteristics (Qin and Zhang, 2003; Qin, 2017; Zhuang et al., 2019). To incorporate the information from both samples, we suggest to use the DRM (1.1) to link $F_0$ and $F_1$. That is, we postulate

$$dF_1(x) = \exp\{\alpha + \boldsymbol{\beta}^\top \mathbf{q}(x)\} dF_0(x) = \exp\{\boldsymbol{\theta}^\top \mathbf{Q}(x)\} dF_0(x), \tag{3.2}$$

where $dF_i(x)$ denotes the density of $F_i(x)$ for $i = 0, 1$; $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top)^\top$ are unknown parameters; and $\boldsymbol{Q}(x) = (1, \boldsymbol{q}(x))^\top$ with a $d$-dimensional basis function $\boldsymbol{q}(x)$. As we discussed in Section 1.3, the inference procedures based on DRMs are more efficient than the fully nonparametric procedures. To the best of our knowledge, the inference procedures for $(J, c)$ under a DRM have not been studied in the existing literature. This chapter fills the void.

The rest of the chapter is organized as follows. In Section 3.2, we propose the MELEs of $J$ and $c$ under a DRM and study their asymptotic results. Confidence intervals of $J$ and $c$ are then constructed based on the asymptotic properties. Simulation studies are presented in Section 3.3 and a real-data application is given in Section 3.4. For the convenience of presentation, detailed review of some existing methods, proofs, and additional simulation results are deferred to Section 3.5.

## 3.2 Main Results

### 3.2.1 Point estimation of $J$ and $c$ under the DRM

Denote $\{X_{01}, \ldots, X_{0n_0}\}$ and $\{X_{11}, \ldots, X_{1n_1}\}$ as two independent random samples from the healthy and diseased populations, respectively. Let $f_0$ and $f_1$ be the probability density functions of $F_0$ and $F_1$, respectively. Following the definition of Youden index in (3.1), the optimal cutoff point $c$ satisfies $f_0(c) = f_1(c)$, which together with (3.2) implies that

$$\boldsymbol{\theta}^\top \boldsymbol{Q}(c) = 0. \tag{3.3}$$

The above equation serves as the basis for estimating $c$.

In the following, we focus on cases where the biomarker has a LLOD, denoted as $r$, and develop estimators for $(J, c)$. Analysis of data without a LLOD amounts to setting $r = -\infty$. Let $m_0$ and $m_1$ be the numbers of observations above the LLOD $r$ in the healthy and diseased groups, respectively. Without loss of generality, we assume the first $m_i$ observations in sample $i$, $X_{i1}, \ldots, X_{im_i}$ are above the LLOD. Let $\zeta_0 = P(x_{01} \geq r)$ and $\zeta_1 = P(x_{11} \geq r)$.

We now discuss the maximum empirical likelihood procedure for estimating the unknown parameters and functions. By the empirical likelihood principle (Owen, 2001) and under the DRM (3.2), the full likelihood can be written as

$$\begin{aligned} L_n &= \prod_{i=0}^{1}\left[(1-\zeta_i)^{n_i-m_i}\prod_{j=1}^{m_i} dF_i(X_{ij})\right] \\ &= \left\{\prod_{i=0}^{1}(1-\zeta_i)^{n_i-m_i}\right\}\left[\prod_{i=0}^{1}\prod_{j=1}^{m_i} p_{ij}\prod_{j=1}^{m_1}\exp\{\boldsymbol{\theta}^\top \mathbf{Q}(X_{1j})\}\right], \end{aligned}$$

where $p_{ij} = dF_0(X_{ij})$ for $i = 0, 1,\ j = 1, \ldots, m_i$ and they satisfy the following constraints:

$$p_{ij} \geq 0,\ 0 < \sum_{i=0}^{1}\sum_{j=1}^{m_i} p_{ij} = \zeta_0 \leq 1,\ 0 < \sum_{i=0}^{1}\sum_{j=1}^{m_i} p_{ij}\exp\{\boldsymbol{\theta}^\top \mathbf{Q}(X_{ij})\} = \zeta_1 \leq 1. \tag{3.4}$$

Let $\boldsymbol{P} = \{p_{ij}\}$. The MELEs of $(\boldsymbol{\theta}, \zeta_0, \zeta_1, \boldsymbol{P})$, denoted as $(\hat{\boldsymbol{\theta}}, \hat{\zeta}_0, \hat{\zeta}_1, \hat{\boldsymbol{P}})$, are defined as the maximizer of $L_n$ subject to the constraints in Equation (3.4). It is shown by Cai and Chen (2018) that

$$\hat{\zeta}_k = m_k/n_k, \quad k = 0, 1,$$

and the $\hat{\boldsymbol{\theta}}$ maximizes the following dual profile empirical log-likelihood function

$$\ell_n(\boldsymbol{\theta}) = \sum_{j=1}^{m_1}\{\boldsymbol{\theta}^\top \mathbf{Q}(X_{1j})\} - \sum_{i=0}^{1}\sum_{j=1}^{m_i}\log\left[1 + \rho\exp\{\boldsymbol{\theta}^\top \mathbf{Q}(X_{ij})\}\right],$$

where $\rho = n_1/n_0$. That is, $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell_n(\boldsymbol{\theta})$. The MELEs of $p_{ij}$'s are given by

$$\hat{p}_{ij} = n_0^{-1}\left[1 + \rho\exp\{\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(X_{ij})\}\right]^{-1}, \quad i = 0, 1,\ j = 1, \cdots, m_i.$$

It follows that, for any $x \geq r$, the MELEs of $F_0$ and $F_1$ are given by

$$\hat{F}_0(x) = (1-\hat{\zeta}_0) + \frac{1}{n_0}\sum_{i=0}^{1}\sum_{j=1}^{m_i}\frac{1}{1+\rho\exp\{\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(X_{ij})\}} I(r \leq X_{ij} \leq x)\,,$$
$$\hat{F}_1(x) = (1-\hat{\zeta}_1) + \frac{1}{n_0}\sum_{i=0}^{1}\sum_{j=1}^{m_i}\frac{\exp\{\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(X_{ij})\}}{1+\rho\exp\{\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(X_{ij})\}} I(r \leq X_{ij} \leq x)\,,$$

where $I(\cdot)$ is the indicator function.

With the MELE $\hat{\boldsymbol{\theta}}$ and Equation (3.3), the MELE of the optimal cutoff point $c$, denoted as $\hat{c}$, can be obtained as the solution to the equation

$$\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(x) = 0. \tag{3.5}$$

Let $X_{min}$ and $X_{max}$ be the minimum and maximum values of the observations in the two samples which are above the LLOD, respectively. If multiple solutions exist for (3.5) in $[X_{min}, X_{max}]$, we choose the one that attains the maximum of $\hat{F}_0(x) - \hat{F}_1(x)$ as $\hat{c}$. If a solution to (3.5) does not exist in the range $[X_{min}, X_{max}]$, we set $\hat{c}$ to be

$$\hat{c} = \arg \max_{x\in\{X_{ij}:\ i=0,1,\ j=1,\ldots,m_i\}} \{\hat{F}_0(x) - \hat{F}_1(x)\}. \tag{3.6}$$

The MELE $\hat{J}$ of $J$ is then given by $\hat{J} = \hat{F}_0(\hat{c}) - \hat{F}_1(\hat{c})$.

We conclude this subsection with a brief discussion on $\hat{c}$. Let $c^*$ be the true value of $c$. According to the proof of Lemma 3.1 in Section 3.5.2, as $n \to \infty$, with probability approaching 1, Equation (3.5) has a solution in the neighbourhood of $c^*$. However, when $n$ is not large and $c^*$ is close to the LLOD, Equation (3.5) may not have a solution in the range $[X_{min}, X_{max}]$. In such situations, Equation (3.6) ensures that $\hat{c}$ is well defined.

### 3.2.2 Asymptotic properties

In this subsection, we study the asymptotic properties of the MELEs $(\hat{J}, \hat{c})$ described in Section 3.2.1. We first introduce some further notation. Let $\boldsymbol{\theta}^*$ be the true value of $\boldsymbol{\theta}$ and

$\omega(x) = \exp\{\boldsymbol{\theta}^{*\top}\mathbf{Q}(x)\}$. Note that $\boldsymbol{Q}(x) = (1, \boldsymbol{q}(x)^\top)^\top$. For $x \ge r$, define

$$\begin{aligned} A_0(x) &= \int_r^x \frac{\omega(x)}{1+\rho\omega(x)} dF_0(x), \\ \boldsymbol{A}_1(x) &= \int_r^x \frac{\omega(x)}{1+\rho\omega(x)} \mathbf{q}(x) dF_0(x), \\ \boldsymbol{A}_2(x) &= \int_r^x \frac{\omega(x)}{1+\rho\omega(x)} \mathbf{q}(x)\mathbf{q}(x)^\top dF_0(x). \end{aligned}$$

Further, let $A_0 = A_0(\infty)$, $\boldsymbol{A}_1 = \boldsymbol{A}_1(\infty)$, $\boldsymbol{A}_2 = \boldsymbol{A}_2(\infty)$, and

$$\boldsymbol{A} = \begin{pmatrix} A_0 & \boldsymbol{A}_1^\top \\ \boldsymbol{A}_1 & \boldsymbol{A}_2 \end{pmatrix}, \quad \boldsymbol{S} = \frac{\rho}{1+\rho}\boldsymbol{A}, \quad \boldsymbol{V} = \boldsymbol{S} - \rho \begin{pmatrix} A_0 \\ \boldsymbol{A}_1 \end{pmatrix} (A_0, \boldsymbol{A}_1^\top).$$

Define $\dot{\mathbf{q}}(x) = d\mathbf{q}(x)/dx$.

**Theorem 3.1.** *Let $J^*$ be the true value of $J$. Suppose the regularity conditions in Section 3.5.2 are satisfied and $c^* > r$. As the total sample size $n = n_0 + n_1$ goes to infinity, we have*
*(a) $\sqrt{n}(\hat{c} - c^*) \to N(0, \sigma_c^2)$ in distribution, where*

$$\sigma_c^2 = \frac{\mathbf{Q}(c^*)^\top \boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{S}^{-1}\mathbf{Q}(c^*)}{\{\boldsymbol{\beta}^{*\top}\dot{\mathbf{q}}(c^*)\}^2} \tag{3.7}$$

*and $\boldsymbol{\beta}^*$ is the true value of $\boldsymbol{\beta}$;*
*(b) $\sqrt{n}(\hat{J} - J^*) \to N(0, \sigma_J^2)$ in distribution, where*

$$\begin{aligned} \sigma_J^2 &= (\rho+1)\{F_0(c^*) - F_0^2(c^*)\} + \frac{\rho+1}{\rho}\{F_1(c^*) - F_1^2(c^*)\} \\ &\quad - \frac{(\rho+1)^3}{\rho}\left\{A_0(c^*) - \begin{pmatrix} A_0(c^*) \\ \boldsymbol{A}_1(c^*) \end{pmatrix}^\top \boldsymbol{A}^{-1} \begin{pmatrix} A_0(c^*) \\ \boldsymbol{A}_1(c^*) \end{pmatrix}\right\}. \end{aligned} \tag{3.8}$$

We provide some comments on the results of Theorem 3.1. Let $(\hat{J}_E, \hat{c}_E)$ and $(\hat{J}_K, \hat{c}_K)$ be the ECDF-based and kernel-based estimators of $(J, c)$, respectively. First, the estimators $\hat{J}$ and $\hat{c}$ both reach the convergence rates of the parametric likelihood based estimators. When there is no LLOD or $r = -\infty$, the convergence rate of $\hat{c}$ is faster than $\hat{c}_E$ and $\hat{c}_K$.

Second, when there is no LLOD or $r = -\infty$, Hsieh and Turnbull (1996) showed that

$$nE\{(\hat{J}_E - J^*)^2\} = \sigma_N^2 + O(n^{-1/3}),\ nE\{(\hat{J}_K - J^*)^2\} = \sigma_N^2 - \gamma n^{-v}\{1 + o(1)\}$$

for some $\gamma > 0$, where

$$\sigma_N^2 = (\rho + 1)\{F_0(c^*) - F_0^2(c^*)\} + \frac{\rho + 1}{\rho}\{F_1(c^*) - F_1^2(c^*)\}.$$

Here the two bandwidths for the kernel method have the order $n^{-v}$ for some $0 < v < 1/3$. According to Theorem 1 in Qin and Zhang (1997), $\sigma_N^2 - \sigma_J^2 \geq 0$. Hence, when $n$ is large, the asymptotic mean square error of $\hat{J}$ is smaller than those of $\hat{J}_E$ and $\hat{J}_K$.

### 3.2.3 Confidence intervals on $J$ and $c$ under the DRM

Replacing $(\boldsymbol{\theta}^*, J^*, c^*, F_0)$ in $\sigma_J^2$ and $\sigma_c^2$ by their respective estimators $(\hat{\boldsymbol{\theta}}, \hat{J}, \hat{c}, \hat{F}_0)$, we obtain the estimators $(\hat{\sigma}_J^2, \hat{\sigma}_c^2)$ for $(\sigma_J^2, \sigma_c^2)$. It can be shown that $\hat{\sigma}_J^2$ and $\hat{\sigma}_c^2$ are both consistent.

**Theorem 3.2.** *Under the conditions of Theorem 3.1, we have*

$$\hat{\sigma}_J^2 \to \sigma_J^2 \quad \text{and} \quad \hat{\sigma}_c^2 \to \sigma_c^2$$

*in probability as $n \to \infty$.*

Because of the asymptotic normality of $\hat{c}$ presented in Theorem 3.1 and the consistency of $\hat{\sigma}_c^2$, the quantity $\sqrt{n}(\hat{c} - c^*)/\hat{\sigma}_c$ is asymptotically pivotal, which leads to the following Wald-type CI for $c$ at level $1 - a$:

$$\mathcal{I}_c = \left[\hat{c} - z_{1-a/2}\hat{\sigma}_c/\sqrt{n}, \hat{c} + z_{1-a/2}\hat{\sigma}_c/\sqrt{n}\right],$$

where $z_{1-a/2}$ is the $100(1 - a/2)$th quantile of the standard normal distribution.

We can similarly construct a Wald-type CI for $J$. However, the Wald-type CI for $J$ based on $\hat{J}$ directly is not range preserving. When $J$ is close to the boundary 0 or 1, the lower or upper bound of the Wald-type CIs could lie outside the range $[0, 1]$. Naturally, we consider a logit transformation on $\hat{J}$ when constructing the CI for $J$. The resulting CI for $J$ is range-preserving. Further, our simulation experience indicates that the logit transformation on $\hat{J}$ leads to a CI for $J$ with better coverage accuracy, especially when $J^*$ is close to 0 or 1. More specifically, using the results in Theorems 3.1 and 3.2, it can be shown that

$$\sqrt{n}\{\text{logit}(\hat{J}) - \text{logit}(J^*)\} \to N\left(0, \frac{\sigma_J^2}{J^{*2}(1-J^*)^2}\right)$$

in distribution as $n \to \infty$, where $\text{logit}(x) = \log\{x/(1-x)\}$ for $0 < x < 1$. Hence $\sqrt{n}\hat{J}(1-\hat{J})\{\text{logit}(\hat{J}) - \text{logit}(J^*)\}/\hat{\sigma}_J$ is also asymptotically pivotal. This suggests the following CI for $J$:

$$\mathcal{I}_J = \left[\text{expit}\left\{\text{logit}(\hat{J}) - \frac{z_{1-a/2}\hat{\sigma}_J}{\sqrt{n}\hat{J}(1-\hat{J})}\right\}, \text{expit}\left\{\text{logit}(\hat{J}) + \frac{z_{1-a/2}\hat{\sigma}_J}{\sqrt{n}\hat{J}(1-\hat{J})}\right\}\right],$$

where $\text{expit}(x) = \exp(x)/\{1+\exp(x)\}$.

## 3.3 Simulation Studies

### 3.3.1 Candidate methods

In this section, we report results from simulation studies to compare the proposed point estimators and CIs of $J$ and $c$ with the following candidate methods.

- The Box-Cox method in Bantis et al. (2019), where the corresponding point estimators and CIs of $J$ and $c$ are denoted as $\hat{J}_B$, $\hat{c}_B$, $\mathcal{I}_{JB}$, and $\mathcal{I}_{cB}$, respectively.

- The ROC-GLM method in Ruopp et al. (2008), where the corresponding point estimators and CIs of $J$ and $c$ are denoted as $\hat{J}_G$, $\hat{c}_G$, $\mathcal{I}_{JG}$, and $\mathcal{I}_{cG}$, respectively.

- The ECDF-based method, where the corresponding point estimators and CIs of $J$ and $c$ are denoted as $\hat{J}_E$, $\hat{c}_E$, $\mathcal{I}_{JE}$, and $\mathcal{I}_{cE}$, respectively.

- The kernel-based method in Bantis et al. (2019), where the corresponding point estimators and CIs of $J$ and $c$ are denoted as $\hat{J}_K$, $\hat{c}_K$, $\mathcal{I}_{JK}$, and $\mathcal{I}_{cK}$, respectively.

- The HCNS method in Bantis et al. (2019), where the corresponding point estimators and CIs of $J$ and $c$ are denoted as $\hat{J}_H$, $\hat{c}_H$, $\mathcal{I}_{JH}$, and $\mathcal{I}_{cH}$, respectively.

For all the above candidate methods, except for $\mathcal{I}_{JB}$, which is obtained via the delta method, the CIs are constructed using the nonparametric bootstrap percentile method. Further details on these methods are deferred to Section 3.5.1.

When there is no LLOD, we compare the proposed method and all the candidate methods listed above. When there is a fixed and finite LLOD, to the best of our knowledge, the kernel-based method and the HCNS method have not been explored in the literature, and hence we do not include these two methods in the comparisons.

### 3.3.2 Simulation setup

We conduct simulation studies under the following two distributional settings from Fluss et al. (2005):

(1) $f_0 \sim Gam(2, 0.5)$ and $f_1 \sim Gam(2, \eta)$;

(2) $f_0 \sim LN(2.5, 0.09)$ and $f_1 \sim LN(\mu, 0.25)$.

Here $Gam(\kappa, \eta)$ denotes the gamma distribution with shape parameter $\kappa$ and rate parameter $\eta$ and $LN(\mu, \sigma^2)$ denotes the lognormal distribution with mean $\mu$ and variance $\sigma^2$, both with respect to the log scale. The proposed estimators are calculated under the correctly specified $\mathbf{q}(x)$. For the setting with gamma distributions, we have $\mathbf{q}(x) = x$, and for the setting with lognormal distributions, we have $\mathbf{q}(x) = (\log x, \log^2 x)^\top$.

For each setting, we choose four values of $\eta$ or $\mu$ such that the corresponding Youden indexes equal 0.2, 0.4, 0.6, and 0.8. The details are given in Table 3.1. For the LLOD, we consider three values: $-\infty$, 15% quantile of $F_0$, and 30% quantile of $F_0$. Note that when the LLOD equals $-\infty$, there is no LLOD. For each simulation scenario, we consider five sample size combinations: $(n_0, n_1) = (50, 50)$, $(100, 100)$, $(200, 200)$, $(150, 50)$, and $(50, 150)$, and results are based on 1,000 repeated simulation runs.

Table 3.1: Parameter values in simulation studies.

| Distribution | $J$ | $\eta/\mu$ | c | $F_0(c)$ | $F_1(c)$ |
|---|---|---|---|---|---|
| Gamma | 0.20 | 0.34 | 4.79 | 0.69 | 0.49 |
| | 0.40 | 0.23 | 5.75 | 0.78 | 0.38 |
| | 0.60 | 0.14 | 7.02 | 0.86 | 0.26 |
| | 0.80 | 0.07 | 9.04 | 0.94 | 0.14 |
| Lognormal | 0.2 | 2.62 | 16.92 | 0.86 | 0.66 |
| | 0.40 | 2.87 | 16.54 | 0.85 | 0.45 |
| | 0.60 | 3.14 | 17.30 | 0.88 | 0.28 |
| | 0.80 | 3.50 | 19.12 | 0.93 | 0.13 |

The simulation results from different simulation scenarios demonstrate similar patterns. To save space, we only report the simulation results under the setting with gamma distribution, without LLOD and with the LLOD equal to the 15% quantile of $F_0$. Other simulation results are provided in Section 3.5.3.

### 3.3.3 Comparison for point estimators

We first examine the point estimators of $(J, c)$. The performance of a point estimator is evaluated through the RB and the MSE. The simulation results are presented in Tables 3.2–3.5.

When there is no LLOD, major observations from Tables 3.2 and 3.3 can be summarized as follows. For estimating the Youden index $J$, the ECDF-based estimator $\hat{J}_E$ has the largest RBs and MSEs in almost all the cases. We also notice that when $J = 0.2$, the RBs of $\hat{J}_B$, $\hat{J}_G$, and $\hat{J}_H$ have greater than 5% RBs, which may not be acceptable, especially when one of $n_0$ and $n_1$ is small. The estimators $\hat{J}$, $\hat{J}_B$, $\hat{J}_G$, and $\hat{J}_K$ have comparable performance in terms of MSE, which are uniformly better than $\hat{J}_E$ and $\hat{J}_H$. When sample sizes are small, the kernel-based estimator $\hat{J}_K$ has slightly smaller MSE than $\hat{J}$; when the sample size increases, the proposed estimator $\hat{J}$ becomes more efficient in terms of MSE. This is in line with our discussion after Theorem 3.1.

For estimating the optimal cutoff point $c$, the proposed estimator $\hat{c}$ outperforms other estimators significantly for the majority of cases. The parametric estimator $\hat{c}_B$ is most competitive. It has larger MSEs than $\hat{c}$ when $J = 0.2, 0.4, 0.6$ and has slightly smaller MSEs than $\hat{c}$ when $J = 0.8$. Among the other four estimators, the estimator $\hat{c}_E$ has the worst performance and $\hat{c}_G$ shows the best performance in most cases. The performances of $\hat{c}_K$ and $\hat{c}_H$ are mixed. There is no obvious trend that one has dominating performance over others.

When the LLOD equals 15% quantile of $F_0$, Tables 3.4 and 3.5 show that the general trend for comparing the proposed method with the Box-Cox method, ROC-GLM method, and ECDF-based method are similar to the case when there is no LLOD. It is worth mentioning that as the LLOD increases, the MSEs of all estimators increase, due to the loss of information under censoring. The estimation of the optimal cutoff point $c$ is more sensitive to the increase of LLOD, especially when $J$ is small.

Table 3.2: RB (%) and MSE (×100) for point estimators of $J$ when there is no LLOD (gamma distributions).

| | $(n_0, n_1)$ | $(50, 50)$ | | $(100, 100)$ | | $(200, 200)$ | | $(50, 150)$ | | $(150, 50)$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $J$ | | RB | MSE | RB | MSE | RB | MSE | RB | MSE | RB | MSE |
| 0.2 | $\hat{J}$ | 4.26 | 0.61 | 1.74 | 0.31 | 0.73 | 0.14 | 3.35 | 0.39 | 2.01 | 0.42 |
| | $\hat{J}_B$ | 8.53 | 0.62 | 4.17 | 0.32 | 2.48 | 0.15 | 6.61 | 0.41 | 5.40 | 0.43 |
| | $\hat{J}_G$ | 10.08 | 0.63 | 5.03 | 0.33 | 2.84 | 0.15 | 8.16 | 0.41 | 6.04 | 0.45 |
| | $\hat{J}_E$ | 40.20 | 1.29 | 26.63 | 0.64 | 17.72 | 0.30 | 33.05 | 0.86 | 32.69 | 0.89 |
| | $\hat{J}_K$ | 9.45 | 0.66 | 5.28 | 0.35 | 3.07 | 0.17 | 6.12 | 0.45 | 8.03 | 0.47 |
| | $\hat{J}_H$ | 17.27 | 0.78 | 7.96 | 0.37 | 5.39 | 0.22 | 12.91 | 0.49 | 9.72 | 0.55 |
| 0.4 | $\hat{J}$ | 2.43 | 0.57 | 0.95 | 0.29 | 0.41 | 0.13 | 1.62 | 0.35 | 1.18 | 0.40 |
| | $\hat{J}_B$ | 4.47 | 0.58 | 2.32 | 0.29 | 1.58 | 0.14 | 3.37 | 0.36 | 2.81 | 0.40 |
| | $\hat{J}_G$ | 3.38 | 0.56 | 1.63 | 0.29 | 1.07 | 0.14 | 2.40 | 0.34 | 2.37 | 0.42 |
| | $\hat{J}_E$ | 16.42 | 1.06 | 10.62 | 0.53 | 6.78 | 0.24 | 13.36 | 0.68 | 12.57 | 0.74 |
| | $\hat{J}_K$ | 2.70 | 0.57 | 1.08 | 0.31 | 0.49 | 0.15 | 1.57 | 0.38 | 1.86 | 0.41 |
| | $\hat{J}_H$ | 4.68 | 0.68 | 1.33 | 0.33 | 0.96 | 0.18 | 3.15 | 0.40 | 1.71 | 0.52 |
| 0.6 | $\hat{J}$ | 1.61 | 0.45 | 0.56 | 0.24 | 0.27 | 0.11 | 0.93 | 0.26 | 0.81 | 0.34 |
| | $\hat{J}_B$ | 2.95 | 0.44 | 1.58 | 0.22 | 1.13 | 0.11 | 2.13 | 0.26 | 1.87 | 0.32 |
| | $\hat{J}_G$ | 1.14 | 0.42 | 0.37 | 0.23 | 0.30 | 0.11 | 0.42 | 0.25 | 1.16 | 0.35 |
| | $\hat{J}_E$ | 8.89 | 0.76 | 5.38 | 0.37 | 3.71 | 0.19 | 7.14 | 0.48 | 6.73 | 0.56 |
| | $\hat{J}_K$ | 0.04 | 0.39 | -0.63 | 0.22 | -0.61 | 0.11 | -0.07 | 0.27 | -0.53 | 0.29 |
| | $\hat{J}_H$ | 1.78 | 0.61 | 0.16 | 0.28 | 0.05 | 0.15 | 0.96 | 0.31 | 0.39 | 0.47 |
| 0.8 | $\hat{J}$ | 1.06 | 0.26 | 0.38 | 0.14 | 0.16 | 0.06 | 0.55 | 0.13 | 0.63 | 0.21 |
| | $\hat{J}_B$ | 1.56 | 0.22 | 0.81 | 0.11 | 0.60 | 0.05 | 1.07 | 0.12 | 0.96 | 0.17 |
| | $\hat{J}_G$ | -0.26 | 0.26 | -0.36 | 0.14 | -0.32 | 0.07 | -0.63 | 0.15 | 0.25 | 0.22 |
| | $\hat{J}_E$ | 4.51 | 0.39 | 2.86 | 0.20 | 1.94 | 0.1 | 3.73 | 0.24 | 3.53 | 0.30 |
| | $\hat{J}_K$ | -2.38 | 0.24 | -2.40 | 0.15 | -2.00 | 0.08 | -1.74 | 0.14 | -2.88 | 0.21 |
| | $\hat{J}_H$ | 1.55 | 0.40 | 0.71 | 0.19 | 0.75 | 0.11 | 1.01 | 0.19 | 0.82 | 0.32 |

Table 3.3: RB (%) and MSE (×100) for point estimators of $c$ when there is no LLOD (gamma distributions).

| | $(n_0, n_1)$ | $(50, 50)$ | | $(100, 100)$ | | $(200, 200)$ | | $(50, 150)$ | | $(150, 50)$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $J$ | | RB | MSE | RB | MSE | RB | MSE | RB | MSE | RB | MSE |
| 0.2 | $\hat{c}$ | -0.39 | 12.22 | -0.15 | 6.09 | -0.03 | 2.78 | -0.40 | 8.56 | -0.02 | 7.68 |
| | $\hat{c}_B$ | -3.03 | 81.16 | -1.97 | 41.56 | -2.33 | 21.17 | -1.83 | 56.07 | -2.85 | 56.79 |
| | $\hat{c}_G$ | -6.33 | 108.4 | -2.55 | 56.42 | -2.20 | 29.56 | -4.64 | 70.96 | -1.79 | 73.30 |
| | $\hat{c}_E$ | -0.92 | 249.25 | 0.49 | 170.90 | -0.76 | 116.8 | -0.78 | 201.93 | 1.89 | 223.79 |
| | $\hat{c}_K$ | 16.99 | 405.52 | 10.39 | 164.68 | 5.99 | 91.84 | 13.63 | 287.86 | 13.00 | 254.68 |
| | $\hat{c}_H$ | 1.75 | 237.90 | 3.16 | 171.47 | 2.71 | 110.91 | -1.09 | 179.73 | 7.6 | 243.96 |
| 0.4 | $\hat{c}$ | -0.32 | 18.88 | -0.13 | 9.35 | 0.00 | 4.27 | -0.42 | 13.68 | 0.02 | 11.05 |
| | $\hat{c}_B$ | -3.25 | 42.84 | -2.37 | 21.78 | -2.22 | 11.46 | -2.58 | 28.62 | -2.53 | 26.45 |
| | $\hat{c}_G$ | -5.73 | 80.28 | -3.26 | 40.06 | -2.15 | 19.16 | -4.16 | 46.09 | -1.75 | 37.58 |
| | $\hat{c}_E$ | -2.89 | 160.26 | -0.23 | 126.83 | -0.45 | 75.26 | -2.66 | 150.21 | 1.51 | 156.93 |
| | $\hat{c}_K$ | 7.67 | 109.91 | 6.77 | 74.57 | 4.68 | 39.14 | 6.38 | 96.15 | 7.36 | 77.39 |
| | $\hat{c}_H$ | 0.40 | 186.56 | 1.30 | 110.77 | 1.12 | 68.91 | -0.36 | 136.04 | 1.29 | 147.84 |
| 0.6 | $\hat{c}$ | -0.35 | 32.29 | -0.18 | 15.91 | 0.01 | 7.38 | -0.53 | 23.77 | 0.06 | 18.20 |
| | $\hat{c}_B$ | -2.75 | 43.66 | -2.04 | 21.86 | -1.76 | 10.88 | -2.45 | 30.09 | -1.89 | 23.40 |
| | $\hat{c}_G$ | -5.87 | 108.83 | -3.71 | 54.12 | -2.30 | 26.28 | -3.94 | 52.16 | -2.82 | 45.86 |
| | $\hat{c}_E$ | -2.29 | 161.37 | -0.67 | 118.08 | 0.16 | 74.77 | -2.40 | 147.73 | 1.22 | 146.35 |
| | $\hat{c}_K$ | 7.57 | 119.52 | 6.43 | 74.85 | 5.08 | 42.21 | 5.93 | 99.47 | 7.62 | 77.68 |
| | $\hat{c}_H$ | 0.40 | 174.95 | 0.78 | 82.95 | 0.60 | 42.69 | 0.16 | 100.73 | 0.25 | 127.29 |
| 0.8 | $\hat{c}$ | -0.41 | 71.84 | -0.31 | 36.11 | 0.02 | 17.53 | -0.80 | 52.89 | 0.20 | 41.00 |
| | $\hat{c}_B$ | -1.53 | 60.09 | -1.11 | 30.31 | -0.78 | 13.91 | -1.71 | 46.46 | -0.66 | 31.67 |
| | $\hat{c}_G$ | -7.14 | 237.07 | -4.14 | 111.72 | -2.58 | 56.25 | -3.34 | 99.57 | -4.87 | 98.48 |
| | $\hat{c}_E$ | -3.17 | 236.85 | -1.72 | 159.92 | -1.24 | 107.54 | -4.03 | 195.30 | 1.09 | 197.79 |
| | $\hat{c}_K$ | 7.09 | 184.22 | 6.21 | 113.61 | 5.07 | 65.86 | 5.56 | 152.15 | 7.02 | 107.26 |
| | $\hat{c}_H$ | -2.30 | 169.65 | -2.76 | 104.87 | -2.90 | 56.83 | -3.58 | 116.50 | -1.74 | 146.35 |

Table 3.4: RB (%) and MSE (×100) for point estimators of $J$ when the LLOD equals 15% quantile of $F_0$ (gamma distributions).

| $J$ | $(n_0, n_1)$ | $(50,50)$ RB | MSE | $(100,100)$ RB | MSE | $(200,200)$ RB | MSE | $(50,150)$ RB | MSE | $(150,50)$ RB | MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\hat{J}$ | 5.92 | 0.61 | 2.49 | 0.31 | 1.11 | 0.14 | 4.62 | 0.40 | 2.85 | 0.42 |
| | $\hat{J}_B$ | 9.59 | 0.64 | 4.75 | 0.32 | 2.85 | 0.15 | 7.38 | 0.41 | 5.85 | 0.43 |
| | $\hat{J}_G$ | 7.58 | 0.69 | 2.67 | 0.36 | -0.34 | 0.17 | 5.99 | 0.46 | 3.50 | 0.50 |
| | $\hat{J}_E$ | 40.17 | 1.29 | 26.63 | 0.64 | 17.72 | 0.30 | 33.04 | 0.86 | 32.69 | 0.89 |
| 0.4 | $\hat{J}$ | 2.69 | 0.57 | 1.10 | 0.30 | 0.51 | 0.13 | 1.88 | 0.36 | 1.31 | 0.40 |
| | $\hat{J}_B$ | 4.92 | 0.59 | 2.59 | 0.29 | 1.75 | 0.14 | 3.59 | 0.36 | 3.12 | 0.41 |
| | $\hat{J}_G$ | 1.50 | 0.61 | -0.21 | 0.32 | -1.20 | 0.15 | 0.68 | 0.39 | -0.01 | 0.46 |
| | $\hat{J}_E$ | 16.42 | 1.06 | 10.62 | 0.53 | 6.78 | 0.24 | 13.36 | 0.68 | 12.57 | 0.74 |
| 0.6 | $\hat{J}$ | 1.74 | 0.46 | 0.6 | 0.24 | 0.29 | 0.11 | 1.04 | 0.27 | 0.85 | 0.34 |
| | $\hat{J}_B$ | 3.17 | 0.44 | 1.70 | 0.22 | 1.19 | 0.11 | 2.16 | 0.27 | 2.09 | 0.32 |
| | $\hat{J}_G$ | 0.11 | 0.47 | -0.64 | 0.26 | -1.04 | 0.12 | -0.56 | 0.29 | -0.35 | 0.38 |
| | $\hat{J}_E$ | 8.89 | 0.76 | 5.38 | 0.37 | 3.71 | 0.19 | 7.14 | 0.48 | 6.73 | 0.56 |
| 0.8 | $\hat{J}$ | 1.10 | 0.26 | 0.40 | 0.14 | 0.16 | 0.06 | 0.58 | 0.14 | 0.63 | 0.21 |
| | $\hat{J}_B$ | 1.58 | 0.22 | 0.80 | 0.11 | 0.56 | 0.05 | 1.01 | 0.12 | 1.02 | 0.17 |
| | $\hat{J}_G$ | -0.60 | 0.30 | -0.65 | 0.16 | -0.80 | 0.08 | -0.94 | 0.17 | -0.39 | 0.24 |
| | $\hat{J}_E$ | 4.51 | 0.39 | 2.86 | 0.20 | 1.94 | 0.10 | 3.73 | 0.24 | 3.53 | 0.30 |

Table 3.5: RB (%) and MSE (×100) for point estimators of $c$ when the LLOD equals 15% quantile of $F_0$ (gamma distributions).

| $J$ | $(n_0, n_1)$ | $(50,50)$ RB | MSE | $(100,100)$ RB | MSE | $(200,200)$ RB | MSE | $(50,150)$ RB | MSE | $(150,50)$ RB | MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\hat{c}$ | -1.58 | 60.04 | -0.86 | 24.25 | -0.44 | 10.84 | -1.53 | 38.67 | -0.60 | 66.67 |
| | $\hat{c}_B$ | 1.04 | 148.39 | 0.49 | 51.29 | -0.30 | 24.87 | 0.29 | 66.35 | -0.07 | 67.79 |
| | $\hat{c}_G$ | 6.30 | 168.58 | 8.65 | 81.29 | 8.51 | 48.7 | 6.72 | 90.4 | 9.00 | 102.15 |
| | $\hat{c}_E$ | -0.65 | 257.51 | 0.49 | 170.8 | -0.76 | 116.8 | -0.79 | 202.38 | 1.89 | 223.79 |
| 0.4 | $\hat{c}$ | -0.59 | 23.70 | -0.24 | 11.48 | -0.04 | 5.58 | -0.67 | 15.51 | -0.15 | 15.28 |
| | $\hat{c}_B$ | -0.82 | 47.68 | -0.54 | 24.65 | -0.57 | 12.19 | -1.06 | 31.41 | -0.33 | 29.38 |
| | $\hat{c}_G$ | 2.35 | 75.60 | 5.21 | 47.69 | 5.98 | 30.66 | 4.04 | 51.81 | 6.34 | 54.94 |
| | $\hat{c}_E$ | -2.89 | 160.26 | -0.23 | 126.83 | -0.45 | 75.26 | -2.66 | 150.21 | 1.51 | 156.93 |
| 0.6 | $\hat{c}$ | -0.39 | 32.90 | -0.22 | 16.35 | -0.01 | 7.71 | -0.63 | 23.81 | 0.01 | 19.68 |
| | $\hat{c}_B$ | -1.05 | 48.30 | -0.81 | 23.92 | -0.64 | 11.46 | -1.48 | 31.25 | -0.27 | 26.75 |
| | $\hat{c}_G$ | 0.18 | 91.92 | 2.33 | 51.66 | 3.77 | 29.23 | 2.04 | 52.36 | 3.34 | 50.15 |
| | $\hat{c}_E$ | -2.29 | 161.37 | -0.67 | 118.08 | 0.16 | 74.77 | -2.40 | 147.73 | 1.22 | 146.35 |
| 0.8 | $\hat{c}$ | -0.45 | 72.29 | -0.34 | 36.11 | 0.02 | 17.60 | -0.87 | 53.45 | 0.17 | 41.41 |
| | $\hat{c}_B$ | -0.65 | 66.17 | -0.46 | 32.77 | -0.17 | 15.11 | -1.24 | 48.59 | 0.25 | 35.50 |
| | $\hat{c}_G$ | -3.53 | 193.47 | -0.26 | 92.54 | 1.59 | 52.64 | 0.28 | 91.43 | -0.43 | 99.37 |
| | $\hat{c}_E$ | -3.17 | 236.85 | -1.72 | 159.92 | -1.24 | 107.54 | -4.03 | 195.30 | 1.09 | 197.79 |

### 3.3.4 Comparison for confidence intervals

We now examine the behaviour of 95% CIs of $J$ and $c$. The performance of a CI is evaluated by the CP and the AL. The simulation results are presented in Tables 3.6–3.9.

Table 3.6: CP (%) and AL for CIs of $J$ when there is no LLOD (gamma distributions).

| | $(n_0, n_1)$ | (50, 50) | | (100, 100) | | (200, 200) | | (50, 150) | | (150, 50) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $J$ | | CP | AL | CP | AL | CP | AL | CP | AL | CP | AL |
| 0.2 | $\mathcal{I}_J$ | 94.1 | 0.32 | 94.9 | 0.22 | 96.4 | 0.15 | 93.6 | 0.25 | 94.7 | 0.26 |
| | $\mathcal{I}_{JB}$ | 94.5 | 0.30 | 93.9 | 0.21 | 95.2 | 0.15 | 92.7 | 0.24 | 94.6 | 0.25 |
| | $\mathcal{I}_{JG}$ | 92.6 | 0.29 | 93.7 | 0.21 | 94.7 | 0.15 | 91.9 | 0.23 | 93.7 | 0.25 |
| | $\mathcal{I}_{JE}$ | 70.2 | 0.30 | 72.1 | 0.22 | 78.0 | 0.16 | 66.7 | 0.24 | 72.6 | 0.26 |
| | $\mathcal{I}_{JK}$ | 93.6 | 0.30 | 93.7 | 0.23 | 94.7 | 0.16 | 92.3 | 0.25 | 93.7 | 0.26 |
| | $\mathcal{I}_{JH}$ | 92.1 | 0.31 | 94.0 | 0.23 | 94.1 | 0.17 | 91.1 | 0.25 | 93.3 | 0.27 |
| 0.4 | $\mathcal{I}_J$ | 95.4 | 0.28 | 94.9 | 0.20 | 96.0 | 0.14 | 93.3 | 0.22 | 94.3 | 0.24 |
| | $\mathcal{I}_{JB}$ | 95.1 | 0.28 | 94.2 | 0.20 | 95.3 | 0.14 | 93.1 | 0.22 | 94.2 | 0.24 |
| | $\mathcal{I}_{JG}$ | 93.1 | 0.29 | 93.6 | 0.20 | 94.7 | 0.14 | 91.6 | 0.22 | 92.9 | 0.25 |
| | $\mathcal{I}_{JE}$ | 79.2 | 0.30 | 80.5 | 0.22 | 83.5 | 0.16 | 75.1 | 0.23 | 80.4 | 0.26 |
| | $\mathcal{I}_{JK}$ | 93.7 | 0.29 | 93.1 | 0.21 | 95.1 | 0.15 | 92.7 | 0.23 | 93.5 | 0.24 |
| | $\mathcal{I}_{JH}$ | 93.7 | 0.31 | 95.0 | 0.22 | 94.7 | 0.16 | 93.3 | 0.23 | 96.5 | 0.28 |
| 0.6 | $\mathcal{I}_J$ | 95.6 | 0.25 | 94.3 | 0.18 | 95.0 | 0.13 | 94.2 | 0.19 | 94.0 | 0.22 |
| | $\mathcal{I}_{JB}$ | 95.8 | 0.25 | 94.2 | 0.18 | 95.1 | 0.13 | 93.8 | 0.19 | 94.2 | 0.22 |
| | $\mathcal{I}_{JG}$ | 93.3 | 0.26 | 93.5 | 0.18 | 94.8 | 0.13 | 92.5 | 0.19 | 92.6 | 0.23 |
| | $\mathcal{I}_{JE}$ | 79.4 | 0.26 | 80.7 | 0.19 | 85.3 | 0.14 | 75.7 | 0.20 | 81.8 | 0.23 |
| | $\mathcal{I}_{JK}$ | 94.4 | 0.24 | 94.1 | 0.17 | 95.4 | 0.13 | 92.4 | 0.19 | 93.5 | 0.21 |
| | $\mathcal{I}_{JH}$ | 94.0 | 0.30 | 95.3 | 0.21 | 95.3 | 0.15 | 92.6 | 0.21 | 95.8 | 0.27 |
| 0.8 | $\mathcal{I}_J$ | 95.1 | 0.20 | 94.0 | 0.14 | 95.0 | 0.10 | 95.1 | 0.14 | 95.9 | 0.18 |
| | $\mathcal{I}_{JB}$ | 96.0 | 0.18 | 94.1 | 0.13 | 95.3 | 0.09 | 94.1 | 0.13 | 95.2 | 0.16 |
| | $\mathcal{I}_{JG}$ | 92.7 | 0.20 | 94.0 | 0.14 | 95.2 | 0.10 | 93.3 | 0.14 | 92.2 | 0.18 |
| | $\mathcal{I}_{JE}$ | 81.6 | 0.19 | 85.5 | 0.14 | 85.9 | 0.10 | 76.1 | 0.14 | 83.5 | 0.17 |
| | $\mathcal{I}_{JK}$ | 95.0 | 0.17 | 91.8 | 0.13 | 91.0 | 0.09 | 93.6 | 0.13 | 92.3 | 0.15 |
| | $\mathcal{I}_{JH}$ | 86.9 | 0.23 | 95.8 | 0.17 | 95.5 | 0.12 | 92.8 | 0.17 | 92.2 | 0.22 |

We first summarize the findings on the CIs for the Youden index $J$. The CPs of $\mathcal{I}_{JE}$ have low coverage probabilities and are not acceptable regardless of the value of LLODs. The proposed CI and the CI based on the Box-Cox method, $\mathcal{I}_J$ and $\mathcal{I}_{JB}$, have comparable and most stable performance in almost all cases. The ROC-GLM based CI, $\mathcal{I}_{JG}$, performs quite well in general but has undercoverage issues in some cases. When there is no LLOD, the two confidence intervals $\mathcal{I}_{JK}$ and $\mathcal{I}_{JH}$ have similar issues as $\mathcal{I}_{JG}$ with undercoverage problems.

We next discuss the findings on the CIs for the optimal cutoff point $c$. When there is no

Table 3.7: CP (%) and AL for CIs of $c$ when there is no LLOD (gamma distributions).

| | $(n_0, n_1)$ | $(50, 50)$ | | $(100, 100)$ | | $(200, 200)$ | | $(50, 150)$ | | $(150, 50)$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $J$ | | CP | AL | CP | AL | CP | AL | CP | AL | CP | AL |
| 0.2 | $\mathcal{I}_c$ | 94.1 | 1.31 | 94.1 | 0.94 | 95.2 | 0.67 | 94.2 | 1.11 | 94.8 | 1.05 |
| | $\mathcal{I}_{cB}$ | 92.8 | 7.26 | 93.9 | 2.64 | 93.8 | 1.77 | 93.7 | 3.25 | 92.6 | 3.25 |
| | $\mathcal{I}_{cG}$ | 94.0 | 4.50 | 94.7 | 3.02 | 94.3 | 2.07 | 93.7 | 3.43 | 94.8 | 3.47 |
| | $\mathcal{I}_{cE}$ | 97.0 | 5.36 | 96.5 | 4.65 | 96.8 | 3.89 | 95.6 | 4.83 | 97.2 | 5.05 |
| | $\mathcal{I}_{cK}$ | 95.6 | 7.02 | 94.3 | 5.70 | 95.0 | 3.89 | 95.2 | 6.68 | 92.6 | 6.36 |
| | $\mathcal{I}_{cH}$ | 97.2 | 5.98 | 97.9 | 4.79 | 95.8 | 3.67 | 96.4 | 4.59 | 95.4 | 5.80 |
| 0.4 | $\mathcal{I}_c$ | 93.9 | 1.63 | 94.5 | 1.16 | 95.6 | 0.82 | 93.3 | 1.40 | 94.5 | 1.27 |
| | $\mathcal{I}_{cB}$ | 92.0 | 2.38 | 92.0 | 1.72 | 92.5 | 1.23 | 92.5 | 1.98 | 92.1 | 1.83 |
| | $\mathcal{I}_{cG}$ | 91.5 | 3.13 | 94.0 | 2.30 | 93.4 | 1.63 | 92.2 | 2.36 | 93.4 | 2.31 |
| | $\mathcal{I}_{cE}$ | 95.1 | 4.52 | 96.6 | 3.80 | 95.0 | 3.18 | 94.2 | 4.02 | 96.8 | 4.32 |
| | $\mathcal{I}_{cK}$ | 93.9 | 4.23 | 93.2 | 2.91 | 93.4 | 2.19 | 94.3 | 3.55 | 92.5 | 3.13 |
| | $\mathcal{I}_{cH}$ | 97.0 | 4.87 | 95.6 | 3.83 | 94.6 | 2.92 | 94.4 | 3.97 | 97.1 | 4.58 |
| 0.6 | $\mathcal{I}_c$ | 93.2 | 2.13 | 94.4 | 1.52 | 95.0 | 1.08 | 93.4 | 1.85 | 94.0 | 1.63 |
| | $\mathcal{I}_{cB}$ | 91.5 | 2.36 | 92.1 | 1.70 | 92.2 | 1.21 | 91.4 | 1.99 | 92.3 | 1.76 |
| | $\mathcal{I}_{cG}$ | 90.2 | 3.57 | 90.8 | 2.59 | 92.4 | 1.86 | 90.9 | 2.57 | 91.1 | 2.44 |
| | $\mathcal{I}_{cE}$ | 94.5 | 4.43 | 95.7 | 3.76 | 97.0 | 3.05 | 92.2 | 3.96 | 95.9 | 4.21 |
| | $\mathcal{I}_{cK}$ | 92.7 | 3.60 | 90.9 | 2.73 | 90.0 | 2.04 | 92.7 | 3.31 | 87.8 | 2.65 |
| | $\mathcal{I}_{cH}$ | 96.7 | 4.80 | 96.6 | 3.76 | 96.4 | 2.78 | 94.7 | 3.80 | 97.6 | 4.50 |
| 0.8 | $\mathcal{I}_c$ | 92.9 | 3.25 | 94.3 | 2.31 | 95.0 | 1.65 | 93.8 | 2.78 | 93.7 | 2.46 |
| | $\mathcal{I}_{cB}$ | 91.9 | 2.87 | 93.3 | 2.03 | 94.5 | 1.44 | 90.6 | 2.49 | 93.8 | 2.14 |
| | $\mathcal{I}_{cG}$ | 87.8 | 5.18 | 91.5 | 3.81 | 90.0 | 2.71 | 92.0 | 3.65 | 86.6 | 3.27 |
| | $\mathcal{I}_{cE}$ | 88.8 | 4.93 | 93.7 | 4.28 | 93.9 | 3.55 | 84.5 | 4.42 | 94.5 | 4.78 |
| | $\mathcal{I}_{cK}$ | 91.3 | 4.34 | 90.9 | 3.29 | 89.7 | 2.49 | 92.1 | 4.06 | 88.8 | 3.16 |
| | $\mathcal{I}_{cH}$ | 95.6 | 4.74 | 97.3 | 3.99 | 94.8 | 2.91 | 93.6 | 3.97 | 98.7 | 4.74 |

LLOD, the proposed CI $\mathcal{I}_c$ has the most stable performance and its CPs are reasonably close to 95% in almost all scenarios. The CPs of $\mathcal{I}_{cE}$ fluctuate around the nominal level 95% while undercoverage problems are associated with the other four CIs $\mathcal{I}_{cB}$, $\mathcal{I}_{cG}$, $\mathcal{I}_{cK}$, and $\mathcal{I}_{cH}$. When there is a fixed and finite LLOD, the ALs of all CIs increase. The proposed CI $\mathcal{I}_c$ and the ECDF-based CI $\mathcal{I}_{cE}$ tend to have an issue with overcoverage, while the CI based on the Box-Cox method has severe undercoverage problem and the ROC-GLM based CI $\mathcal{I}_{cG}$ also has the same issue for some cases. When $J = 0.4, 0.6, 0.8$, the proposed CI $\mathcal{I}_c$ becomes quite stable in almost all cases. The performance of $\mathcal{I}_{cB}$ improves as $J$ increases. The CPs of $\mathcal{I}_{cG}$ are reasonably close to the nominal level. However, $\mathcal{I}_{cG}$ has longer ALs compared to $\mathcal{I}_c$.

Table 3.8: CP (%) and AL for CIs of $J$ when the LLOD equals 15% quantile of $F_0$ (gamma distributions).

| $J$ | $(n_0, n_1)$ | $(50, 50)$ CP | AL | $(100, 100)$ CP | AL | $(200, 200)$ CP | AL | $(50, 150)$ CP | AL | $(150, 50)$ CP | AL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\mathcal{I}_J$ | 93.4 | 0.31 | 94.2 | 0.21 | 96.0 | 0.15 | 93.8 | 0.25 | 94.3 | 0.25 |
| | $\mathcal{I}_{JB}$ | 94.3 | 0.30 | 93.9 | 0.21 | 95.2 | 0.15 | 92.6 | 0.24 | 94.5 | 0.25 |
| | $\mathcal{I}_{JG}$ | 93.2 | 0.31 | 94.0 | 0.23 | 94.7 | 0.16 | 92.9 | 0.25 | 93.2 | 0.26 |
| | $\mathcal{I}_{JE}$ | 64.8 | 0.28 | 68.2 | 0.21 | 73.4 | 0.15 | 61.4 | 0.23 | 65.0 | 0.24 |
| 0.4 | $\mathcal{I}_J$ | 95.3 | 0.28 | 94.9 | 0.20 | 95.8 | 0.15 | 93.7 | 0.22 | 94.5 | 0.24 |
| | $\mathcal{I}_{JB}$ | 95.4 | 0.28 | 94.0 | 0.20 | 95.4 | 0.14 | 93.3 | 0.22 | 94.6 | 0.24 |
| | $\mathcal{I}_{JG}$ | 94.6 | 0.30 | 94.8 | 0.22 | 95.2 | 0.15 | 92.7 | 0.24 | 93.3 | 0.26 |
| | $\mathcal{I}_{JE}$ | 77.4 | 0.28 | 78.5 | 0.21 | 81.7 | 0.15 | 73.8 | 0.22 | 78.1 | 0.25 |
| 0.6 | $\mathcal{I}_J$ | 95.7 | 0.26 | 94.6 | 0.18 | 94.7 | 0.13 | 94.0 | 0.19 | 94.2 | 0.22 |
| | $\mathcal{I}_{JB}$ | 95.7 | 0.25 | 94.2 | 0.18 | 94.7 | 0.13 | 93.8 | 0.19 | 94.2 | 0.22 |
| | $\mathcal{I}_{JG}$ | 94.3 | 0.27 | 93.9 | 0.19 | 94.5 | 0.14 | 92.9 | 0.20 | 93.4 | 0.23 |
| | $\mathcal{I}_{JE}$ | 77.6 | 0.25 | 79.3 | 0.19 | 82.8 | 0.14 | 74.3 | 0.19 | 80.3 | 0.22 |
| 0.8 | $\mathcal{I}_J$ | 95.5 | 0.20 | 94.2 | 0.14 | 95.3 | 0.10 | 94.9 | 0.14 | 95.7 | 0.18 |
| | $\mathcal{I}_{JB}$ | 96.2 | 0.18 | 94.8 | 0.13 | 94.6 | 0.09 | 94.6 | 0.13 | 95.2 | 0.16 |
| | $\mathcal{I}_{JG}$ | 93.7 | 0.21 | 93.5 | 0.15 | 95.4 | 0.11 | 94.3 | 0.15 | 93.8 | 0.19 |
| | $\mathcal{I}_{JE}$ | 80.2 | 0.18 | 85.3 | 0.14 | 86.4 | 0.10 | 75.1 | 0.14 | 83.1 | 0.17 |

Table 3.9: CP (%) and AL for CIs of $c$ when the LLOD equals 15% quantile of $F_0$ (gamma distributions).

| $J$ | $(n_0, n_1)$ | $(50, 50)$ CP | AL | $(100, 100)$ CP | AL | $(200, 200)$ CP | AL | $(50, 150)$ CP | AL | $(150, 50)$ CP | AL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\mathcal{I}_c$ | 97.5 | 5.97 | 97.5 | 1.95 | 96.5 | 1.26 | 96.9 | 2.73 | 96.8 | 4.02 |
| | $\mathcal{I}_{cB}$ | 83.6 | 5.06 | 83.7 | 2.10 | 81.8 | 1.43 | 77.4 | 2.35 | 84.8 | 2.74 |
| | $\mathcal{I}_{cG}$ | 95.3 | 4.83 | 93.1 | 3.27 | 89.5 | 2.23 | 91.1 | 3.80 | 91.9 | 3.77 |
| | $\mathcal{I}_{cE}$ | 96.6 | 5.07 | 95.4 | 4.41 | 96.9 | 3.75 | 94.6 | 4.55 | 96.4 | 4.82 |
| 0.4 | $\mathcal{I}_c$ | 95.3 | 1.87 | 95.0 | 1.32 | 95.7 | 0.93 | 94.0 | 1.53 | 95.5 | 1.52 |
| | $\mathcal{I}_{cB}$ | 86.8 | 2.19 | 87.9 | 1.58 | 87.8 | 1.12 | 83.5 | 1.58 | 91.3 | 1.84 |
| | $\mathcal{I}_{cG}$ | 94.3 | 3.20 | 92.7 | 2.36 | 87.3 | 1.68 | 91.2 | 2.48 | 92.0 | 2.41 |
| | $\mathcal{I}_{cE}$ | 95.1 | 4.43 | 95.9 | 3.74 | 95.3 | 3.14 | 93.6 | 3.93 | 97.3 | 4.27 |
| 0.6 | $\mathcal{I}_c$ | 93.4 | 2.17 | 94.3 | 1.55 | 95.3 | 1.10 | 93.5 | 1.85 | 94.3 | 1.70 |
| | $\mathcal{I}_{cB}$ | 91.0 | 2.39 | 91.3 | 1.71 | 92.1 | 1.22 | 87.8 | 1.81 | 93.0 | 1.91 |
| | $\mathcal{I}_{cG}$ | 94.0 | 3.59 | 94.2 | 2.62 | 92.3 | 1.88 | 93.7 | 2.65 | 93.6 | 2.51 |
| | $\mathcal{I}_{cE}$ | 94.4 | 4.40 | 95.5 | 3.72 | 96.5 | 3.04 | 91.6 | 3.94 | 95.9 | 4.17 |
| 0.8 | $\mathcal{I}_c$ | 92.9 | 3.24 | 94.4 | 2.31 | 94.9 | 1.65 | 93.2 | 2.78 | 93.7 | 2.48 |
| | $\mathcal{I}_{cB}$ | 92.8 | 2.99 | 94.9 | 2.14 | 94.7 | 1.52 | 91.1 | 2.51 | 94.2 | 2.27 |
| | $\mathcal{I}_{cG}$ | 92.5 | 5.28 | 95.0 | 3.79 | 94.8 | 2.71 | 94.2 | 3.74 | 92.8 | 3.39 |
| | $\mathcal{I}_{cE}$ | 88.5 | 4.90 | 93.5 | 4.26 | 94.3 | 3.54 | 84.7 | 4.39 | 94.3 | 4.76 |

## 3.4 Real Data Analysis

In this section, we illustrate the performance of the proposed method by analyzing a dataset on Duchenne Muscular Dystrophy (DMD). The DMD is a genetic disorder characterized by progressive muscle degeneration and weakness. A particular gene on the X chromosome, when mutated, leads to DMD. This disease is transmitted from a mother to her children genetically. Affected male offsprings usually develop the disease and die at a young age while the mutated gene does not affect the health of female offsprings. Therefore, detection of potential affected females is of great interest.

Percy et al. (1982) pointed out that carriers of DMD tend to exhibit high levels of certain biomarkers even though they do not show any symptoms. Andrews and Herzberg (2012) collected the complete data of four biomarkers, namely, creatine kinase (CK), hemopexin (H), lactate dehydroginase (LD), and pyruvate kinase (PK), from the blood serum samples of a healthy group of people ($n_0 = 127$) and a group of carriers ($n_1 = 67$). Our goal is to choose the most appropriate biomarker to distinguish healthy individuals from diseased ones.

We choose $\mathbf{q}(x) = x$ in the proposed method for each biomarker, which is equivalent to assuming a logistic regression model for an individual's disease status and the biomarker (Qin and Zhang, 1997). Table 3.10 presents Qin and Zhang (1997)'s test statistics along with the p-values for the goodness-of-fit of the DRM in (3.2) with $\mathbf{q}(x) = x$. It shows that for each biomarker, the data does not provide evidence to reject the DRM in (3.2) with $\mathbf{q}(x) = x$.

Table 3.10: Qin and Zhang (1997)'s test statistics and their p-values when $\mathbf{q}(x) = x$.

| Biomarker | CK | LD | PH | H |
|---|---|---|---|---|
| Test statistic | 0.138 | 0.247 | 0.226 | 0.222 |
| $P$-value | 0.912 | 0.291 | 0.507 | 0.676 |

Table 3.11 provides the point estimates and the CIs (in parentheses) from the proposed method and all the competitive methods listed in Section 3.3. As we can see, for all biomarkers, the point estimates of Youden index are similar for all methods: they differ only in the second digit. For the CIs of the Youden index, the methods with $\hat{J}$, $\hat{J}_B$, $\hat{J}_G$, and $\hat{J}_K$ have similar performances for all biomarkers; the CIs with $\hat{J}_E$ and $\hat{J}_H$ tend to be wider than other four methods. For the optimal cutoff point, the point estimates have substantial differences, especially for the biomarker LD, compared with the estimates of the Youden index. For all biomarkers, the proposed method has the shortest CIs, while the ECDF-based method and HCNS method tend to have the widest CIs. The performances

of other three CIs are mixed: the CI based on the Box-Cox method has shorter length for biomarkers CK and LD, while the CIs based on ROC-GLM and kernel methods have shorter length for biomarkers PK and H. Furthermore, we find that the biomarker CK gives the largest estimated Youden index which is around 0.6. Therefore, the biomarker CK performs the best among these four biomarkers to distinguish the diseased individuals and the healthy ones. The estimated optimal cutoff point for the biomarker CK using the proposed method is 61.13 with the 95% CI being $(54.59, 67.68)$.

Table 3.11: Estimation of the Youden index and the optimal cutoff point with the DMD dataset.

| | CK | LD | PK | H |
|---|---|---|---|---|
| $\hat{J}$ | 0.59 (0.48, 0.69) | 0.55 (0.45, 0.65) | 0.49 (0.38, 0.59) | 0.36 (0.26, 0.48) |
| $\hat{J}_B$ | 0.62 (0.51, 0.70) | 0.56 (0.46, 0.66) | 0.48 (0.37, 0.58) | 0.37 (0.26, 0.48) |
| $\hat{J}_G$ | 0.60 (0.50, 0.71) | 0.57 (0.47, 0.68) | 0.48 (0.38, 0.61) | 0.39 (0.29, 0.50) |
| $\hat{J}_E$ | 0.61 (0.52, 0.73) | 0.58 (0.50, 0.72) | 0.51 (0.42, 0.65) | 0.42 (0.34, 0.57) |
| $\hat{J}_K$ | 0.59 (0.51, 0.67) | 0.55 (0.45, 0.66) | 0.47 (0.37, 0.58) | 0.37 (0.25, 0.49) |
| $\hat{J}_H$ | 0.61 (0.52, 0.80) | 0.57 (0.46, 0.70) | 0.48 (0.35, 0.62) | 0.40 (0.31, 0.56) |
| $\hat{c}$ | 61.13 (54.59, 67.68) | 198.56 (190.34, 206.78) | 15.54 (14.65, 16.43) | 87.74 (86.09, 89.39) |
| $\hat{c}_B$ | 58.01 (51.17, 65.42) | 200.01 (188.99, 209.41) | 16.56 (14.83, 18.24) | 86.73 (83.59, 89.35) |
| $\hat{c}_G$ | 55.60 (48.83, 68.41) | 197.54 (183.47, 211.64) | 15.81 (14.58, 16.79) | 85.25 (82.31, 87.90) |
| $\hat{c}_E$ | 56.00 (43.00, 75.00) | 187.00 (181.00, 232.00) | 16.60 (14.00, 18.20) | 87.20 (80.50, 88.50) |
| $\hat{c}_K$ | 73.36 (54.15, 79.16) | 202.32 (188.31, 216.94) | 17.22 (15.87, 18.28) | 85.52 (82.84, 88.36) |
| $\hat{c}_H$ | 52.02 (43.01, 68.50) | 202.92 (179.20, 221.22) | 14.37 (12.34, 18.05) | 82.90 (80.26, 92.10) |

## 3.5 Technical Details and Additional Simulation Results

### 3.5.1 Review of existing methods

In simulation studies, we compare the proposed method with five candidate methods: the Box-Cox method (Bantis et al., 2019), the ROC-GLM method (Ruopp et al., 2008), the ECDF-based method, the kernel-based method (Bantis et al., 2019), and the HCNS method (Bantis et al., 2019). In the following, we provide detailed review of each method.

#### Methods for data without a LLOD

We start with the case when there is no LLOD.

*The Box-Cox method*

For any $x > 0$, define the Box-Cox transformation of $x$ as

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log x, & \lambda = 0 \end{cases}.$$

The Box-Cox method assumes that

$$X_{01}^{(\lambda)}, \ldots, X_{0n_0}^{(\lambda)} \sim N(\mu_0, \sigma_0^2) \quad \text{and} \quad X_{11}^{(\lambda)}, \ldots, X_{1n_1}^{(\lambda)} \sim N(\mu_1, \sigma_1^2). \tag{3.9}$$

Under (3.9), we have

$$f_0(x) = \frac{\lambda x^{\lambda-1}}{\sigma_0} \phi\left(\frac{x^{(\lambda)} - \mu_0}{\sigma_0}\right), \quad f_1(x) = \frac{\lambda x^{\lambda-1}}{\sigma_1} \phi\left(\frac{x^{(\lambda)} - \mu_1}{\sigma_1}\right), \tag{3.10}$$

and

$$F_0(x) = \Phi\left(\frac{x^{(\lambda)} - \mu_0}{\sigma_0}\right), \quad F_1(x) = \Phi\left(\frac{x^{(\lambda)} - \mu_1}{\sigma_1}\right),$$

where $\phi(x)$ and $\Phi(x)$ are the probability density function and CDF of $N(0,1)$, respectively.

With (3.10) and the available data, Bantis et al. (2019) suggest to estimate unknown parameters $(\lambda, \mu_0, \sigma_0, \mu_1, \sigma_1)$ by the maximum likelihood method, where the corresponding estimators are denoted by $(\hat{\lambda}, \hat{\mu}_0, \hat{\sigma}_0, \hat{\mu}_1, \hat{\sigma}_1)$. The Youden index $J$ and the optimal cutoff point $c$ are then estimated by

$$\begin{aligned} \hat{J}_B &= \max_{x>0} \left\{ \Phi\left(\frac{x^{(\hat{\lambda})} - \hat{\mu}_0}{\hat{\sigma}_0}\right) - \Phi\left(\frac{x^{(\hat{\lambda})} - \hat{\mu}_1}{\hat{\sigma}_1}\right) \right\}, \\ \hat{c}_B &= \arg\max_{x>0} \left\{ \Phi\left(\frac{x^{(\hat{\lambda})} - \hat{\mu}_0}{\hat{\sigma}_0}\right) - \Phi\left(\frac{x^{(\hat{\lambda})} - \hat{\mu}_1}{\hat{\sigma}_1}\right) \right\}. \end{aligned}$$

*The ROC-GLM method*

Let $S_0 = 1 - F_0$ and $S_1 = 1 - F_1$ be the survival functions for the healthy population and the diseased population, respectively. For any $x \in [0,1]$, the ROC curve is given by

$$ROC(x) = P\big(S_0(X_1) \leq x\big),$$

where $X_1$ is a random variable from diseased distribution $F_1$. The ROC-GLM method assumes a parametric model for $ROC(x)$:

$$\Phi^{-1}\big(ROC(x)\big) = a_0 + a_1\Phi^{-1}(x), \tag{3.11}$$

where $a_0$ and $a_1$ are unknown parameters (Ruopp et al., 2008).

Let

$$\hat{F}_{E0}(x) = \frac{1}{n_0}\sum_{j=1}^{n_0} I(X_{0j} \le x) \quad \text{and} \quad \hat{F}_{E1}(x) = \frac{1}{n_1}\sum_{j=1}^{n_1} I(X_{1j} \le x)$$

be the ECDFs of the sample from the healthy population and that from the diseased population, respectively. Here $I(\cdot)$ is the indicator function. Further define $\hat{S}_0(x) = 1 - \hat{F}_{E0}(x)$ and $\hat{S}_1(x) = 1 - \hat{F}_{E1}(x)$. Ruopp et al. (2008) suggest to apply a binary regression with the probit link to

$$\Big\{\Big(I\big(\hat{S}_0(X_{1i}) \le j/n_0\big), \Phi^{-1}\big(j/n_0\big)\Big), i = 1, \dots, n_1, \quad j = 1, \dots, n_0 - 1\Big\}$$

with $I\big(\hat{S}_0(X_{1i}) \le j/n_0\big)$ and $\Phi^{-1}\big(j/n_0\big)$ being the response and covariate, respectively, to obtain the estimators $\hat{a}_0$ and $\hat{a}_1$ of $a_0$ and $a_1$.

With $\hat{a}_0$ and $\hat{a}_1$, Ruopp et al. (2008) suggest to estimate $ROC(x)$ by

$$\widehat{ROC}(x) = \Phi\left(\hat{a}_0 + \hat{a}_1\Phi^{-1}(x)\right)$$

and the Youden index $J$ by

$$\hat{J}_G = \max_{x\in[0,1]}\{\widehat{ROC}(x) - x\} = \widehat{ROC}(\hat{x}) - \hat{x},$$

where

$$\hat{x} = \arg\max_{x\in[0,1]}\{\widehat{ROC}(x) - x\}.$$

Having $\widehat{ROC}(x)$ and $\hat{x}$, the estimator of the optimal cutoff point $c$ can be obtained by mapping $\widehat{ROC}(\hat{x})$ back to a chosen population (Ruopp et al., 2008). We illustrate the estimation method for $c$ by mapping back to the diseased population. Suppose the $n_1$ observations from the diseased population are sorted in an descending order such that $X_{1,(1)} \ge \cdots \ge X_{1,(n_1)}$. Let $i^*$ be the index such that

$$\hat{S}_1\big(X_{1,(i^*)}\big) \le \widehat{ROC}(\hat{x}) \le \hat{S}_1\big(X_{1,(i^*+1)}\big).$$

Then ROC-GLM estimator of the optimal cutoff point is defined as (Ruopp et al., 2008)

$$\hat{c}_G = X_{1,(i^*)} + \big(X_{1,(i^*+1)} - X_{1,(i^*)}\big)\,\frac{\widehat{ROC}(\hat{x}) - \hat{S}_1\big(X_{1,(i^*)}\big)}{\hat{S}_1\big(X_{1,(i^*+1)}\big) - \hat{S}_1\big(X_{1,(i^*)}\big)}.$$

Note that the above estimation method for $c$ can be similarly performed by mapping back to the healthy population. In numerical calculation, if $n_1 \geq n_0$, $\hat{c}_G$ is obtained by mapping back to the diseased population, otherwise, it is obtained by mapping back to the healthy population.

*The ECDF-based method*

Recall that $\hat{F}_{E0}(x)$ and $\hat{F}_{E1}(x)$ are the ECDFs of the sample from the healthy population and that from the diseased population, respectively. The ECDF-based estimator of the Youden index $J$ is defined as

$$\hat{J}_E = \max_{x\in\{X_{ij}:\ i=0,1,\ j=1,\cdots,n_k\}}\{\hat{F}_{E0}(x) - \hat{F}_{E1}(x)\}.$$

The corresponding estimator $\hat{c}_E$ of the optimal cutoff point $c$ is obtained at $x$ where $\hat{J}_E$ is determined.

*The kernel-based method*

The kernel-based method uses the kernel method to estimate $F_0$ and $F_1$. Bantis et al. (2019) suggest to estimate $F_0(x)$ and $F_1(x)$ as

$$\hat{F}_{K0}(x) = \frac{1}{n_0}\sum_{i=1}^{n_0}\Phi\left(\frac{x - X_{0i}}{h_0}\right), \quad \hat{F}_{K1}(x) = \frac{1}{n_1}\sum_{i=1}^{n_1}\Phi\left(\frac{x - X_{1i}}{h_1}\right),$$

where $h_0$ and $h_1$ are are two bandwidths. They further recommended using

$$h_0 = 0.9\min\{s_0, q_0/1.34\}n_0^{-0.2} \quad \text{and} \quad h_1 = 0.9\min\{s_1, q_1/1.34\}n_1^{-0.2}.$$

Here $s_0$ and $q_0$ are the sample standard deviation and interquartile range of the sample from the healthy population, and $s_1$ and $q_1$ are for the sample from the diseased population. Let $X_{min}$ and $X_{max}$ be the minimum and maximum values of samples $\{X_{ij}:\ i=0,1,\ j=1,\cdots,n_k\}$. The estimators of $J$ and $c$ are then defined as

$$\hat{J}_K = \max_{x\in[X_{min},X_{max}]}\{\hat{F}_{K0}(x) - \hat{F}_{K1}(x)\}$$

and

$$\hat{c}_K = \arg\max_{x\in[X_{min},X_{max}]}\{\hat{F}_{K0}(x) - \hat{F}_{K1}(x)\}.$$

*The HCNS method*

Let $x_+ = \max(x, 0)$ and $H_0(x)$ be the cumulative hazard for the healthy population. The HCNS method models $H_0(x)$ by a cubic spline with parameters $(\theta_1, \cdots, \theta_{K-2})$:

$$H_0(x) = \theta_1(x-\tau_1)_+^3 + \cdots + \theta_{K-2}(x-\tau_{K-2})_+^3 + \theta_{K-1}(x-\tau_{K-1})_+^3 + \theta_K(x-\tau_K)_+^3, \quad (3.12)$$

where $\tau_1 < \cdots < \tau_K$ are the $K$ knots,

$$\theta_{K-1} = \frac{\sum_{i=1}^{K-2}\theta_i(\tau_i - \tau_K)}{\tau_K - \tau_{K-1}}, \quad \theta_K = \frac{\sum_{i=1}^{K-2}\theta_i(\tau_i - \tau_{K-1})}{\tau_{K-1} - \tau_K}.$$

Let $\hat{H}_{KM}(x)$ be the Kaplan-Meier estimator of $H_0(x)$. The estimators $(\hat{\theta}_1, \cdots, \hat{\theta}_{K-2})$ of $(\theta_1, \cdots, \theta_{K-2})$ minimize

$$\Psi(\theta_1, \cdots, \theta_{K-2}) = \sum_{i=1}^{n_0}\left\{H_0(X_{0i}) - \hat{H}_{KM}(X_{0i})\right\}^2$$

subject to the constraints such that $H_0(x)$ is a monotonically increasing function (Bantis et al., 2012).

Regarding the knot selection, six is the most preferable number of knots as recommended by Bantis et al. (2019). The six knots are selected from following eight values of the sample: minimum, $2.5^{th}$ percentile, $5^{th}$ percentile, $10^{th}$ percentile, $20^{th}$ percentile, $50^{th}$ percentile, $80^{th}$ percentile, and maximum. Bantis et al. (2019) suggest to explore all possible $\binom{8}{6}$ knot schemes and then choose the knot scheme that provides the smallest value of $\Psi(\hat{\theta}_1, \cdots, \hat{\theta}_{K-2})$.

With the estimators $(\hat{\theta}_1, \cdots, \hat{\theta}_{K-2})$, we plug them to (3.12) to obtain the estimator $\hat{H}_0(x)$ for $H_0(x)$. The CDF $F_0(x)$ is estimated subsequently as

$$\hat{F}_{H0}(x) = 1 - \exp\left\{-\hat{H}_0(x)\right\}.$$

The CDF $F_1$ can be estimated in a similar manner and we denote the correponding estimator as $\hat{F}_{H1}(x)$. The estimators of $J$ and $c$ are then defined as

$$\hat{J}_H = \max_{x \in [X_{min}, X_{max}]} \{\hat{F}_{H0}(x) - \hat{F}_{H1}(x)\}$$

and

$$\hat{c}_H = \arg \max_{x \in [X_{min}, X_{max}]} \{\hat{F}_{H0}(x) - \hat{F}_{H1}(x)\}.$$

**Methods for data with a fixed and finite LLOD**

When the LLOD exists, the measurements of the biomarker below the LLOD cannot be observed, which results in left-censored data in both samples. In the literature, only three methods are available for this case: the Box-Cox method, the ROC-GLM method, and the ECDF-based method.

The idea for the Box-Cox method in this case is similar to the one discussed before. The unknown parameters $(\lambda, \mu_0, \sigma_0, \mu_1, \sigma_1)$ in the Box-Cox transformation model can still be estimated by the maximum likelihood method. We need to take the LLOD, i.e., left censoring, into account when defining the likelihood function of the unknown parameters.

For the ROC-GLM method, Ruopp et al. (2008) suggest to slightly modify the estimation method for the unknown parameters $(a_0, a_1)$ in (3.11). Recall that we use $X_{11}, \ldots, X_{1m_1}$ to denote the observations in the sample from the diseased population which are above the LLOD. Ruopp et al. (2008) suggest to apply a binary regression with the probit link to

$$\left\{ \left( I\big(\hat{S}_0(X_{1i}) \le j/n_0\big), \Phi^{-1}\big(j/n_0\big) \right), i = 1, \ldots, m_1, \quad j = 1, \ldots, n_0 - 1 \right\}$$

with $I\big(\hat{S}_0(X_{1i}) \le j/n_0\big)$ and $\Phi^{-1}\big(j/n_0\big)$ being the response and covariate, respectively, to obtain the estimators $\hat{a}_0$ and $\hat{a}_1$ of $a_0$ and $a_1$. Once $\hat{a}_0$ and $\hat{a}_1$ are available, the procedure for estimating $ROC(t)$, $J$, and $c$ is the same as the case when there is no LLOD.

When applying the ECDF-based method, Ruopp et al. (2008) suggest an ad hoc procedure to first replace the unobserved values by a constant lower than the LLOD, for example, half of LLOD or LLOD divided by $\sqrt{2}$, and then apply the ECDF-based method discussed before.

### 3.5.2 Proofs

#### Regularity Conditions

The asymptotic properties of $(\hat{J}, \hat{c})$ are established under the following regularity conditions.

C1. For any $\epsilon > 0$, $J_\epsilon = \sup_{|x-c^*|\geq\epsilon} \{F_0(x) - F_1(x)\} < J^*$.

C2. The first and second derivatives of $F_0(x)$ and $F_1(x)$ are continuous in the neighbourhood of $c^*$, with $F_0'(c^*) - F_1'(c^*) = 0$ and $F_0''(c^*) - F_1''(c^*) < 0$.

C3. The total sample size $n = n_0 + n_1 \to \infty$ and $n_1/n_0 \to \rho$ for some constant.

C4. The two CDFs $F_0$ and $F_1$ satisfy the DRM (1.1) with a true parameter value $\boldsymbol{\theta}^*$ and $\int_r^\infty \exp\{\boldsymbol{\theta}^\top \mathbf{Q}(x)\} dF_0 < \infty$ in a neighbourhood of $\boldsymbol{\theta}^*$, and $\int_r^\infty \mathbf{Q}(x)\mathbf{Q}(x)^\top dF_0(x)$ is positive definite.

Condition C1 is from Hsieh and Turnbull (1996), which ensures $c^*$ is unique. Condition C2 comes from the definitions of the Youden index and its corresponding optimal cutoff point. Condition C3 indicates that both $n_0$ and $n_1$ go to infinity at the same rate. For simplicity and convenience of presentation, we write $\rho = n_1/n_0$ and assume that it is a constant. This does not affect our technical development. Conditions C3 and C4 guarantee that the asymptotic results in Cai and Chen (2018) can be applied.

#### Proof of Theorem 3.1

We first present some preliminary results, which serve as preparations for the proof of Theorem 3.1. We need some further notation. Let

$$H(x) = F_0(x) - F_1(x), \quad \hat{H}(x) = \hat{F}_0(x) - \hat{F}_1(x).$$

Then $J^* = H(c^*)$ and $\hat{J} = \hat{H}(\hat{c})$. Further let

$$\Delta_{n0} = \sup_{x\geq r} |\hat{F}_0(x) - F_0(x)|, \quad \Delta_{n1} = \sup_{x\geq r} |\hat{F}_1(x) - F_1(x)|, \quad \Delta_n = \sup_{x\geq r} |\hat{H}(x) - H(x)|.$$

Following the proof of Lemma 3 in Cai and Chen (2018), we have $\Delta_{n0} = O_p(n^{-1/2})$ and $\Delta_{n1} = O_p(n^{-1/2})$. Hence $\Delta_n = O_p(n^{-1/2})$.

We can establish the consistency of $\hat{c}$ and argue that, with the probability goes to 1, the estimator $\hat{c}$ is the solution to $\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(x) = 0$.

**Lemma 3.1.** *Assume Conditions C1–C4 are satisfied. Then, as $n \to \infty$, we have*

$$\hat{c} \to c^* \quad \textit{in probability} \tag{3.13}$$

*and*

$$P\left(\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(\hat{c}) = 0\right) \to 1. \tag{3.14}$$

*Proof.* For (3.13) to hold, it is sufficient to show that for any $0 < \epsilon < c^* - r$,

$$\lim_{n\to\infty} P(\hat{c} > c^* + \epsilon) = 0, \tag{3.15}$$

$$\lim_{n\to\infty} P(\hat{c} < c^* - \epsilon) = 0. \tag{3.16}$$

We focus on proving (3.15). The other part in (3.16) can be similarly proved. Let $\sup_{|x-c^*|>\epsilon} H(x) = J_\epsilon$. We choose $\epsilon^* < \epsilon$ such that

(a) $H(x) \geq \frac{J^*+J_\epsilon}{2}$, for $x \in [c^* - \epsilon^*, c^* + \epsilon^*]$;

(b) $\boldsymbol{\theta}^{*\top}\mathbf{Q}(c^* - \epsilon^*) < 0$ and $\boldsymbol{\theta}^{*\top}\mathbf{Q}(c^* + \epsilon^*) > 0$.

By Conditions C1 and C2, the existence of such $\epsilon^*$ is obvious. We further define a subset of the sample space as $A_{n,\epsilon} = A_{n1,\epsilon} \cap A_{n2,\epsilon} \cap A_{n3,\epsilon}$, where

$$\begin{aligned}
A_{n1,\epsilon} &= \left\{\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(c^* - \epsilon^*) < \frac{1}{2}\boldsymbol{\theta}^{*\top}\mathbf{Q}(c^* - \epsilon^*)\right\}, \\
A_{n2,\epsilon} &= \left\{\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(c^* + \epsilon^*) > \frac{1}{2}\boldsymbol{\theta}^{*\top}\mathbf{Q}(c^* + \epsilon^*)\right\}, \\
A_{n3,\epsilon} &= \left\{\inf_{x\in[c^*-\epsilon^*, c^*+\epsilon^*]} \hat{H}(x) \geq \frac{J^* + 3J_\epsilon}{4}\right\}.
\end{aligned}$$

The two subsets $A_{n1,\epsilon}$ and $A_{n2,\epsilon}$ together ensure that there exists a solution $\hat{c}^*$ to $\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(x) = 0$ in $[c^* - \epsilon^*, c^* + \epsilon^*]$, and $A_{n3,\epsilon}$ implies that $\hat{H}(\hat{c}^*)$ is very close to $J^*$.

With the choice of $\epsilon^*$, the consistency of $\hat{\boldsymbol{\theta}}$ (Cai and Chen, 2018), and the fact that $\Delta_n = O_p(n^{-1/2})$, it can be shown that

$$\lim_{n\to\infty} P(A_{n1,\epsilon}) = \lim_{n\to\infty} P(A_{n2,\epsilon}) = \lim_{n\to\infty} P(A_{n3,\epsilon}) = 1. \tag{3.17}$$

The details are sketched as follows. By the choice of $\epsilon^*$,

$$\begin{aligned} P(A_{n1,\epsilon}) &= P\left(\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(c^*-\epsilon^*) - \boldsymbol{\theta}^{*\top}\mathbf{Q}(c^*-\epsilon^*) < -\frac{1}{2}\boldsymbol{\theta}^{*\top}\mathbf{Q}(c^*-\epsilon^*)\right) \\ &\geq P\left(\left|\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(c^*-\epsilon^*) - \boldsymbol{\theta}^{*\top}\mathbf{Q}(c^*-\epsilon^*)\right| < -\frac{1}{2}\boldsymbol{\theta}^{*\top}\mathbf{Q}(c^*-\epsilon^*)\right). \end{aligned}$$

Then by the consistency of $\hat{\boldsymbol{\theta}}$ (Cai and Chen, 2018), we have $\lim_{n\to\infty} P(A_{n1,\epsilon}) = 1$. Similarly, we also have $\lim_{n\to\infty} P(A_{n2,\epsilon}) = 1$. As for the third term $A_{n3,\epsilon}$, again by the choice of $\epsilon^*$, when $x \in [c^*-\epsilon^*, c^*+\epsilon^*]$, we have

$$\hat{H}(x) = \{\hat{H}(x) - H(x) + H(x)\} \geq -\Delta_n + \frac{J^* + J_\epsilon}{2}.$$

Therefore,

$$P(A_{n3,\epsilon}) \geq P\left(-\Delta_n + \frac{J^*+J_\epsilon}{2} \geq \frac{J^*+3J_\epsilon}{4}\right) = P\left(\Delta_n \leq \frac{J^*-J_\epsilon}{4}\right).$$

Since $\Delta_n = O_p(n^{-\frac{1}{2}})$, we have $\lim_{n\to\infty} P(A_{n3,\epsilon}) = 1$.

We are now ready to prove (3.15). Note that

$$\begin{aligned} P(\hat{c} > c^* + \epsilon) &\leq P\big(H(\hat{c}) \leq J_\epsilon\big) \leq P\big(\hat{H}(\hat{c}) \leq J_\epsilon + \Delta_n\big) \\ &\leq P\Big(\{\hat{H}(\hat{c}) \leq J_\epsilon + \Delta_n\} \cap A_{n,\epsilon}\Big) + P(A_{n,\epsilon}^c). \end{aligned}$$

By the definition of $A_{n,\epsilon}$, if $\{\hat{H}(\hat{c}) \leq J_\epsilon + \Delta_n\}$ and $A_{n,\epsilon}$ both occur, we have

$$J_\epsilon + \Delta_n \geq \hat{H}(\hat{c}) \geq \hat{H}(\hat{c}^*) \geq \inf_{x\in[c^*-\epsilon^*, c^*+\epsilon^*]} \hat{H}(x) \geq \frac{J^*+3J_\epsilon}{4},$$

which implies $\Delta_n \geq (J^* - J_\epsilon)/4$. Hence,

$$P(\hat{c} > c^* + \epsilon) \leq P\Big(\Delta_n \geq \frac{J^*-J_\epsilon}{4}\Big) + P(A_{n,\epsilon}^c) \to 0,$$

where the last step follows from (3.17) and $\Delta_n = O_p(n^{-\frac{1}{2}})$. This finishes the proof of (3.15) and the consistency of $\hat{c}$ stated in (3.13).

For (3.14), we note that

$$A_{n1,\epsilon} \cap A_{n2,\epsilon} \subset \left\{ \hat{\boldsymbol{\theta}}^\top \mathbf{Q}(\hat{c}) = 0 \right\},$$

which, together with (3.17), implies that

$$\lim_{n\to\infty} P\left( \hat{\boldsymbol{\theta}}^\top \mathbf{Q}(\hat{c}) = 0 \right) = 1.$$

This completes the proof of (3.14). □

*Proof of Theorem 3.1.* We first consider Part (a). By (3.14) of Lemma 3.1 and the Slutsky's theorem, we can derive the asymptotic normality of $\hat{c}$ from $\hat{\boldsymbol{\theta}}^\top \mathbf{Q}(\hat{c}) = 0$. Applying the first-order Taylor expansion on $\mathbf{q}(\hat{c})$ at the point $x = c^*$ and using the consistency result of $\hat{c}$ in (3.13) of Lemma 3.1, we have

$$0 = \hat{\alpha} + \hat{\boldsymbol{\beta}}^\top \mathbf{q}(c^*) + \hat{\boldsymbol{\beta}}^\top \dot{\mathbf{q}}(c^*)(\hat{c} - c^*) + o_p(1) \cdot (\hat{c} - c^*).$$

By Theorem 1 of Cai and Chen (2018), we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \to N(0, \boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{S}^{-1}) \tag{3.18}$$

in distribution as $n \to \infty$. This together with the fact $\boldsymbol{\theta}^{*\top}\mathbf{Q}(c^*) = 0$ implies that

$$\sqrt{n}(\hat{c} - c^*) = -\frac{\mathbf{Q}(c^*)^\top}{\boldsymbol{\beta}^{*\top}\dot{\mathbf{q}}(c^*)} \left\{ \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \right\} + o_p(1) \to N(0, \sigma_c^2)$$

in distribution as $n \to \infty$, where $\sigma_c^2$ is defined in (3.7).

We next consider Part (b). Recall that

$$\hat{J} - J^* \ = \ \{\hat{F}_0(\hat{c}) - \hat{F}_1(\hat{c})\} - \{F_0(c^*) - F_1(c^*)\}.$$

Let

$$\begin{aligned}
M_{n0} &= \hat{F}_0(c^*) - F_0(c^*), \quad M_{n1} = \hat{F}_1(c^*) - F_1(c^*), \\
e_{n0} &= \{\hat{F}_0(\hat{c}) - \hat{F}_0(c^*)\} - \{F_0(\hat{c}) - F_0(c^*)\}, \\
e_{n1} &= \{\hat{F}_1(\hat{c}) - \hat{F}_1(c^*)\} - \{F_1(\hat{c}) - F_1(c^*)\}, \\
e_{n2} &= \{F_0(\hat{c}) - F_1(\hat{c})\} - \{F_0(c^*) - F_1(c^*)\}.
\end{aligned}$$

It can be shown that

$$\hat{J} - J^* = M_{n0} - M_{n1} + e_{n0} - e_{n1} + e_{n2}. \tag{3.19}$$

One of the key technical arguments is to show that $e_{n0}$, $e_{n1}$, and $e_{n2}$ are all of order $o_p(n^{-1/2})$.

By Lemma 4 of Cai and Chen (2018), we have for any $b > 0$,

$$\begin{aligned} & \sup_{x:|x-c^*|<bn^{-1/2}} |\{\hat{F}_0(x) - \hat{F}_0(c^*)\} - \{F_0(x) - F_0(c^*)\}| \\ = \; & O_p(n^{-3/4}(\log(n))^{1/2}) = o_p(n^{-1/2}). \end{aligned} \tag{3.20}$$

The result in Part (a) implies that $\hat{c} - c^* = O_p(n^{-1/2})$, which, together with (3.20), leads to $e_{n0} = o_p(n^{-1/2})$. Similarly, we also have $e_{n1} = o_p(n^{-1/2})$. By the second order Taylor expansion and Condition A2, we have $e_{n2} = o_p(n^{-1/2})$. It follows that

$$\sqrt{n}\left(\hat{J} - J^*\right) = \sqrt{n}(M_{n0} - M_{n1}) + o_p(1). \tag{3.21}$$

Applying Theorem 2 of Cai and Chen (2018), we have

$$\sqrt{n}\begin{pmatrix} M_{n0} \\ M_{n1} \end{pmatrix} = \sqrt{n}\begin{pmatrix} \hat{F}_0(c^*) - F_0(c^*) \\ \hat{F}_1(c^*) - F_1(c^*) \end{pmatrix} \to N\left(\mathbf{0}, \begin{pmatrix} \sigma_{00}^2 & \sigma_{01}^2 \\ \sigma_{01}^2 & \sigma_{11}^2 \end{pmatrix}\right) \tag{3.22}$$

in distribution as $n \to \infty$, where

$$\begin{aligned} \sigma_{00}^2 &= (1+\rho)\{F_0(c^*) - F_0^2(c^*)\} \\ &\quad -\rho(1+\rho)\left\{A_0(c^*) - \begin{pmatrix} A_0(c^*) \\ \boldsymbol{A}_1(c^*) \end{pmatrix}^\top \boldsymbol{A}^{-1} \begin{pmatrix} A_0(c^*) \\ \boldsymbol{A}_1(c^*) \end{pmatrix}\right\}, \\ \sigma_{01}^2 &= (1+\rho)\left\{A_0(c^*) - \begin{pmatrix} A_0(c^*) \\ \boldsymbol{A}_1(c^*) \end{pmatrix}^\top \boldsymbol{A}^{-1} \begin{pmatrix} A_0(c^*) \\ \boldsymbol{A}_1(c^*) \end{pmatrix}\right\}, \\ \sigma_{11}^2 &= \frac{1+\rho}{\rho}\{F_1(c^*) - F_1^2(c)\} \\ &\quad -\frac{1+\rho}{\rho}\left\{A_0(c^*) - \begin{pmatrix} A_0(c^*) \\ \boldsymbol{A}_1(c^*) \end{pmatrix}^\top \boldsymbol{A}^{-1} \begin{pmatrix} A_0(c^*) \\ \boldsymbol{A}_1(c^*) \end{pmatrix}\right\}. \end{aligned}$$

It immediately follows that, as $n \to \infty$,

$$\sqrt{n}\,(M_{n0} - M_{n1}) \to N(0, \sigma_J^2)$$

in distribution, where $\sigma_J^2$ is defined in (3.8). Recall that $\sqrt{n}(\hat{J} - J^*) = \sqrt{n}\,(M_{n0} - M_{n1}) + o_p(1)$. By the Slusky's theorem, we have

$$\sqrt{n}(\hat{J} - J^*) \to N(0, \sigma_J^2)$$

in distribution as $n \to \infty$. This completes the proof of the theorem. □

### 3.5.3 Additional simulation studies

#### Additional simulations for the gamma distributional setting

Tables 3.12 and 3.13 compare the RBs and MSEs of point estimators of $(J, c)$ under gamma setting when the LLOD equals 30% quantile of $F_0$. Tables 3.14–3.15 present the CPs and ALs of the CIs of $(J, c)$ under the same setting. The general trend for comparing the proposed method and all candidate methods is similar to the case when the LLOD is equal to the 15% quantile of $F_0$. Hence, we omit the comparison results here.

Table 3.12: RB (%) and MSE ($\times 100$) for point estimators of $J$ when the LLOD equals 30% quantile of $F_0$ (gamma distributions).

| $J$ | $(n_0, n_1)$ | (50, 50) RB | (50, 50) MSE | (100, 100) RB | (100, 100) MSE | (200, 200) RB | (200, 200) MSE | (50, 150) RB | (50, 150) MSE | (150, 50) RB | (150, 50) MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\hat{J}$ | 8.62 | 0.64 | 3.78 | 0.31 | 1.71 | 0.14 | 6.89 | 0.41 | 4.69 | 0.41 |
| | $\hat{J}_B$ | 10.52 | 0.65 | 5.12 | 0.32 | 2.94 | 0.16 | 8.24 | 0.42 | 6.35 | 0.43 |
| | $\hat{J}_G$ | 5.85 | 0.80 | 0.77 | 0.40 | -2.84 | 0.20 | 4.31 | 0.53 | 0.83 | 0.57 |
| | $\hat{J}_E$ | 39.99 | 1.29 | 26.6 | 0.64 | 17.72 | 0.30 | 32.97 | 0.86 | 32.52 | 0.89 |
| 0.4 | $\hat{J}$ | 3.16 | 0.58 | 1.36 | 0.30 | 0.62 | 0.13 | 2.25 | 0.37 | 1.55 | 0.40 |
| | $\hat{J}_B$ | 5.06 | 0.59 | 2.57 | 0.29 | 1.67 | 0.14 | 3.74 | 0.37 | 3.01 | 0.41 |
| | $\hat{J}_G$ | -0.34 | 0.73 | -2.06 | 0.38 | -3.18 | 0.19 | -1.13 | 0.47 | -2.07 | 0.53 |
| | $\hat{J}_E$ | 16.42 | 1.06 | 10.62 | 0.53 | 6.78 | 0.24 | 13.36 | 0.68 | 12.57 | 0.74 |
| 0.6 | $\hat{J}$ | 1.81 | 0.46 | 0.66 | 0.24 | 0.33 | 0.11 | 1.11 | 0.28 | 0.89 | 0.35 |
| | $\hat{J}_B$ | 3.14 | 0.44 | 1.64 | 0.23 | 1.11 | 0.11 | 2.16 | 0.27 | 2.03 | 0.33 |
| | $\hat{J}_G$ | -1.12 | 0.56 | -1.91 | 0.30 | -2.42 | 0.15 | -1.72 | 0.37 | -1.90 | 0.42 |
| | $\hat{J}_E$ | 8.89 | 0.76 | 5.38 | 0.37 | 3.70 | 0.19 | 7.14 | 0.48 | 6.73 | 0.56 |
| 0.8 | $\hat{J}$ | 1.15 | 0.27 | 0.42 | 0.14 | 0.17 | 0.07 | 0.62 | 0.14 | 0.64 | 0.21 |
| | $\hat{J}_B$ | 1.55 | 0.22 | 0.78 | 0.12 | 0.53 | 0.05 | 0.98 | 0.13 | 1.00 | 0.17 |
| | $\hat{J}_G$ | -1.17 | 0.36 | -1.27 | 0.18 | -1.50 | 0.10 | -1.46 | 0.22 | -1.21 | 0.27 |
| | $\hat{J}_E$ | 4.51 | 0.39 | 2.86 | 0.20 | 1.94 | 0.10 | 3.73 | 0.24 | 3.52 | 0.30 |

Table 3.13: RB (%) and MSE ($\times 100$) for point estimators of $c$ when the LLOD equals 30% quantile of $F_0$ (gamma distributions).

| $J$ | $(n_0, n_1)$ | (50, 50) RB | (50, 50) MSE | (100, 100) RB | (100, 100) MSE | (200, 200) RB | (200, 200) MSE | (50, 150) RB | (50, 150) MSE | (150, 50) RB | (150, 50) MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\hat{c}$ | 1.72 | 166.59 | -0.14 | 76.36 | -1.11 | 25.61 | -0.17 | 100.80 | 0.16 | 99.95 |
| | $\hat{c}_B$ | 1.22 | 120.35 | 1.62 | 64.03 | 0.53 | 30.39 | 1.06 | 84.80 | 1.17 | 83.72 |
| | $\hat{c}_G$ | 18.38 | 241.37 | 20.51 | 170.02 | 19.84 | 126.64 | 19.37 | 187.87 | 22.02 | 267.68 |
| | $\hat{c}_E$ | 0.21 | 246.68 | 0.65 | 169.08 | -0.76 | 116.80 | -0.59 | 204.52 | 2.51 | 215.93 |
| 0.4 | $\hat{c}$ | -1.39 | 40.03 | -0.42 | 18.20 | -0.11 | 7.98 | -1.23 | 25.73 | -0.24 | 25.09 |
| | $\hat{c}_B$ | 0.14 | 56.14 | 0.33 | 29.60 | 0.10 | 13.84 | -0.47 | 38.37 | 0.81 | 35.86 |
| | $\hat{c}_G$ | 11.39 | 121.93 | 14.37 | 109.64 | 14.58 | 91.16 | 13.27 | 108.65 | 14.72 | 115.49 |
| | $\hat{c}_E$ | -2.89 | 160.26 | -0.23 | 126.83 | -0.45 | 75.26 | -2.65 | 149.91 | 1.53 | 156.36 |
| 0.6 | $\hat{c}$ | -0.59 | 35.72 | -0.27 | 17.64 | -0.04 | 8.36 | -0.81 | 24.8 | -0.01 | 22.01 |
| | $\hat{c}_B$ | -0.32 | 51.67 | -0.27 | 26.14 | -0.20 | 12.4 | -1.09 | 33.34 | 0.40 | 30.83 |
| | $\hat{c}_G$ | 6.94 | 123.88 | 9.08 | 90.88 | 10.55 | 78.96 | 8.81 | 93.44 | 9.86 | 98.53 |
| | $\hat{c}_E$ | -2.29 | 161.37 | -0.67 | 118.08 | 0.20 | 74.7 | -2.40 | 147.73 | 1.22 | 146.64 |
| 0.8 | $\hat{c}$ | -0.52 | 72.34 | -0.34 | 36.15 | 0.01 | 17.71 | -0.95 | 53.74 | 0.17 | 41.95 |
| | $\hat{c}_B$ | -0.33 | 69.75 | -0.29 | 34.76 | -0.04 | 16.50 | -1.12 | 49.95 | 0.50 | 39.56 |
| | $\hat{c}_G$ | 1.13 | 200.79 | 4.49 | 108.25 | 6.35 | 85.81 | 5.11 | 124.86 | 4.12 | 117.41 |
| | $\hat{c}_E$ | -3.18 | 236.72 | -1.72 | 159.92 | -1.25 | 107.68 | -4.03 | 195.3 | 1.14 | 197.76 |

Table 3.14: CP (%) and AL for CIs of $J$ when the LLOD equals 30% quantile of $F_0$ (gamma distributions).

| $J$ | $(n_0, n_1)$ | $(50, 50)$ CP | AL | $(100, 100)$ CP | AL | $(200, 200)$ CP | AL | $(50, 150)$ CP | AL | $(150, 50)$ CP | AL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\mathcal{I}_J$ | 92.1 | 0.31 | 93.6 | 0.21 | 96.0 | 0.15 | 92.8 | 0.24 | 93.7 | 0.25 |
| | $\mathcal{I}_{JB}$ | 93.8 | 0.30 | 94.1 | 0.21 | 95.3 | 0.15 | 92.8 | 0.24 | 94.6 | 0.25 |
| | $\mathcal{I}_{JG}$ | 93.2 | 0.33 | 95.2 | 0.24 | 95.3 | 0.17 | 93.6 | 0.27 | 94.0 | 0.27 |
| | $\mathcal{I}_{JE}$ | 55.5 | 0.25 | 59.9 | 0.19 | 67.3 | 0.14 | 52.9 | 0.21 | 55.6 | 0.21 |
| 0.4 | $\mathcal{I}_J$ | 95.0 | 0.28 | 95.1 | 0.20 | 95.8 | 0.15 | 93.8 | 0.23 | 94.28 | 0.24 |
| | $\mathcal{I}_{JB}$ | 95.6 | 0.29 | 93.7 | 0.20 | 95.4 | 0.14 | 92.9 | 0.22 | 94.08 | 0.24 |
| | $\mathcal{I}_{JG}$ | 93.8 | 0.33 | 94.3 | 0.23 | 94.9 | 0.16 | 94.3 | 0.26 | 93.18 | 0.27 |
| | $\mathcal{I}_{JE}$ | 72.1 | 0.26 | 75.0 | 0.19 | 78.8 | 0.14 | 69.2 | 0.21 | 73.22 | 0.23 |
| 0.6 | $\mathcal{I}_J$ | 95.6 | 0.26 | 94.5 | 0.18 | 95.28 | 0.13 | 94.3 | 0.20 | 93.79 | 0.22 |
| | $\mathcal{I}_{JB}$ | 95.8 | 0.25 | 94.1 | 0.18 | 95.08 | 0.13 | 94.2 | 0.19 | 94.29 | 0.22 |
| | $\mathcal{I}_{JG}$ | 94.8 | 0.29 | 94.8 | 0.20 | 94.57 | 0.14 | 94.1 | 0.22 | 93.29 | 0.25 |
| | $\mathcal{I}_{JE}$ | 76.0 | 0.24 | 78.1 | 0.18 | 81.61 | 0.13 | 72.8 | 0.19 | 77.35 | 0.21 |
| 0.8 | $\mathcal{I}_J$ | 95.5 | 0.20 | 94.6 | 0.14 | 95.18 | 0.10 | 95.3 | 0.14 | 95.48 | 0.18 |
| | $\mathcal{I}_{JB}$ | 95.8 | 0.18 | 95.0 | 0.13 | 95.08 | 0.09 | 94.9 | 0.14 | 95.58 | 0.17 |
| | $\mathcal{I}_{JG}$ | 94.4 | 0.23 | 93.8 | 0.16 | 94.18 | 0.11 | 94.1 | 0.16 | 94.18 | 0.20 |
| | $\mathcal{I}_{JE}$ | 80.2 | 0.18 | 84.7 | 0.14 | 84.94 | 0.10 | 73.7 | 0.13 | 81.22 | 0.16 |

Table 3.15: CP (%) and AL for CIs of $c$ when the LLOD equals 30% quantile of $F_0$ (gamma distributions).

| $J$ | $(n_0, n_1)$ | $(50, 50)$ CP | AL | $(100, 100)$ CP | AL | $(200, 200)$ CP | AL | $(50, 150)$ CP | AL | $(150, 50)$ CP | AL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\mathcal{I}_c$ | 96.6 | 4.37 | 95.3 | 3.27 | 96.4 | 2.02 | 97.7 | 4.67 | 94.7 | 4.42 |
| | $\mathcal{I}_{cB}$ | 81.5 | 4.53 | 81.3 | 2.14 | 83.7 | 1.53 | 70.4 | 2.19 | 85.1 | 2.69 |
| | $\mathcal{I}_{cG}$ | 89.8 | 5.34 | 75.7 | 3.59 | 63.6 | 2.42 | 80.9 | 4.42 | 80.6 | 4.17 |
| | $\mathcal{I}_{cE}$ | 95.9 | 4.72 | 94.5 | 4.13 | 95.5 | 3.55 | 93.8 | 4.25 | 95.0 | 4.52 |
| 0.4 | $\mathcal{I}_c$ | 95.4 | 2.37 | 94.7 | 1.61 | 95.7 | 1.12 | 94.8 | 1.86 | 95.6 | 1.91 |
| | $\mathcal{I}_{cB}$ | 84.4 | 2.19 | 84.0 | 1.57 | 86.8 | 1.13 | 75.4 | 1.47 | 89.1 | 1.94 |
| | $\mathcal{I}_{cG}$ | 89.5 | 3.39 | 75.2 | 2.45 | 51.2 | 1.75 | 78.1 | 2.68 | 73.8 | 2.49 |
| | $\mathcal{I}_{cE}$ | 94.1 | 4.27 | 95.4 | 3.66 | 95.1 | 3.09 | 93.0 | 3.78 | 96.4 | 4.18 |
| 0.6 | $\mathcal{I}_c$ | 92.4 | 2.28 | 94.3 | 1.62 | 95.1 | 1.14 | 92.9 | 1.90 | 94.8 | 1.83 |
| | $\mathcal{I}_{cB}$ | 89.5 | 2.34 | 89.2 | 1.67 | 91.1 | 1.19 | 83.9 | 1.65 | 93.3 | 1.98 |
| | $\mathcal{I}_{cG}$ | 93.7 | 3.67 | 86.4 | 2.68 | 69.5 | 1.92 | 87.1 | 2.80 | 83.4 | 2.59 |
| | $\mathcal{I}_{cE}$ | 94.4 | 4.31 | 95.1 | 3.68 | 96.4 | 3.03 | 91.8 | 3.82 | 94.6 | 4.14 |
| 0.8 | $\mathcal{I}_c$ | 93.0 | 3.25 | 94.0 | 2.31 | 95.0 | 1.65 | 93.0 | 2.78 | 94.28 | 2.51 |
| | $\mathcal{I}_{cB}$ | 92.4 | 3.01 | 93.0 | 2.14 | 93.47 | 1.52 | 89.9 | 2.43 | 93.88 | 2.35 |
| | $\mathcal{I}_{cG}$ | 94.9 | 5.43 | 94.9 | 3.86 | 89.46 | 2.75 | 94.2 | 3.90 | 94.58 | 3.44 |
| | $\mathcal{I}_{cE}$ | 88.5 | 4.88 | 93.7 | 4.24 | 94.28 | 3.53 | 84.7 | 4.34 | 94.48 | 4.73 |

**Additional simulations for the lognormal distributional setting**

We present the simulation results under the lognormal distributional setting. Tables 3.16–3.19 provide the simulation results of the point estimators and CIs of $(J, c)$ when there is no LLOD. Tables 3.20–3.23 summarize the simulation results of the point estimators and CIs of $(J, c)$ when the LLOD equals 15% quantile of $F_0$. Tables 3.24–3.27 summarize the simulation results of the point estimators and CIs of $(J, c)$ when the LLOD equals 30% quantile of $F_0$. We only summarize the comparison results between the proposed method and the Box-Cox method. The general trend for comparing our method and other candidate methods is similar to the gamma distributional setting. Hence we omit their comparison.

First, we discuss the point estimators of $(J, c)$. For estimating the Youden index, the RBs and MSEs of the estimators $\hat{J}$ and $\hat{J}_B$ are very close and small in majority cases. For estimating the optimal cutoff point, the estimator $\hat{c}_B$ is uniformly better than our estimator in terms of MSE. This is expected because the parametric assumption for the Box-Cox method is satisfied.

Next, we discuss the findings for the CIs of $(J, c)$. In general, the ALs of both $\mathcal{I}_J$ and $\mathcal{I}_{JB}$ are comparable and small, while both CIs encounter slight overcoverage in some cases especially in the cases that one of the sample sizes is small. The performance of the CI $\mathcal{I}_c$ is stable with short ALs and reasonable CPs when there is no LLOD or when the LLOD equals 15% quantile of $F_0$. When the LLOD increases to 30% quantile of $F_0$, the CI $\mathcal{I}_c$ tends to have undercoverage and longer AL especially in the cases when one of small sample sizes is small or when the Youden index is small. When there is no LLOD, the CI $\mathcal{I}_{cB}$ has similar performance as $\mathcal{I}_c$. However, with the existence of a fixed and finite LLOD, the CI $\mathcal{I}_{cB}$ experiences severe undercoverage when $J = 0.2$ and 0.4. Consequently, the CPs of $\mathcal{I}_{cB}$ are much worse than those of $\mathcal{I}_c$ in those cases.

Table 3.16: RB (%) and MSE (×100) for point estimators of $J$ when there is no LLOD (lognormal distributions).

| $J$ | $(n_0, n_1)$ | $(50, 50)$ RB | $(50, 50)$ MSE | $(100, 100)$ RB | $(100, 100)$ MSE | $(200, 200)$ RB | $(200, 200)$ MSE | $(50, 150)$ RB | $(50, 150)$ MSE | $(150, 50)$ RB | $(150, 50)$ MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\hat{J}$ | 5.37 | 0.48 | 3.22 | 0.25 | 0.83 | 0.14 | 4.15 | 0.26 | 3.34 | 0.41 |
| | $\hat{J}_B$ | 2.80 | 0.46 | 1.61 | 0.23 | 0.19 | 0.13 | 3.03 | 0.25 | 0.31 | 0.38 |
| | $\hat{J}_G$ | 2.10 | 0.48 | 1.42 | 0.25 | -0.27 | 0.14 | 1.84 | 0.26 | 2.14 | 0.42 |
| | $\hat{J}_E$ | 29.91 | 0.88 | 20.20 | 0.46 | 12.17 | 0.22 | 24.49 | 0.54 | 24.86 | 0.69 |
| | $\hat{J}_K$ | 8.50 | 0.49 | 6.00 | 0.28 | 3.12 | 0.16 | 5.59 | 0.29 | 8.20 | 0.43 |
| | $\hat{J}_H$ | 9.48 | 0.60 | 5.96 | 0.29 | 2.89 | 0.16 | 7.85 | 0.34 | 6.54 | 0.48 |
| 0.4 | $\hat{J}$ | 2.87 | 0.55 | 1.55 | 0.26 | 0.27 | 0.15 | 1.99 | 0.29 | 1.36 | 0.44 |
| | $\hat{J}_B$ | 2.60 | 0.52 | 1.38 | 0.25 | 0.23 | 0.14 | 1.90 | 0.28 | 0.91 | 0.40 |
| | $\hat{J}_G$ | 1.06 | 0.53 | 0.54 | 0.25 | -0.27 | 0.15 | 0.55 | 0.28 | 0.64 | 0.44 |
| | $\hat{J}_E$ | 14.10 | 0.91 | 9.31 | 0.43 | 5.54 | 0.23 | 11.46 | 0.54 | 10.99 | 0.67 |
| | $\hat{J}_K$ | 2.08 | 0.49 | 1.18 | 0.25 | 0.17 | 0.16 | 1.03 | 0.29 | 1.72 | 0.40 |
| | $\hat{J}_H$ | 2.12 | 0.64 | 2.01 | 0.31 | 0.97 | 0.18 | 3.04 | 0.35 | 1.29 | 0.50 |
| 0.6 | $\hat{J}$ | 1.90 | 0.45 | 0.92 | 0.21 | 0.20 | 0.12 | 1.15 | 0.24 | 0.76 | 0.35 |
| | $\hat{J}_B$ | 2.18 | 0.41 | 1.10 | 0.19 | 0.26 | 0.11 | 1.35 | 0.22 | 0.95 | 0.31 |
| | $\hat{J}_G$ | 0.32 | 0.43 | 0.13 | 0.21 | -0.26 | 0.12 | -0.11 | 0.23 | 0.23 | 0.36 |
| | $\hat{J}_E$ | 7.97 | 0.69 | 5.16 | 0.34 | 3.16 | 0.17 | 6.48 | 0.41 | 5.91 | 0.51 |
| | $\hat{J}_K$ | -1.15 | 0.37 | -1.38 | 0.19 | -1.43 | 0.12 | -1.21 | 0.23 | -1.60 | 0.29 |
| | $\hat{J}_H$ | 0.65 | 0.52 | 0.26 | 0.25 | -0.38 | 0.14 | 0.84 | 0.28 | 0.05 | 0.43 |
| 0.8 | $\hat{J}$ | 1.42 | 0.25 | 0.64 | 0.12 | 0.18 | 0.07 | 0.72 | 0.13 | 0.63 | 0.19 |
| | $\hat{J}_B$ | 1.36 | 0.21 | 0.66 | 0.10 | 0.19 | 0.05 | 0.78 | 0.11 | 0.61 | 0.16 |
| | $\hat{J}_G$ | -0.12 | 0.26 | -0.24 | 0.13 | -0.31 | 0.07 | -0.59 | 0.14 | 0.03 | 0.20 |
| | $\hat{J}_E$ | 4.61 | 0.39 | 2.91 | 0.19 | 1.86 | 0.10 | 3.70 | 0.23 | 3.20 | 0.28 |
| | $\hat{J}_K$ | -3.27 | 0.28 | -3.02 | 0.16 | -2.61 | 0.10 | -2.56 | 0.17 | -3.77 | 0.25 |
| | $\hat{J}_H$ | 1.38 | 0.38 | 1.00 | 0.19 | 0.35 | 0.10 | 1.04 | 0.19 | 0.92 | 0.31 |

Table 3.17: RB (%) and MSE (×100) for point estimators of $c$ when there is no LLOD (lognormal distributions).

| $J$ | $(n_0, n_1)$ | (50, 50) RB | (50, 50) MSE | (100, 100) RB | (100, 100) MSE | (200, 200) RB | (200, 200) MSE | (50, 150) RB | (50, 150) MSE | (150, 50) RB | (150, 50) MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\hat{c}$ | 0.20 | 165.43 | 0.10 | 74.31 | 0.22 | 38.72 | -0.08 | 106.52 | 0.58 | 92.00 |
| | $\hat{c}_B$ | -0.57 | 150.88 | -0.31 | 64.87 | 0.00 | 32.17 | -0.39 | 99.33 | -0.10 | 72.64 |
| | $\hat{c}_G$ | -1.93 | 249.45 | -0.76 | 100.65 | -0.30 | 51.49 | -1.29 | 130.09 | -0.21 | 99.08 |
| | $\hat{c}_E$ | -2.90 | 547.53 | -1.50 | 343.76 | -0.34 | 228.25 | -3.14 | 448.08 | -0.34 | 448.97 |
| | $\hat{c}_K$ | 2.86 | 480.18 | 1.95 | 253.76 | 1.87 | 155.61 | 2.65 | 404.36 | 2.33 | 285.09 |
| | $\hat{c}_H$ | 1.98 | 300.68 | 0.09 | 225.39 | 0.21 | 110.75 | -0.32 | 298.09 | 0.99 | 258.55 |
| 0.4 | $\hat{c}$ | -0.26 | 86.21 | -0.05 | 41.36 | -0.02 | 20.08 | -0.16 | 56.25 | 0.14 | 53.16 |
| | $\hat{c}_B$ | -0.85 | 77.38 | -0.38 | 38.58 | -0.18 | 18.08 | -0.43 | 53.77 | -0.42 | 42.94 |
| | $\hat{c}_G$ | -2.17 | 159.00 | -1.10 | 71.71 | -0.54 | 36.52 | -1.19 | 80.69 | -0.51 | 72.77 |
| | $\hat{c}_E$ | -1.58 | 347.41 | -0.57 | 242.58 | -0.44 | 151.00 | -1.55 | 285.24 | 0.25 | 299.00 |
| | $\hat{c}_K$ | 3.33 | 211.66 | 2.62 | 129.96 | 2.10 | 75.34 | 2.67 | 187.62 | 3.06 | 130.84 |
| | $\hat{c}_H$ | 0.79 | 279.02 | 0.60 | 141.91 | 0.47 | 66.56 | 0.38 | 181.29 | 1.00 | 178.57 |
| 0.6 | $\hat{c}$ | -0.66 | 76.57 | -0.29 | 36.89 | -0.20 | 17.93 | -0.43 | 48.70 | -0.24 | 47.57 |
| | $\hat{c}_B$ | -0.56 | 67.27 | -0.25 | 32.86 | -0.19 | 16.09 | -0.38 | 45.19 | -0.22 | 36.87 |
| | $\hat{c}_G$ | -2.46 | 174.05 | -1.07 | 82.60 | -0.60 | 37.54 | -1.15 | 80.66 | -0.82 | 77.85 |
| | $\hat{c}_E$ | -1.20 | 282.61 | -0.63 | 183.82 | -0.09 | 115.00 | -1.50 | 237.71 | 0.39 | 230.26 |
| | $\hat{c}_K$ | 4.32 | 200.28 | 3.44 | 112.79 | 2.63 | 66.22 | 3.32 | 154.10 | 3.89 | 116.61 |
| | $\hat{c}_H$ | 0.57 | 244.48 | 1.43 | 132.73 | 0.97 | 57.63 | 1.00 | 155.92 | 1.20 | 181.91 |
| 0.8 | $\hat{c}$ | -0.88 | 99.18 | -0.56 | 48.62 | -0.41 | 24.13 | -0.84 | 65.16 | -0.48 | 60.61 |
| | $\hat{c}_B$ | -0.22 | 70.70 | -0.13 | 34.19 | -0.16 | 16.96 | -0.36 | 51.67 | 0.02 | 37.30 |
| | $\hat{c}_G$ | -3.48 | 287.68 | -1.58 | 137.17 | -0.77 | 63.57 | -1.01 | 116.65 | -1.94 | 134.35 |
| | $\hat{c}_E$ | -1.37 | 291.59 | -0.55 | 189.65 | -0.74 | 128.34 | -1.46 | 232.99 | 0.43 | 233.55 |
| | $\hat{c}_K$ | 3.66 | 199.67 | 3.16 | 114.55 | 2.37 | 65.77 | 3.04 | 163.09 | 3.33 | 114.45 |
| | $\hat{c}_H$ | -1.78 | 247.56 | -1.45 | 135.38 | -1.38 | 78.60 | -1.95 | 156.22 | -1.49 | 194.25 |

Table 3.18: CP (%) and AL for CIs of $J$ when there is no LLOD (lognormal distributions).

| $J$ | $(n_0, n_1)$ | (50, 50) CP | (50, 50) AL | (100, 100) CP | (100, 100) AL | (200, 200) CP | (200, 200) AL | (50, 150) CP | (50, 150) AL | (150, 50) CP | (150, 50) AL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\mathcal{I}_J$ | 96.0 | 0.28 | 95.2 | 0.20 | 94.2 | 0.14 | 94.8 | 0.20 | 94.4 | 0.25 |
| | $\mathcal{I}_{JB}$ | 95.2 | 0.27 | 94.0 | 0.19 | 93.8 | 0.13 | 95.8 | 0.20 | 93.8 | 0.23 |
| | $\mathcal{I}_{JG}$ | 95.4 | 0.27 | 93.8 | 0.20 | 92.8 | 0.14 | 95.0 | 0.20 | 94.3 | 0.25 |
| | $\mathcal{I}_{JE}$ | 80.6 | 0.29 | 82.7 | 0.21 | 82.8 | 0.15 | 73.5 | 0.21 | 83.7 | 0.26 |
| | $\mathcal{I}_{JK}$ | 94.9 | 0.27 | 92.7 | 0.20 | 92.5 | 0.14 | 93.3 | 0.20 | 93.4 | 0.24 |
| | $\mathcal{I}_{JH}$ | 93.4 | 0.28 | 93 | 0.20 | 94.8 | 0.15 | 90 | 0.21 | 94.8 | 0.25 |
| 0.4 | $\mathcal{I}_J$ | 96.2 | 0.28 | 96.0 | 0.20 | 94.0 | 0.15 | 95.2 | 0.21 | 95.2 | 0.25 |
| | $\mathcal{I}_{JB}$ | 95.4 | 0.29 | 95.6 | 0.20 | 93.5 | 0.14 | 95.7 | 0.21 | 95.2 | 0.25 |
| | $\mathcal{I}_{JG}$ | 95.9 | 0.29 | 94.6 | 0.20 | 93.9 | 0.14 | 95.3 | 0.21 | 93.9 | 0.26 |
| | $\mathcal{I}_{JE}$ | 82.3 | 0.29 | 86.0 | 0.22 | 85.4 | 0.16 | 77.7 | 0.22 | 85.7 | 0.27 |
| | $\mathcal{I}_{JK}$ | 94.8 | 0.28 | 94.7 | 0.20 | 92.7 | 0.14 | 94.3 | 0.21 | 93.7 | 0.24 |
| | $\mathcal{I}_{JH}$ | 94.0 | 0.30 | 94.9 | 0.22 | 95.6 | 0.16 | 95.2 | 5.66 | 94.4 | 0.27 |
| 0.6 | $\mathcal{I}_J$ | 95.7 | 0.26 | 96.2 | 0.18 | 94.8 | 0.13 | 95.8 | 0.19 | 95.3 | 0.23 |
| | $\mathcal{I}_{JB}$ | 95.6 | 0.25 | 96.0 | 0.18 | 94.7 | 0.13 | 95.7 | 0.19 | 95.4 | 0.22 |
| | $\mathcal{I}_{JG}$ | 94.3 | 0.26 | 94.9 | 0.18 | 93.7 | 0.13 | 94.4 | 0.19 | 93.9 | 0.23 |
| | $\mathcal{I}_{JE}$ | 84.2 | 0.26 | 83.7 | 0.19 | 85.7 | 0.14 | 77.7 | 0.19 | 84.5 | 0.24 |
| | $\mathcal{I}_{JK}$ | 94.3 | 0.24 | 95.8 | 0.17 | 93.1 | 0.13 | 95.1 | 0.19 | 94.4 | 0.21 |
| | $\mathcal{I}_{JH}$ | 93.9 | 0.29 | 94.2 | 0.20 | 95.0 | 0.14 | 92.1 | 0.20 | 94.7 | 0.27 |
| 0.8 | $\mathcal{I}_J$ | 96.9 | 0.20 | 95.7 | 0.14 | 94.9 | 0.10 | 96.0 | 0.15 | 94.9 | 0.17 |
| | $\mathcal{I}_{JB}$ | 96.0 | 0.18 | 96.2 | 0.13 | 94.9 | 0.09 | 95.2 | 0.14 | 95.6 | 0.16 |
| | $\mathcal{I}_{JG}$ | 92.8 | 0.20 | 94.5 | 0.14 | 94.6 | 0.10 | 94.0 | 0.14 | 92.5 | 0.18 |
| | $\mathcal{I}_{JE}$ | 82.3 | 0.18 | 85.6 | 0.14 | 86 | 0.10 | 77.7 | 0.14 | 84.5 | 0.17 |
| | $\mathcal{I}_{JK}$ | 94.2 | 0.18 | 92.5 | 0.13 | 88.9 | 0.09 | 94.1 | 0.14 | 91.6 | 0.16 |
| | $\mathcal{I}_{JH}$ | 89.5 | 0.23 | 95.7 | 0.17 | 94.7 | 0.12 | 92.4 | 0.17 | 91.3 | 0.22 |

Table 3.19: CP (%) and AL for CIs of $c$ when there is no LLOD (lognormal distributions).

| | $(n_0, n_1)$ | $(50, 50)$ | | $(100, 100)$ | | $(200, 200)$ | | $(50, 150)$ | | $(150, 50)$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $J$ | | CP | AL | CP | AL | CP | AL | CP | AL | CP | AL |
| | $\mathcal{I}_c$ | 93.9 | 5.06 | 95.3 | 3.44 | 95.0 | 2.44 | 94.3 | 4.14 | 95.3 | 3.82 |
| | $\mathcal{I}_{cB}$ | 93.7 | 6.57 | 94.3 | 3.33 | 94.7 | 2.26 | 93.4 | 4.12 | 93.5 | 4.56 |
| 0.2 | $\mathcal{I}_{cG}$ | 93.4 | 6.95 | 94.6 | 4.22 | 95.1 | 2.89 | 93.1 | 4.67 | 95.8 | 4.42 |
| | $\mathcal{I}_{cE}$ | 93.4 | 8.05 | 95.7 | 6.70 | 96.5 | 5.56 | 91.6 | 7.08 | 96.6 | 7.49 |
| | $\mathcal{I}_{cK}$ | 95.5 | 8.95 | 95.0 | 6.48 | 94.8 | 4.67 | 95.6 | 7.53 | 95.2 | 7.21 |
| | $\mathcal{I}_{cH}$ | 94.9 | 8.44 | 96.2 | 5.60 | 94.6 | 4.26 | 93.7 | 6.23 | 95.0 | 7.22 |
| | $\mathcal{I}_c$ | 94.1 | 3.57 | 94.5 | 2.51 | 95.4 | 1.77 | 93.7 | 2.92 | 94.6 | 2.84 |
| | $\mathcal{I}_{cB}$ | 93.8 | 3.36 | 92.9 | 2.37 | 94.6 | 1.68 | 93.7 | 2.84 | 92.8 | 2.52 |
| 0.4 | $\mathcal{I}_{cG}$ | 92.4 | 4.70 | 93.1 | 3.30 | 94.8 | 2.33 | 94.1 | 3.43 | 94.6 | 3.31 |
| | $\mathcal{I}_{cE}$ | 94.7 | 6.48 | 96.8 | 5.39 | 97.0 | 4.45 | 94.0 | 5.78 | 96.7 | 6.05 |
| | $\mathcal{I}_{cK}$ | 94.8 | 5.74 | 93.5 | 4.00 | 93.2 | 3.07 | 94.6 | 4.89 | 92.1 | 4.16 |
| | $\mathcal{I}_{cH}$ | 96.1 | 6.38 | 96.0 | 4.71 | 96.7 | 3.46 | 95.4 | 5.09 | 95.2 | 5.66 |
| | $\mathcal{I}_c$ | 92.9 | 3.35 | 94.1 | 2.38 | 95.2 | 1.67 | 93.9 | 2.69 | 94.7 | 2.73 |
| | $\mathcal{I}_{cB}$ | 92.9 | 3.09 | 94.6 | 2.21 | 94.3 | 1.57 | 93.6 | 2.60 | 93.7 | 2.34 |
| 0.6 | $\mathcal{I}_{cG}$ | 91.6 | 4.71 | 92.9 | 3.37 | 93.9 | 2.39 | 92.2 | 3.33 | 92.3 | 3.28 |
| | $\mathcal{I}_{cE}$ | 94.2 | 5.78 | 95.5 | 4.84 | 96.7 | 3.90 | 92.7 | 5.16 | 96.4 | 5.43 |
| | $\mathcal{I}_{cK}$ | 91.9 | 4.86 | 90.7 | 3.46 | 91.2 | 2.56 | 92.8 | 4.26 | 88.8 | 3.37 |
| | $\mathcal{I}_{cH}$ | 96.9 | 6.02 | 96.1 | 4.46 | 96.6 | 3.18 | 94.7 | 4.72 | 98.0 | 5.38 |
| | $\mathcal{I}_c$ | 93.2 | 3.87 | 94.9 | 2.78 | 96.1 | 1.97 | 94.4 | 3.14 | 95.0 | 3.16 |
| | $\mathcal{I}_{cB}$ | 94.0 | 3.26 | 94.9 | 2.31 | 94.7 | 1.64 | 93.8 | 2.79 | 95.3 | 2.49 |
| 0.8 | $\mathcal{I}_{cG}$ | 90.2 | 5.81 | 92.8 | 4.34 | 93.0 | 3.05 | 92.8 | 4.15 | 89.1 | 3.72 |
| | $\mathcal{I}_{cE}$ | 88.3 | 5.61 | 95.3 | 4.87 | 96.5 | 3.96 | 86.9 | 5.08 | 96.1 | 5.45 |
| | $\mathcal{I}_{cK}$ | 92.8 | 5.12 | 91.6 | 3.59 | 90.2 | 2.68 | 92.3 | 4.39 | 89.4 | 3.43 |
| | $\mathcal{I}_{cH}$ | 97.6 | 5.75 | 96.2 | 4.48 | 95.3 | 3.31 | 92.4 | 4.44 | 98.6 | 5.40 |

Table 3.20: RB (%) and MSE (×100) for point estimators of $J$ when the LLOD equals 15% quantile of $F_0$ (lognormal distributions).

| | $(n_0, n_1)$ | (50, 50) | | (100, 100) | | (200, 200) | | (50, 150) | | (150, 50) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $J$ | | RB | MSE | RB | MSE | RB | MSE | RB | MSE | RB | MSE |
| 0.2 | $\hat{J}$ | 7.56 | 0.50 | 4.49 | 0.26 | 1.35 | 0.14 | 5.47 | 0.27 | 4.94 | 0.41 |
| | $\hat{J}_B$ | 5.45 | 0.48 | 3.00 | 0.24 | 0.80 | 0.13 | 4.20 | 0.26 | 2.02 | 0.38 |
| | $\hat{J}_G$ | 34.64 | 1.13 | 33.44 | 0.80 | 30.59 | 0.57 | 33.94 | 0.83 | 34.23 | 1.02 |
| | $\hat{J}_E$ | 29.86 | 0.88 | 20.13 | 0.46 | 12.08 | 0.22 | 24.42 | 0.54 | 24.80 | 0.68 |
| 0.4 | $\hat{J}$ | 3.76 | 0.56 | 2.21 | 0.27 | 0.62 | 0.15 | 2.71 | 0.30 | 1.92 | 0.45 |
| | $\hat{J}_B$ | 3.29 | 0.52 | 1.79 | 0.25 | 0.51 | 0.14 | 2.26 | 0.28 | 1.26 | 0.40 |
| | $\hat{J}_G$ | 8.26 | 0.69 | 8.27 | 0.39 | 6.94 | 0.25 | 7.99 | 0.42 | 8.23 | 0.59 |
| | $\hat{J}_E$ | 14.11 | 0.91 | 9.35 | 0.44 | 5.61 | 0.23 | 11.48 | 0.54 | 10.90 | 0.67 |
| 0.6 | $\hat{J}$ | 2.43 | 0.46 | 1.27 | 0.22 | 0.42 | 0.12 | 1.59 | 0.24 | 1.08 | 0.35 |
| | $\hat{J}_B$ | 2.21 | 0.41 | 1.12 | 0.19 | 0.28 | 0.11 | 1.33 | 0.22 | 1.05 | 0.31 |
| | $\hat{J}_G$ | 1.04 | 0.48 | 0.96 | 0.24 | 0.26 | 0.13 | 0.65 | 0.27 | 0.79 | 0.39 |
| | $\hat{J}_E$ | 7.97 | 0.69 | 5.15 | 0.34 | 3.18 | 0.18 | 6.48 | 0.41 | 5.96 | 0.51 |
| 0.8 | $\hat{J}$ | 1.73 | 0.26 | 0.84 | 0.13 | 0.33 | 0.07 | 1.03 | 0.14 | 0.82 | 0.20 |
| | $\hat{J}_B$ | 1.27 | 0.21 | 0.63 | 0.10 | 0.17 | 0.06 | 0.73 | 0.12 | 0.68 | 0.16 |
| | $\hat{J}_G$ | -0.64 | 0.29 | -0.62 | 0.15 | -0.92 | 0.08 | -1.00 | 0.17 | -0.71 | 0.23 |
| | $\hat{J}_E$ | 4.60 | 0.39 | 2.91 | 0.19 | 1.87 | 0.10 | 3.70 | 0.23 | 3.24 | 0.28 |

Table 3.21: RB (%) and MSE (×100) for point estimators of $c$ when the LLOD equals 15% quantile of $F_0$ (lognormal distributions).

| | $(n_0, n_1)$ | (50, 50) | | (100, 100) | | (200, 200) | | (50, 150) | | (150, 50) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $J$ | | RB | MSE | RB | MSE | RB | MSE | RB | MSE | RB | MSE |
| 0.2 | $\hat{c}$ | -0.17 | 225.03 | -0.22 | 112.48 | 0.09 | 58.83 | -0.54 | 148.20 | 0.37 | 143.62 |
| | $\hat{c}_B$ | -0.25 | 164.45 | -0.15 | 75.43 | 0.09 | 37.46 | -0.30 | 108.84 | 0.08 | 88.84 |
| | $\hat{c}_G$ | -6.76 | 321.47 | -5.66 | 181.59 | -5.08 | 118.25 | -6.36 | 244.43 | -5.32 | 181.49 |
| | $\hat{c}_E$ | -2.94 | 554.67 | -1.49 | 346.36 | -0.38 | 225.67 | -3.20 | 450.15 | -0.39 | 452.45 |
| 0.4 | $\hat{c}$ | 0.07 | 116.35 | 0.06 | 61.77 | 0.06 | 30.91 | -0.25 | 81.46 | 0.51 | 80.19 |
| | $\hat{c}_B$ | -0.23 | 92.05 | 0.06 | 45.31 | 0.02 | 22.06 | -0.16 | 60.59 | 0.15 | 54.55 |
| | $\hat{c}_G$ | -1.49 | 148.7 | -0.54 | 72.78 | 0.02 | 36.59 | -0.84 | 82.54 | 0.04 | 70.48 |
| | $\hat{c}_E$ | -1.59 | 346.94 | -0.53 | 244.02 | -0.37 | 150.03 | -1.53 | 285.84 | 0.21 | 300.37 |
| 0.6 | $\hat{c}$ | -0.22 | 90.66 | -0.02 | 46.55 | -0.11 | 24.47 | -0.42 | 61.71 | 0.34 | 58.21 |
| | $\hat{c}_B$ | -0.06 | 72.49 | 0.05 | 35.93 | -0.06 | 17.92 | -0.18 | 47.64 | 0.16 | 44.06 |
| | $\hat{c}_G$ | -0.53 | 151.34 | 0.84 | 79.10 | 1.38 | 41.43 | 0.75 | 80.73 | 1.23 | 75.74 |
| | $\hat{c}_E$ | -1.20 | 282.61 | -0.57 | 182.46 | -0.12 | 115.12 | -1.50 | 237.71 | 0.37 | 230.93 |
| 0.8 | $\hat{c}$ | -0.41 | 105.32 | -0.23 | 51.60 | -0.26 | 27.89 | -0.68 | 68.33 | 0.09 | 67.62 |
| | $\hat{c}_B$ | 0.00 | 74.60 | -0.01 | 36.17 | -0.12 | 17.85 | -0.29 | 53.10 | 0.25 | 41.60 |
| | $\hat{c}_G$ | -1.29 | 235.86 | 0.54 | 116.04 | 1.16 | 65.54 | 1.00 | 114.92 | 0.40 | 120.53 |
| | $\hat{c}_E$ | -1.37 | 292.03 | -0.55 | 189.9 | -0.77 | 128.77 | -1.46 | 232.99 | 0.44 | 235.25 |

Table 3.22: CP (%) and AL for CIs of $J$ when the LLOD equals 15% quantile of $F_0$ (lognormal distributions).

| | $(n_0, n_1)$ | $(50, 50)$ | | $(100, 100)$ | | $(200, 200)$ | | $(50, 150)$ | | $(150, 50)$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $J$ | | CP | AL | CP | AL | CP | AL | CP | AL | CP | AL |
| 0.2 | $\mathcal{I}_J$ | 95.4 | 0.29 | 94.3 | 0.20 | 93.8 | 0.14 | 94.1 | 0.21 | 94.2 | 0.25 |
| | $\mathcal{I}_{JB}$ | 95.5 | 0.28 | 94.5 | 0.20 | 93.7 | 0.14 | 95.3 | 0.20 | 93.8 | 0.25 |
| | $\mathcal{I}_{JG}$ | 86.3 | 0.32 | 76.2 | 0.23 | 65.0 | 0.16 | 75.4 | 0.24 | 82.6 | 0.29 |
| | $\mathcal{I}_{JE}$ | 75.4 | 0.26 | 77.7 | 0.19 | 79.0 | 0.14 | 70.2 | 0.20 | 77.8 | 0.24 |
| 0.4 | $\mathcal{I}_J$ | 95.8 | 0.29 | 95.9 | 0.21 | 94.5 | 0.15 | 95.2 | 0.21 | 95.2 | 0.26 |
| | $\mathcal{I}_{JB}$ | 95.8 | 0.29 | 95.4 | 0.21 | 93.6 | 0.14 | 95.6 | 0.21 | 95.0 | 0.26 |
| | $\mathcal{I}_{JG}$ | 92.3 | 0.30 | 89.7 | 0.22 | 86.1 | 0.15 | 88.7 | 0.23 | 89.8 | 0.27 |
| | $\mathcal{I}_{JE}$ | 80.0 | 0.27 | 82.5 | 0.20 | 83.6 | 0.15 | 75.2 | 0.21 | 82.8 | 0.25 |
| 0.6 | $\mathcal{I}_J$ | 95.6 | 0.26 | 96.0 | 0.19 | 94.7 | 0.13 | 94.9 | 0.19 | 95.3 | 0.23 |
| | $\mathcal{I}_{JB}$ | 95.3 | 0.25 | 95.9 | 0.18 | 94.5 | 0.13 | 95.6 | 0.19 | 96.0 | 0.22 |
| | $\mathcal{I}_{JG}$ | 93.4 | 0.27 | 94.5 | 0.19 | 93.6 | 0.14 | 93.7 | 0.20 | 92.5 | 0.24 |
| | $\mathcal{I}_{JE}$ | 81.4 | 0.25 | 81.7 | 0.18 | 83.5 | 0.14 | 75.9 | 0.19 | 82.3 | 0.23 |
| 0.8 | $\mathcal{I}_J$ | 96.8 | 0.20 | 95.6 | 0.14 | 95 | 0.10 | 95.5 | 0.15 | 95.1 | 0.18 |
| | $\mathcal{I}_{JB}$ | 94.9 | 0.18 | 96.2 | 0.13 | 95.2 | 0.09 | 96.1 | 0.14 | 95.1 | 0.16 |
| | $\mathcal{I}_{JG}$ | 94.2 | 0.21 | 95.0 | 0.15 | 95.2 | 0.11 | 95.1 | 0.15 | 93.7 | 0.19 |
| | $\mathcal{I}_{JE}$ | 81.5 | 0.18 | 85.2 | 0.14 | 85.8 | 0.10 | 77.6 | 0.14 | 83.7 | 0.17 |

Table 3.23: CP (%) and AL for CIs of $c$ when the LLOD equals 15% quantile of $F_0$ (lognormal distributions).

| | $(n_0, n_1)$ | $(50, 50)$ | | $(100, 100)$ | | $(200, 200)$ | | $(50, 150)$ | | $(150, 50)$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $J$ | | CP | AL | CP | AL | CP | AL | CP | AL | CP | AL |
| 0.2 | $\mathcal{I}_c$ | 93.7 | 5.96 | 94.3 | 4.22 | 95.0 | 3.05 | 93.5 | 4.84 | 93.5 | 4.83 |
| | $\mathcal{I}_{cB}$ | 84.6 | 5.35 | 84.1 | 2.50 | 83.5 | 1.70 | 84.9 | 3.20 | 79.1 | 2.95 |
| | $\mathcal{I}_{cG}$ | 82.8 | 5.55 | 81.6 | 3.73 | 71.5 | 2.57 | 81.9 | 4.41 | 82.2 | 4.22 |
| | $\mathcal{I}_{cE}$ | 92.9 | 7.55 | 95.3 | 6.47 | 95.3 | 5.42 | 90.8 | 6.77 | 95.4 | 7.16 |
| 0.4 | $\mathcal{I}_c$ | 93.0 | 4.27 | 93.4 | 3.11 | 95.3 | 2.27 | 92.6 | 3.52 | 91.9 | 3.55 |
| | $\mathcal{I}_{cB}$ | 88.8 | 3.03 | 88.9 | 2.12 | 88.7 | 1.51 | 87.1 | 2.35 | 89.2 | 2.41 |
| | $\mathcal{I}_{cG}$ | 94.4 | 4.58 | 93.2 | 3.29 | 94.3 | 2.34 | 93.6 | 3.51 | 94.2 | 3.22 |
| | $\mathcal{I}_{cE}$ | 94.5 | 6.31 | 96.2 | 5.25 | 96.8 | 4.38 | 93.1 | 5.64 | 96.5 | 5.96 |
| 0.6 | $\mathcal{I}_c$ | 91.7 | 3.72 | 93.2 | 2.72 | 94.5 | 1.97 | 93.5 | 3.04 | 94.3 | 3.12 |
| | $\mathcal{I}_{cB}$ | 91.6 | 3.04 | 92.2 | 2.15 | 92.0 | 1.52 | 90.9 | 2.36 | 93.4 | 2.44 |
| | $\mathcal{I}_{cG}$ | 94.5 | 4.70 | 94.0 | 3.36 | 93.7 | 2.42 | 94.3 | 3.40 | 94.6 | 3.26 |
| | $\mathcal{I}_{cE}$ | 93.9 | 5.74 | 95.3 | 4.79 | 96.5 | 3.88 | 92.3 | 5.10 | 95.8 | 5.40 |
| 0.8 | $\mathcal{I}_c$ | 93.6 | 3.97 | 94.8 | 2.91 | 95.0 | 2.10 | 93.5 | 3.23 | 94.1 | 3.33 |
| | $\mathcal{I}_{cB}$ | 94.2 | 3.34 | 95.8 | 2.36 | 95.6 | 1.67 | 94.3 | 2.77 | 94.4 | 2.60 |
| | $\mathcal{I}_{cG}$ | 94.8 | 5.86 | 94.9 | 4.29 | 94.7 | 3.03 | 95.6 | 4.21 | 94.7 | 3.84 |
| | $\mathcal{I}_{cE}$ | 88.6 | 5.57 | 95.1 | 4.84 | 96.0 | 3.94 | 86.5 | 5.06 | 95.5 | 5.48 |

Table 3.24: RB (%) and MSE (×100) for point estimators of $J$ when the LLOD equals 30% quantile of $F_0$ (lognormal distributions).

| $J$ | $(n_0, n_1)$ | (50, 50) RB | (50, 50) MSE | (100, 100) RB | (100, 100) MSE | (200, 200) RB | (200, 200) MSE | (50, 150) RB | (50, 150) MSE | (150, 50) RB | (150, 50) MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\hat{J}$ | 8.39 | 0.52 | 4.76 | 0.26 | 1.57 | 0.14 | 6.10 | 0.28 | 5.56 | 0.43 |
| | $\hat{J}_B$ | 4.89 | 0.50 | 2.83 | 0.25 | 0.80 | 0.14 | 4.59 | 0.26 | -1.52 | 0.44 |
| | $\hat{J}_G$ | 47.93 | 1.70 | 45.32 | 1.25 | 42.55 | 0.94 | 46.74 | 1.32 | 46.29 | 1.50 |
| | $\hat{J}_E$ | 29.86 | 0.89 | 20.03 | 0.46 | 12.13 | 0.22 | 24.31 | 0.53 | 24.86 | 0.69 |
| 0.4 | $\hat{J}$ | 3.81 | 0.57 | 2.26 | 0.27 | 0.59 | 0.16 | 2.83 | 0.31 | 1.99 | 0.46 |
| | $\hat{J}_B$ | 3.36 | 0.52 | 1.95 | 0.25 | 0.45 | 0.14 | 2.35 | 0.28 | 1.45 | 0.43 |
| | $\hat{J}_G$ | 10.94 | 0.85 | 10.33 | 0.49 | 9.08 | 0.32 | 10.50 | 0.55 | 10.25 | 0.68 |
| | $\hat{J}_E$ | 14.00 | 0.90 | 9.36 | 0.43 | 5.54 | 0.23 | 11.46 | 0.54 | 10.91 | 0.67 |
| 0.6 | $\hat{J}$ | 2.52 | 0.48 | 1.28 | 0.22 | 0.42 | 0.12 | 1.73 | 0.26 | 1.06 | 0.37 |
| | $\hat{J}_B$ | 2.21 | 0.41 | 1.10 | 0.19 | 0.27 | 0.11 | 1.36 | 0.22 | 1.05 | 0.31 |
| | $\hat{J}_G$ | 0.65 | 0.54 | 0.41 | 0.26 | -0.31 | 0.14 | 0.30 | 0.32 | 0.18 | 0.42 |
| | $\hat{J}_E$ | 7.95 | 0.69 | 5.15 | 0.34 | 3.16 | 0.17 | 6.51 | 0.41 | 5.95 | 0.51 |
| 0.8 | $\hat{J}$ | 1.85 | 0.27 | 0.93 | 0.13 | 0.35 | 0.07 | 1.14 | 0.14 | 0.77 | 0.20 |
| | $\hat{J}_B$ | 1.16 | 0.21 | 0.56 | 0.10 | 0.15 | 0.06 | 0.62 | 0.12 | 0.52 | 0.16 |
| | $\hat{J}_G$ | -1.38 | 0.36 | -1.51 | 0.18 | -1.92 | 0.11 | -1.78 | 0.22 | -1.84 | 0.26 |
| | $\hat{J}_E$ | 4.57 | 0.38 | 2.91 | 0.19 | 1.86 | 0.10 | 3.65 | 0.23 | 3.17 | 0.28 |

Table 3.25: RB (%) and MSE (×100) for point estimators of $c$ when the LLOD equals 30% quantile of $F_0$ (lognormal distributions).

| $J$ | $(n_0, n_1)$ | (50, 50) RB | (50, 50) MSE | (100, 100) RB | (100, 100) MSE | (200, 200) RB | (200, 200) MSE | (50, 150) RB | (50, 150) MSE | (150, 50) RB | (150, 50) MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\hat{c}$ | 0.65 | 258.41 | 0.43 | 136.56 | 0.48 | 73.65 | 0.18 | 174.52 | 1.09 | 184.57 |
| | $\hat{c}_B$ | -0.50 | 281.55 | -0.31 | 93.98 | 0.06 | 43.73 | -0.48 | 127.55 | -1.24 | 154.60 |
| | $\hat{c}_G$ | -3.91 | 233.83 | -2.66 | 117.81 | -2.33 | 62.51 | -3.21 | 154.52 | -2.20 | 121.42 |
| | $\hat{c}_E$ | -2.88 | 543.75 | -1.48 | 340.90 | -0.36 | 227.91 | -3.10 | 447.73 | -0.34 | 448.97 |
| 0.4 | $\hat{c}$ | 0.91 | 138.59 | 0.60 | 75.44 | 0.29 | 39.32 | 0.36 | 96.11 | 1.06 | 96.88 |
| | $\hat{c}_B$ | 0.09 | 101.96 | 0.15 | 52.63 | 0.08 | 24.67 | -0.20 | 69.73 | 0.22 | 69.45 |
| | $\hat{c}_G$ | 1.86 | 155.21 | 2.88 | 95.61 | 3.35 | 69.84 | 2.76 | 103.54 | 3.45 | 103.53 |
| | $\hat{c}_E$ | -1.56 | 349.31 | -0.54 | 243.76 | -0.41 | 151.41 | -1.56 | 283.63 | 0.22 | 300.01 |
| 0.6 | $\hat{c}$ | 0.38 | 103.45 | 0.34 | 51.80 | 0.14 | 28.56 | 0.00 | 69.41 | 0.80 | 68.51 |
| | $\hat{c}_B$ | 0.13 | 79.98 | 0.14 | 38.93 | -0.01 | 19.06 | -0.21 | 51.75 | 0.47 | 49.46 |
| | $\hat{c}_G$ | 2.69 | 171.73 | 4.13 | 122.81 | 4.62 | 100.98 | 4.08 | 130.54 | 4.46 | 135.43 |
| | $\hat{c}_E$ | -1.12 | 281.90 | -0.61 | 183.18 | -0.09 | 115.00 | -1.54 | 239.01 | 0.35 | 228.98 |
| 0.8 | $\hat{c}$ | 0.00 | 111.15 | 0.02 | 56.67 | -0.10 | 29.70 | -0.44 | 74.27 | 0.41 | 75.64 |
| | $\hat{c}_B$ | 0.13 | 77.44 | 0.06 | 37.09 | -0.07 | 18.56 | -0.21 | 52.89 | 0.39 | 44.46 |
| | $\hat{c}_G$ | 1.28 | 236.85 | 3.12 | 156.06 | 3.86 | 113.08 | 3.47 | 165.46 | 2.88 | 155.10 |
| | $\hat{c}_E$ | -1.36 | 292.78 | -0.56 | 189.16 | -0.74 | 128.34 | -1.43 | 232.44 | 0.44 | 233.30 |

Table 3.26: CP (%) and AL for CIs of $J$ when the LLOD equals 30% quantile of $F_0$ (lognormal distributions).

| $J$ | $(n_0, n_1)$ | $(50, 50)$ CP | AL | $(100, 100)$ CP | AL | $(200, 200)$ CP | AL | $(50, 150)$ CP | AL | $(150, 50)$ CP | AL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\mathcal{I}_J$ | 94.7 | 0.29 | 94.2 | 0.20 | 94.1 | 0.14 | 93.8 | 0.21 | 94.5 | 0.25 |
| | $\mathcal{I}_{JB}$ | 94.5 | 0.28 | 93.5 | 0.20 | 93.5 | 0.14 | 95.4 | 0.21 | 90.5 | 0.23 |
| | $\mathcal{I}_{JG}$ | 78.3 | 0.35 | 67.7 | 0.25 | 48.2 | 0.18 | 64.5 | 0.27 | 74.7 | 0.31 |
| | $\mathcal{I}_{JE}$ | 71.2 | 0.24 | 75.1 | 0.18 | 76.0 | 0.13 | 67.2 | 0.19 | 72.8 | 0.22 |
| 0.4 | $\mathcal{I}_J$ | 96.5 | 0.29 | 96.0 | 0.21 | 94.4 | 0.15 | 95.3 | 0.21 | 95.2 | 0.26 |
| | $\mathcal{I}_{JB}$ | 96.0 | 0.29 | 95.2 | 0.20 | 93.9 | 0.14 | 95.5 | 0.21 | 94.1 | 0.25 |
| | $\mathcal{I}_{JG}$ | 89.2 | 0.32 | 87.8 | 0.23 | 83.5 | 0.16 | 86.4 | 0.24 | 88.2 | 0.28 |
| | $\mathcal{I}_{JE}$ | 75.3 | 0.25 | 78.9 | 0.19 | 80.7 | 0.14 | 72.5 | 0.20 | 78.3 | 0.23 |
| 0.6 | $\mathcal{I}_J$ | 95.2 | 0.26 | 96.0 | 0.19 | 94.7 | 0.13 | 94.8 | 0.19 | 94.7 | 0.23 |
| | $\mathcal{I}_{JB}$ | 95.4 | 0.25 | 96.0 | 0.18 | 94.8 | 0.13 | 95.5 | 0.19 | 95.6 | 0.23 |
| | $\mathcal{I}_{JG}$ | 94.0 | 0.29 | 94.3 | 0.20 | 93.8 | 0.14 | 94.1 | 0.22 | 92.3 | 0.25 |
| | $\mathcal{I}_{JE}$ | 78.4 | 0.24 | 80.3 | 0.18 | 82.1 | 0.13 | 75.2 | 0.18 | 80.4 | 0.21 |
| 0.8 | $\mathcal{I}_J$ | 96.4 | 0.20 | 95.2 | 0.14 | 95.2 | 0.10 | 95.2 | 0.15 | 95.4 | 0.18 |
| | $\mathcal{I}_{JB}$ | 95.4 | 0.19 | 96.1 | 0.14 | 94.5 | 0.10 | 96.6 | 0.14 | 94.4 | 0.17 |
| | $\mathcal{I}_{JG}$ | 94.6 | 0.23 | 95.2 | 0.16 | 93.5 | 0.11 | 94.9 | 0.17 | 95.0 | 0.20 |
| | $\mathcal{I}_{JE}$ | 80.3 | 0.18 | 83.9 | 0.14 | 84.4 | 0.10 | 76.6 | 0.14 | 82.2 | 0.16 |

Table 3.27: CP (%) and AL for CIs of $c$ when the LLOD equals 30% quantile of $F_0$ (lognormal distributions).

| $J$ | $(n_0, n_1)$ | $(50, 50)$ CP | AL | $(100, 100)$ CP | AL | $(200, 200)$ CP | AL | $(50, 150)$ CP | AL | $(150, 50)$ CP | AL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.2 | $\mathcal{I}_c$ | 90.0 | 6.50 | 91.2 | 4.63 | 93.7 | 3.41 | 92.0 | 5.26 | 89.9 | 5.26 |
| | $\mathcal{I}_{cB}$ | 80.9 | 5.72 | 80.6 | 2.76 | 79.3 | 1.73 | 75.9 | 2.88 | 80.6 | 4.26 |
| | $\mathcal{I}_{cG}$ | 90.2 | 5.64 | 91.6 | 3.84 | 90.5 | 2.69 | 91.2 | 4.38 | 91.3 | 4.15 |
| | $\mathcal{I}_{cE}$ | 92.4 | 7.15 | 95.0 | 6.28 | 95.0 | 5.27 | 90.3 | 6.44 | 94.4 | 6.88 |
| 0.4 | $\mathcal{I}_c$ | 91.1 | 4.67 | 93.2 | 3.40 | 93.1 | 2.51 | 91.9 | 3.89 | 91.4 | 3.83 |
| | $\mathcal{I}_{cB}$ | 86.6 | 3.16 | 86.5 | 2.12 | 86.0 | 1.49 | 80.2 | 2.12 | 90.3 | 3.02 |
| | $\mathcal{I}_{cG}$ | 95.1 | 4.72 | 92.5 | 3.36 | 85.7 | 2.39 | 92.6 | 3.62 | 90.1 | 3.28 |
| | $\mathcal{I}_{cE}$ | 94.0 | 6.15 | 95.8 | 5.17 | 96.2 | 4.34 | 92.7 | 5.48 | 96.8 | 5.85 |
| 0.6 | $\mathcal{I}_c$ | 92.2 | 3.97 | 94.0 | 2.94 | 94.0 | 2.14 | 93.7 | 3.30 | 93.7 | 3.33 |
| | $\mathcal{I}_{cB}$ | 89.6 | 2.98 | 91.0 | 2.08 | 89.6 | 1.47 | 85.3 | 2.14 | 92.0 | 2.57 |
| | $\mathcal{I}_{cG}$ | 95.0 | 4.77 | 89.3 | 3.39 | 78.0 | 2.43 | 89.6 | 3.55 | 87.2 | 3.29 |
| | $\mathcal{I}_{cE}$ | 93.6 | 5.61 | 95.5 | 4.75 | 96.4 | 3.85 | 92.1 | 5.00 | 96.0 | 5.35 |
| 0.8 | $\mathcal{I}_c$ | 92.7 | 4.08 | 94.1 | 3.02 | 95.0 | 2.20 | 92.5 | 3.34 | 93.3 | 3.45 |
| | $\mathcal{I}_{cB}$ | 93.1 | 3.30 | 95.2 | 2.33 | 94.7 | 1.64 | 93.2 | 2.63 | 95.0 | 2.66 |
| | $\mathcal{I}_{cG}$ | 96.3 | 5.97 | 94.2 | 4.25 | 87.6 | 3.04 | 93.4 | 4.34 | 94.2 | 3.91 |
| | $\mathcal{I}_{cE}$ | 88.3 | 5.55 | 94.6 | 4.81 | 96.1 | 3.95 | 86.2 | 5.00 | 95.4 | 5.42 |

# Chapter 4

# Empirical Likelihood Inference on Gini Indices of Two Semicontinuous Populations under Density Ratio Models

## 4.1 Introduction

The Gini index, first proposed by Gini (1912), has been widely used to measure population inequality. In economic studies, it is an important measure of income or wealth inequality among individuals or households in a particular population (Wang et al., 2016; Peng, 2011). In life expectancy studies, it is used to describe the concentration of survival times and to evaluate inequality among people in the target population (Bonetti et al., 2009; Lv et al., 2017). The index is closely related to the Lorenz curve (Lorenz, 1905), a widely used measure for the size distribution of income or wealth. It is the ratio of the area between the Lorenz curve and the 45-degree line to the area under the 45-degree line. Hence, the Gini index ranges from 0 to 1, with 0 indicating perfect equality and 1 for extreme inequality.

Study variables such as income and survival time are often modelled by using a positive continuous distribution. One important scenario in applications is that there are two related populations, each containing a sizeable zero values for the study variable. The inferential problems can be on the Gini index for each population separately or the difference of the two Gini indices. The scenario is quite common in practice but efficient inferential procedures are not available in the existing literature.

In this chapter, we propose new semiparametric inference procedures for the Gini indices of two semicontinuous populations. Specifically, suppose that we have two independent samples from two related populations with values of the study variable $X$ generated from the following mixture models:

$$(X_{i1}, \cdots, X_{in_i}) \sim F_i(x) = \nu_i I(x \geq 0) + (1 - \nu_i) I(x > 0) G_i(x), \quad \text{for} \quad i = 0, 1, \qquad (4.1)$$

where $\nu_i$ is the zero proportion in population $i$, $n_i$ is the sample size for sample $i$, $I(\cdot)$ is an indicator function, and $G_i(\cdot)$ is the CDF of the positive observations in sample $i$. For population $i = 0, 1$, the Gini index can be equivalently defined (David, 1968) as

$$\mathcal{G}_i = \frac{D_i}{2\mu_i}, \qquad (4.2)$$

where $D_i = E(|X_{i1} - X_{i2}|)$ is the expected absolute difference of $X$ for two randomly selected units from population $i$ and $\mu_i = E(X_{i1})$ is the expectation of population $i$. Our discussions in this chapter focus on statistical inferences on $\mathcal{G}_0$, $\mathcal{G}_1$, and $\mathcal{G}_0 - \mathcal{G}_1$. It is worth mentioning that although our results are presented for cases where the two populations contain excess zeros for the study variable, the proposed methods and the theoretical results are also applicable to cases without excess zeros, i.e., $\nu_i = 0$ in model (4.1). In addition, inferences on a general function of $\mathcal{G}_0$ and $\mathcal{G}_1$ can also be conducted. See Section 4.2 for further discussion.

Samples with positive outcomes only, i.e., $\nu_i = 0$ in model (4.1), are common in studies of family income or wealth or a country's gross domestic products (Gastwirth, 1972; Cowell, 2011). For instance, all the household incomes are positive in the 1997 Family and Income and Expenditure Survey conducted by the Philippine Statistics Authority. More details can be found in Section 4.4. Samples with a mixture of excess zero values and skewed positive outcomes, i.e., $\nu_i > 0$ in model (4.1), naturally arise in studies of expenditure data and health cost data (Zhou and Tu, 1999, 2000; Zhou and Cheng, 2008). For example, Zhou and Cheng (2008) presented a dataset from the assessment of inpatient charges (see Section 4.4), and most patients with uncomplicated hypertension had no hospitalization and therefore zero costs. This chapter systematically studies both cases ($\nu_i = 0$ and $\nu_i > 0$) in a unified framework via model (4.1).

Many studies of the Gini index have applied nonparametric methods. For example, point estimators of $\mathcal{G}_0$, $\mathcal{G}_1$, and $\mathcal{G}_0 - \mathcal{G}_1$ and their asymptotic variance estimation have been discussed in Hoeffding (1948), Anand (1983), Ogwang (2000), Giles (2004), Modarres and Gastwirth (2006), Yitzhaki (1991), Karagiannis and Kovacevic (2000), and Davidson (2009). See Wang and Zhao (2016) for a detailed review. Qin et al. (2010) and Peng

(2011) used the empirical likelihood method (Owen, 2001) to construct CIs for the index. More recently, Wang et al. (2016) derived the jackknife empirical likelihood (JEL). Peng (2011) and Wang and Zhao (2016) compared two Gini indices of independent or correlated populations using the empirical likelihood method and the JEL method, respectively.

Fully nonparametric methods enjoy the robustness against the model misspecifications. However, these methods ignore the characteristics common to the two samples and/or the relation between the two populations, which have been shown to be useful for more efficient statistical inferences; see, for instance, the studies on the strengths of lumber produced in Canada in different years (Chen and Liu, 2013; Cai et al., 2017; Cai and Chen, 2018). Ignoring such information may result in less efficient inference procedures.

To combine the information from the two samples without making risky parametric distributional assumptions, we use the DRM (1.1) to link the CDFs of the positive observations $G_0$ and $G_1$ in model (4.1). Let $dG_i$ be the probability density function of $G_i$, $i = 0, 1$. The DRM postulates that

$$dG_1(x) = \exp\{\alpha + \boldsymbol{\beta}^\top \boldsymbol{q}(x)\} dG_0(x) = \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} dG_0(x)\,, \tag{4.3}$$

where $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top)^\top$ is the vector of unknown parameters and $\boldsymbol{Q}(x) = (1, \boldsymbol{q}(x)^\top)^\top$ with $\boldsymbol{q}(x)$ being a $d$-dimensional basis function. As far as we are aware, inferential procedures for two Gini indices $\mathcal{G}_0$, $\mathcal{G}_1$ and their difference $\mathcal{G}_0 - \mathcal{G}_1$ have not been explored under the mixture model (4.1) and the DRM (4.3). This chapter aims to fill this void.

The rest of the chapter is organized as follows. In Section 4.2, we first present some preliminary results for the MELEs of all unknown parameters and the MELEs of general functionals. These results serve as preparations for studying the estimators $\mathcal{G}_0$ and $\mathcal{G}_1$. After that, we propose the MELEs of $\mathcal{G}_0$, $\mathcal{G}_1$, and $\mathcal{G}_0 - \mathcal{G}_1$ and investigate their asymptotic properties . We construct CIs and conduct hypothesis tests on $\mathcal{G}_0$, $\mathcal{G}_1$, and $\mathcal{G}_0 - \mathcal{G}_1$ based on the theoretical results. Results from simulation studies are presented in Section 4.3, and applications to two real-world datasets are given in Section 4.4. Proofs, technical details and additional simulation results are provided in Section 4.5.

## 4.2 Main Results

Let $n_{i0}$ and $n_{i1}$ be the (random) numbers of zero observations and positive observations, respectively, in each sample $i = 0, 1$. Clearly, $n_i = n_{i0} + n_{i1}$, for $i = 0, 1$. Without loss of generality, we assume that the first $n_{i1}$ observations in group $i$, $X_{i1}, \cdots, X_{in_{i1}}$, are positive, and the remaining $n_{i0}$ observations are 0. Let $n$ be the total (fixed) sample size, i.e., $n = n_0 + n_1$.

We first investigate the estimators of parameters in models (4.1) and (4.3) and the estimators of a class of functionals. These help to obtain the estimators of $\mathcal{G}_0$ and $\mathcal{G}_1$, and study the asymptotic properties of the estimators of numerator and denominator of the Gini index in (4.2).

### 4.2.1 Estimation of model parameters

With the two samples of observations from model (4.1), the full likelihood function is

$$\mathcal{L}_n = \prod_{i=0}^{1} \left\{ \nu_i^{n_{i0}} (1-\nu_i)^{n_{i1}} \prod_{j=1}^{n_{i1}} dG_i(X_{ij}) \right\}.$$

Following the empirical likelihood principle (Owen, 2001), we estimate the baseline distribution $G_0(x)$ as

$$G_0^*(x) = \sum_{i=0}^{1} \sum_{j=1}^{n_{i1}} p_{ij} I(X_{ij} \le x), \tag{4.4}$$

where $p_{ij} = dG_0(X_{ij})$ for $i = 0, 1$ and $j = 1, \ldots, n_{i1}$. With (4.4) and under the DRM (4.3), the full likelihood function can be rewritten as

$$\mathcal{L}_n = \prod_{i=0}^{1} \nu_i^{n_{i0}} (1-\nu_i)^{n_{i1}} \cdot \left\{ \prod_{i=0}^{1} \prod_{j=1}^{n_{i1}} p_{ij} \right\} \cdot \left[ \prod_{j=1}^{n_{11}} \exp\left\{ \boldsymbol{\theta}^\top \boldsymbol{Q}(X_{1j}) \right\} \right],$$

where the $p_{ij}$'s satisfy the constraints

$$p_{ij} > 0, \quad \sum_{i=0}^{1} \sum_{j=1}^{n_{i1}} p_{ij} = 1, \quad \text{and} \quad \sum_{i=0}^{1} \sum_{j=1}^{n_{i1}} p_{ij} \exp\left\{ \boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij}) \right\} = 1. \tag{4.5}$$

These constraints ensure that the estimates of $G_0$ and $G_1$ are CDFs.

Let $\boldsymbol{P} = \{p_{ij}\}$ and $\boldsymbol{\nu} = (\nu_0, \nu_1)^\top$. The MELE of $(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{P})$ is then defined as

$$(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{P}}) = \arg\max_{\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{P}} \mathcal{L}_n$$

subject to the constraints in (4.5). We write the logarithm of the empirical likelihood function $\mathcal{L}_n$ as

$$\tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) = \ell_0\left(\boldsymbol{\nu}\right) + \tilde{\ell}_1\left(\boldsymbol{\theta}, \boldsymbol{P}\right), \tag{4.6}$$

where

$$\ell_0\left(\boldsymbol{\nu}\right) = \sum_{i=0}^{1} \log\left\{\nu_i^{n_{i0}}\left(1-\nu_i\right)^{n_{i1}}\right\} \text{ and } \tilde{\ell}_1\left(\boldsymbol{\theta}, \boldsymbol{P}\right) = \sum_{j=1}^{n_{11}} \left\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{1j})\right\} + \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \log p_{ij}.$$

Here $\ell_0\left(\boldsymbol{\nu}\right)$ is the binomial log-likelihood function corresponding to the number of zero observations, and $\tilde{\ell}_1\left(\boldsymbol{\theta}, \boldsymbol{P}\right)$ represents the empirical log-likelihood function associated with the positive observations.

Following Wang et al. (2017a), we have $\hat{\boldsymbol{\nu}} = \arg\max_{\boldsymbol{\nu}} \ell_0(\boldsymbol{\nu})$ and

$$(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{P}}) = \arg\max_{\boldsymbol{\theta},\boldsymbol{P}} \left\{\tilde{\ell}_1\left(\boldsymbol{\theta}, \boldsymbol{P}\right): \ p_{ij} > 0,\ \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} p_{ij} = 1,\ \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} p_{ij} \exp\left\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\right\} = 1\right\}.$$

By the method of Lagrange multipliers, $\hat{\boldsymbol{\theta}}$ can be obtained by maximizing the following dual empirical log-likelihood function (Cai et al., 2017):

$$\ell_1(\boldsymbol{\theta}) = -\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \log\left\{1 + \hat{\rho}[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1]\right\} + \sum_{j=1}^{n_{11}} \{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\}, \tag{4.7}$$

where $\hat{\rho} = n_{11}(n_{01} + n_{11})^{-1}$. That is, $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell_1(\boldsymbol{\theta})$. Note that $\hat{\rho}$ is a random variable in our setup. This is fundamentally different from the case where there is no excess of zeros in the data (Qin and Zhang, 1997), and it creates theoretical challenges for our asymptotic development.

Once $\hat{\boldsymbol{\theta}}$ is obtained, the MELEs of the $\hat{p}_{ij}$'s are

$$\hat{p}_{ij} = (n_{01} + n_{11})^{-1} \left\{1 + \hat{\rho}[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} - 1]\right\}^{-1}, \tag{4.8}$$

and the MELEs of $G_0(x)$ and $G_1(x)$ are

$$\hat{G}_0(x) = \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \hat{p}_{ij} I(X_{ij} \le x), \quad \hat{G}_1(x) = \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \hat{p}_{ij} \exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} I(X_{ij} \le x). \tag{4.9}$$

We now study the asymptotic properties of $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\nu}}^\top, \hat{\rho}, \hat{\boldsymbol{\theta}}^\top)^\top$. For ease of presentation, we introduce some notation. We use $\boldsymbol{\nu}^*$ and $\boldsymbol{\theta}^*$ to denote the true values of $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$, respectively. Let $w_i = n_i/n$ and

$$\Delta^* = \sum_{i=0}^{1} w_i(1-\nu_i^*),\ \rho^* = \frac{w_1(1-\nu_1^*)}{\Delta^*},\ \omega(x) = \exp\{\boldsymbol{\theta}^{*\top}\boldsymbol{Q}(x)\},$$
$$h(x) = 1+\rho^*\{\omega(x)-1\},\ h_1(x) = \rho^*\omega(x)/h(x),$$
$$\boldsymbol{A}_{\boldsymbol{\nu}} = \text{diag}\left\{\frac{w_0}{\nu_0^*(1-\nu_0^*)}, \frac{w_1}{\nu_1^*(1-\nu_1^*)}\right\},\ \boldsymbol{A}_{\boldsymbol{\theta}} = \Delta^*(1-\rho^*)E_0\left\{h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}(X)^\top\right\},$$

where $E_0(\cdot)$ represents the expectation operator with respect to $G_0$ and $X$ refers to a random variable from $G_0$. Noting that although $\omega(\cdot)$, $h(\cdot)$, and $h_1(\cdot)$ also depend on $\boldsymbol{\theta}^*$ and/or $\rho^*$, we drop these redundant parameters for notational simplicity.

The following theorem establishes the asymptotic normality of $\hat{\boldsymbol{\eta}}$.

**Theorem 4.1.** *Let $\boldsymbol{\eta}^* = (\boldsymbol{\nu}^{*\top}, \rho^*, \boldsymbol{\theta}^{*\top})^\top$. Assume Conditions C1–C4 in Section 4.5.1 are satisfied. As the total sample size $n\to\infty$,*

$$n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \to N(\boldsymbol{0}, \boldsymbol{\Lambda}),$$

*in distribution, where*

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\nu}}^{-1} & \rho^*(1-\rho^*)\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^\top & \boldsymbol{0} \\ \rho^*(1-\rho^*)\boldsymbol{W}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1} & (\Delta^*)^{-1}\rho^*(1-\rho^*)\{\rho^*\nu_0^* + (1-\rho^*)\nu_1^*\} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)} \end{pmatrix}$$

*with $\boldsymbol{W} = ((1-\nu_0^*)^{-1}, -(1-\nu_1^*)^{-1})$ and $\boldsymbol{e} = (1, \boldsymbol{0}_{d\times 1}^\top)^\top$.*

Qin and Zhang (1997) considered the asymptotic normality of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ when there is no excess of zeros in the data. Theorem 4.1 generalizes their results to the case where the data contain excessive zeros. Furthermore, it establishes the joint limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$, $\sqrt{n}(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*)$, and $\sqrt{n}(\hat{\rho} - \rho^*)$, where the latter two are induced by the semicontinuous data structure.

### 4.2.2 Estimation of a class of functionals

Under the mixture model (4.1) and the DRM (4.3), we consider a class of functionals $\boldsymbol{\gamma}$ of length $p$, defined as

$$\boldsymbol{\gamma} = \int_0^\infty \boldsymbol{u}(x; \boldsymbol{\nu}, \boldsymbol{\theta}) dG_0(x), \tag{4.10}$$

where $\boldsymbol{u}(x; \boldsymbol{\nu}, \boldsymbol{\theta}) = (u_1(x; \boldsymbol{\nu}, \boldsymbol{\theta}), \ldots, u_p(x; \boldsymbol{\nu}, \boldsymbol{\theta}))^\top$ is a given $(p \times 1)$-dimensional function.

The functional $\boldsymbol{\gamma}$ include $\mu_i$'s in (4.2) as special cases. Let

$$u_1(x; \boldsymbol{\nu}, \boldsymbol{\theta}) = (1 - \nu_0)x \quad \text{and} \quad u_2(x; \boldsymbol{\nu}, \boldsymbol{\theta}) = (1 - \nu_1)x \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\}.$$

Then $\boldsymbol{\gamma} = (\mu_0, \mu_1)^\top$. More examples of $\boldsymbol{\gamma}$ can be found in Yuan et al. (2021d). As we will show in Section 4.5.1, the MELEs of $D_i$'s in (4.2) can be approximated by the MELEs of some specific $\boldsymbol{\gamma}$. In the following, we construct the MELE of $\boldsymbol{\gamma}$ in (4.10) and study its asymptotic properties. This will pave our road to study the MELEs of $\mathcal{G}_0$ and $\mathcal{G}_1$.

By the definition of $\boldsymbol{\gamma}$ in (4.10), $\boldsymbol{\gamma}$ is a function of $(\boldsymbol{\nu}, \boldsymbol{\theta})$ and $G_0$. Replacing them with $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})$ and $\hat{G}_0$, the MELE of $\boldsymbol{\gamma}$ is

$$\hat{\boldsymbol{\gamma}} = \int_0^\infty \boldsymbol{u}(x; \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}) d\hat{G}_0(x) = \sum_{i=0}^{1} \sum_{j=1}^{n_{i1}} \hat{p}_{ij} \boldsymbol{u}(X_{ij}; \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}). \tag{4.11}$$

Based on the results in Theorem 4.1, we have the following theorem.

**Theorem 4.2.** *Let $\boldsymbol{\gamma}^*$ be the true value of $\boldsymbol{\gamma}$. Assume that $\boldsymbol{u}(x; \boldsymbol{\nu}, \boldsymbol{\theta})$ is twice continuously differentiable with respect to $(\boldsymbol{\nu}, \boldsymbol{\theta})$, $||\partial^2 \boldsymbol{u}(x; \boldsymbol{\nu}, \boldsymbol{\theta})/\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^\top||$, $||\partial^2 \boldsymbol{u}(x; \boldsymbol{\nu}, \boldsymbol{\theta})/\partial \boldsymbol{\nu} \partial \boldsymbol{\theta}^\top||$, $||\partial^2 \boldsymbol{u}(x; \boldsymbol{\nu}, \boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top||$ are bounded by some integrable function with respect to $G_0$ in the neighbourhood of $(\boldsymbol{\nu}^*, \boldsymbol{\theta}^*)$, the variance-covariance matrix of $\boldsymbol{u}(X; \boldsymbol{\nu}^*, \boldsymbol{\theta}^*)$ with respect to $G_0$ is positive definite, and $E_0\{\partial \boldsymbol{u}(X; \boldsymbol{\nu}^*, \boldsymbol{\theta}^*)/\partial \boldsymbol{\nu}\}$ and $E_0\{\partial \boldsymbol{u}(X; \boldsymbol{\nu}^*, \boldsymbol{\theta}^*)/\partial \boldsymbol{\theta}\}$ have ranks two and $d$ respectively. Under the conditions of Theorem 4.1, as $n \to \infty$, we have*

$$\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*) \to N(\boldsymbol{0}, \boldsymbol{\Gamma})$$

*in distribution, where*

$$\begin{aligned} \boldsymbol{\Gamma} &= \frac{1}{\Delta^*} E_0 \left\{ \frac{\boldsymbol{u}(X; \boldsymbol{\nu}^*, \boldsymbol{\theta}^*) \boldsymbol{u}(X; \boldsymbol{\nu}^*, \boldsymbol{\theta}^*)^\top}{h(X)} \right\} - \frac{\boldsymbol{\gamma}^* \boldsymbol{\gamma}^{*\top}}{\Delta^*} \\ &\quad + \mathcal{M}_1 \boldsymbol{A}_{\boldsymbol{\nu}}^{-1} \mathcal{M}_1^\top - \frac{\mathcal{M}_2 \mathcal{M}_2^\top}{\Delta^* \rho^* (1 - \rho^*)} + \mathcal{M}_3 \boldsymbol{A}_{\boldsymbol{\theta}}^{-1} \mathcal{M}_3^\top, \end{aligned} \tag{4.12}$$

*with*

$$\begin{aligned}
\mathcal{M}_1 &= E_0\left\{\frac{\partial \boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)}{\partial \boldsymbol{\nu}}\right\},\\
\mathcal{M}_2 &= E_0\left[\{\partial \boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}\}\,\boldsymbol{e}\right] - \rho^*\boldsymbol{\gamma}^*,\\
\mathcal{M}_3 &= E_0\left\{\partial \boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta} - h_1(X)\boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)\boldsymbol{Q}(X)^\top\right\}.
\end{aligned}$$

Li et al. (2018) derived a similar result in their Theorem 2.1 for $\hat{\boldsymbol{\gamma}}$ when there is no excess of zeros in the data and $p = 1$. Theorem 4.2 covers the case with excessive zeros. The two results complement each other to cover both cases.

### 4.2.3 Estimation of Gini indices

We now move to the estimation of the Gini indices $\mathcal{G}_0$ and $\mathcal{G}_1$. In Section 4.5.1, we show that $\mathcal{G}_i$ in (4.2) can be equivalently expressed as

$$\mathcal{G}_i = (2\nu_i - 1) + (1-\nu_i)\psi_i/m_i, \tag{4.13}$$

where $m_i = \int_0^\infty x dG_i(x)$ and $\psi_i = \int_0^\infty \{2xG_i(x)\}dG_i(x)$. Using the alternative form given in (4.13) and the MELEs of $G_0$ and $G_1$ in (4.9), the MELE of the two Gini indices are given by

$$\hat{\mathcal{G}}_i = (2\hat{\nu}_i - 1) + (1-\hat{\nu}_i)\hat{\psi}_i/\hat{m}_i, \quad i = 0, 1, \tag{4.14}$$

where

$$\hat{m}_i = \int_0^\infty x d\hat{G}_i(x) \ \text{ and } \ \hat{\psi}_i = \int_0^\infty \left\{2x\hat{G}_i(x)\right\}d\hat{G}_i(x).$$

**Remark.** *We comment that the MELEs of the two Gini indices in (4.14) are also applicable to the case where there is no excess of zero values, i.e.,* $\boldsymbol{\nu} = (0,0)$ *and* $n_{i1} = n_i$*. We need to set* $\hat{\nu}_i = 0$ *and obtain* $\boldsymbol{\theta}$ *by maximizing* $\ell_1(\boldsymbol{\theta})$ *in (4.7) with* $\hat{\rho} = n_1(n_0+n_1)^{-1}$*; then the MELEs in (4.9) and (4.14) can be directly applied.*

Let $\mathcal{G}_0^*$ and $\mathcal{G}_1^*$ be the true values of Gini indices $\mathcal{G}_0$ and $\mathcal{G}_1$. Define

$$\boldsymbol{J} = \begin{pmatrix} -\frac{\mathcal{G}_0^*}{m_0} & \frac{1}{m_0} & 0 & 0 \\ 0 & 0 & -\frac{\mathcal{G}_1^*}{m_1} & \frac{1}{m_1} \end{pmatrix},$$

$$u_0(x) = (2\nu_0^* - 1)x + (1-\nu_0^*)\left[2\left\{xG_0(x) + \int_x^\infty y dG_0(y)\right\} - \psi_0\right],$$

$$u_1(x) = (2\nu_1^* - 1)x + (1-\nu_1^*)\left[2\left\{xG_1(x) + \int_x^\infty y dG_1(y)\right\} - \psi_1\right].$$

The following theorem establishes the asymptotic normality of the MELEs $(\hat{\mathcal{G}}_0, \hat{\mathcal{G}}_1)$.

**Theorem 4.3.** *Assume Conditions C1–C5 in Section 4.5.1 are satisfied. As the total sample size $n \to \infty$,*

$$n^{1/2}\begin{pmatrix} \hat{\mathcal{G}}_0 - \mathcal{G}_0^* \\ \hat{\mathcal{G}}_1 - \mathcal{G}_1^* \end{pmatrix} \to N(\boldsymbol{0}, \boldsymbol{\Sigma})$$

*in distribution with the asymptotic variance-covariance matrix*

$$\boldsymbol{\Sigma} = \frac{1}{\Delta^*}\boldsymbol{J}\left[E_0\left\{\frac{\boldsymbol{u}(X)\boldsymbol{u}(X)^\top}{h(X)}\right\} + \frac{1}{(\rho^*)^2}\boldsymbol{B}\right]\boldsymbol{J}^\top + diag\left\{\frac{\nu_0^*(1-\mathcal{G}_0^*)^2}{\Delta^*(1-\rho^*)}, \frac{\nu_1^*(1-\mathcal{G}_1^*)^2}{\Delta^*\rho^*}\right\}, \quad (4.15)$$

*where*

$$\boldsymbol{u}(x) = (x, u_0(x), \omega(x)x, \omega(x)u_0(x))^\top, \ \tilde{\boldsymbol{u}}_0(x) = -\rho^*(x, u_0(x))^\top,$$
$$\tilde{\boldsymbol{u}}_1(x) = (1-\rho^*)(x, u_1(x))^\top, \ \tilde{\boldsymbol{u}}(x) = \left(\tilde{\boldsymbol{u}}_0(x)^\top, \tilde{\boldsymbol{u}}_1(x)^\top\right)^\top,$$
$$\boldsymbol{B} = E_0\{h_1(X)\tilde{\boldsymbol{u}}(X)\boldsymbol{Q}(X)^\top\}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}E_0\{h_1(X)\boldsymbol{Q}(X)\tilde{\boldsymbol{u}}(X)^\top\}.$$

Repeating all the steps in the proof of Theorem 4.3, we obtain a similar result for cases where $\nu_i^* = 0$ for $i = 0, 1$ in the following theorem.

**Theorem 4.4.** *Assume that Conditions C2–C5 are satisfied. When there is no excess of zeros, i.e., $\nu_i^* = 0$ for $i = 0, 1$, the joint distribution of $\sqrt{n}(\hat{\mathcal{G}}_0 - \mathcal{G}_0^*)$ and $\sqrt{n}(\hat{\mathcal{G}}_1 - \mathcal{G}_1^*)$ asymptotically follows a bivariate normal distribution with mean zero and variance in (4.15) with $\nu_i^*$ being replaced by 0.*

Since the proposed method utilizes more information to obtain the MELEs of Gini indices, we expect that the proposed MELEs are more efficient than fully nonparametric estimators. With the alternative form of the Gini index in (4.13), the fully nonparametric estimators of the two Gini indices for sample $i = 0, 1$ are

$$\tilde{\mathcal{G}}_i = (2\hat{\nu}_i - 1) + (1 - \hat{\nu}_i)\tilde{\psi}_i/\tilde{m}_i, \quad i = 0, 1, \tag{4.16}$$

where

$$\tilde{m}_i = {n_{i1}}^{-1}\sum_{j=1}^{n_{i1}} X_{ij}, \quad \tilde{\psi}_i = \int_0^\infty \{2x\tilde{G}_i(x)\}d\tilde{G}_i(x),$$

and $\tilde{G}_i(x) = {n_{i1}}^{-1}\sum_{j=1}^{n_{i1}} I(X_{ij} \le x)$ is the empirical CDF of the positive observations in sample $i$.

The following theorem compares the proposed estimators MELEs $\hat{\mathcal{G}}_i$ and the nonparametric estimators $\tilde{\mathcal{G}}_i$ in terms of their asymptotic variance-covariance matrices. Recall that $\mathbf{\Sigma}$ is the asymptotic variance-covariance matrix of $(\hat{\mathcal{G}}_0, \hat{\mathcal{G}}_1)$ given in Theorem 4.3.

**Theorem 4.5.** *Assume that Conditions C2–C5 in Section 4.5.1 are satisfied.*

*(a) For the nonparametric estimators $(\tilde{\mathcal{G}}_0, \tilde{\mathcal{G}}_1)$ and as $n \to \infty$, we have*

$$\sqrt{n}\begin{pmatrix} \tilde{\mathcal{G}}_0 - \mathcal{G}_0^* \\ \tilde{\mathcal{G}}_1 - \mathcal{G}_1^* \end{pmatrix} \to N(\mathbf{0}, \mathbf{\Sigma}_{non})$$

*in distribution, where the variance-covariance matrix*

$$\begin{aligned}\mathbf{\Sigma}_{non} &= \boldsymbol{J}\, diag\left\{\frac{E_0\{\tilde{\boldsymbol{u}}_0(X)\tilde{\boldsymbol{u}}_0(X)^\top\}}{\Delta^*(\rho^*)^2(1-\rho)}, \frac{E_0\{\omega(X)\tilde{\boldsymbol{u}}_1(X)\tilde{\boldsymbol{u}}_1(X)^\top\}}{\Delta^*\rho^*(1-\rho)^2}\right\}\boldsymbol{J}^\top \\ &\quad + diag\left\{\frac{\nu_0^*(1-\mathcal{G}_0^*)^2}{\Delta^*(1-\rho^*)}, \frac{\nu_1^*(1-\mathcal{G}_1^*)^2}{\Delta^*\rho^*}\right\}.\end{aligned}$$

*(b) The two asymptotic variance-covariance matrices $\mathbf{\Sigma}_{non}$ and $\mathbf{\Sigma}$ satisfy*

$$\mathbf{\Sigma}_{non} - \mathbf{\Sigma} = \frac{1}{\Delta^*(\rho^*)^2(1-\rho^*)}\boldsymbol{J}E_0\{h_1(X)\boldsymbol{D}(X)\boldsymbol{D}(X)^\top\}\boldsymbol{J}^\top \ge \mathbf{0},$$

*where $\boldsymbol{D}(x) = \left(\boldsymbol{D}_0(x)^\top, \boldsymbol{D}_1(x)^\top\right)^\top$ for $x > 0$ and*

$$\boldsymbol{D}_i(x) = \tilde{\boldsymbol{u}}_i(x) - \Delta^*(1-\rho^*)E_0\left\{h_1(X)\tilde{\boldsymbol{u}}_i(X)\boldsymbol{Q}(X)^\top\right\}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{Q}(x), \; i = 0, 1.$$

Note that $\mathbf{\Sigma}_{non} - \mathbf{\Sigma}$ is a positive semidefinite matrix, which implies that the proposed MELEs for the Gini indices are at least as efficient as the nonparametric estimators. Our simulation results reported in Section 4.3 confirm this result. It is worth mentioning that the theorem is applicable whether or not there are excess zero values.

### 4.2.4 Inference on functions of Gini indices

Under the current setting of two samples, we may be interested in performing inference on the Gini index for only one of the samples or other functions of the two Gini indices, such as their difference. The results of Theorems 4.3 and 4.4 can be used to develop the following theorem for parameters which are a general function of the two Gini indices.

**Theorem 4.6.** *Assume the conditions of Theorem 4.5 hold. Let $\phi(\cdot,\cdot)$ be a bivariate smooth function. As $n \to \infty$, we have $\sqrt{n}\{\phi(\hat{\mathcal{G}}_0, \hat{\mathcal{G}}_1) - \phi(\mathcal{G}_0^*, \mathcal{G}_1^*)\} \to N(0, \sigma_\phi^2)$ in distribution with*

$$\sigma_\phi^2 = \left( \frac{\partial\phi(\mathcal{G}_0^*, \mathcal{G}_1^*)}{\partial\mathcal{G}_0}, \frac{\partial\phi(\mathcal{G}_0^*, \mathcal{G}_1^*)}{\partial\mathcal{G}_1} \right) \mathbf{\Sigma} \left( \frac{\partial\phi(\mathcal{G}_0^*, \mathcal{G}_1^*)}{\partial\mathcal{G}_0}, \frac{\partial\phi(\mathcal{G}_0^*, \mathcal{G}_1^*)}{\partial\mathcal{G}_1} \right)^\top .$$

With the results in Theorems 4.5 and 4.6, we can easily show that $\sigma_\phi^2$ is no larger than the asymptotic variance of the fully nonparametric estimator $\phi(\tilde{\mathcal{G}}_0, \tilde{\mathcal{G}}_1)$. That is, utilizing the information from both samples via the DRM (4.3) improves the estimation of $\phi(\mathcal{G}_0, \mathcal{G}_1)$.

The general form $\phi(\cdot,\cdot)$ covers many interesting functions of $\mathcal{G}_0$ and $\mathcal{G}_1$. For example, when $\phi(x_1, x_2) = \text{logit}(x_1) = \log\{x_1/(1-x_1)\}$, the parameter $\phi(\mathcal{G}_0, \mathcal{G}_1)$ represents the logit transformation of the Gini index $\mathcal{G}_0$; when $\phi(x_1, x_2) = x_1 - x_2$, the parameter $\phi(\mathcal{G}_0, \mathcal{G}_1)$ refers to the difference of two Gini indices.

The variance $\sigma_\phi^2$ may depend on $\mathcal{G}_0^*$, $\mathcal{G}_1^*$, and $(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{P})$. Replacing these unknown quantities by their MELEs leads to a consistent estimator $\hat{\sigma}_\phi^2$ of $\sigma_\phi^2$. Together with the result in Theorem 4.6, we have as $n \to \infty$,

$$\sqrt{n}\{\phi(\hat{\mathcal{G}}_0, \hat{\mathcal{G}}_1) - \phi(\mathcal{G}_0^*, \mathcal{G}_1^*)\}/\hat{\sigma}_\phi \to N(0, 1)$$

in distribution. Hence, $\sqrt{n}\{\phi(\hat{\mathcal{G}}_0, \hat{\mathcal{G}}_1) - \phi(\mathcal{G}_0^*, \mathcal{G}_1^*)\}/\hat{\sigma}_\phi$ is asymptotically pivotal and can be used to construct CIs and to conduct hypothesis tests on $\phi(\mathcal{G}_0, \mathcal{G}_1)$.

For ease of presentation, we use $\hat{\phi}$ and $\phi$ to denote $\phi(\hat{\mathcal{G}}_0, \hat{\mathcal{G}}_1)$ and $\phi(\mathcal{G}_0, \mathcal{G}_1)$. Then the $100(1-a)\%$ Wald-type CI for $\phi$ is given by

$$[\hat{\phi} - z_{1-a/2}\hat{\sigma}_\phi/\sqrt{n}, \hat{\phi} + z_{1-a/2}\hat{\sigma}_\phi/\sqrt{n}],$$

where $z_{1-a/2}$ is the $(1-a/2)$ quantile of the standard normal distribution. When testing $H_0 : \phi = 0$, we reject the null hypothesis if $|\sqrt{n}\hat{\phi}/\hat{\sigma}_\phi| > z_{1-a/2}$ at the significance level $a$.

## 4.3 Simulation Studies

In this section, we compare finite-sample performances of our semiparametric methods with existing methods of inferences on the Gini indices through simulation studies. We focus on three inferential problems:

(1) Point estimation for $\mathcal{G}_0$, $\mathcal{G}_1$, and $\mathcal{G}_0 - \mathcal{G}_1$;

(2) Confidence intervals on $\mathcal{G}_0$, $\mathcal{G}_1$, and $\mathcal{G}_0 - \mathcal{G}_1$;

(3) Hypothesis testing on $H_0 : \mathcal{G}_0 = \mathcal{G}_1$.

We conduct the simulation studies under two distributional settings: (i) $G_0$ and $G_1$ are the CDFs of $\chi^2_3$ and $\chi^2_4$; and (ii) $G_0$ and $G_1$ are the CDFs of $Exp(0.5)$ and $Exp(1)$. Here $\chi^2_k$ represents the chi-square distribution with $k$ degrees of freedom, and $Exp(k)$ refers to the exponential distribution with the rate parameter $k$. The proposed inference procedures under the DRM are implemented with the correctly specified $\boldsymbol{q}(x)$, where $\boldsymbol{q}(x) = \log(x)$ in the $\chi^2$ setting and $\boldsymbol{q}(x) = x$ in the exponential setting. For each scenario, we consider two combinations of sample sizes, $(n_0, n_1) = (100, 100)$, $(300, 300)$, and the results are based on 2,000 simulation runs.

### 4.3.1 Performance of point estimators

We start by exploring the performance of the point estimators. We consider the following three estimators:

- *EMP:* $\tilde{\mathcal{G}}_0$, $\tilde{\mathcal{G}}_1$, and $\tilde{\mathcal{G}}_0 - \tilde{\mathcal{G}}_1$, where $\tilde{\mathcal{G}}_i$ is the nonparametric estimator given in (4.16) for $i = 0, 1$;

- *JEL:* $\bar{\mathcal{G}}_0$, $\bar{\mathcal{G}}_1$, and $\bar{\mathcal{G}}_0 - \bar{\mathcal{G}}_1$, which are the JEL estimators defined in Wang et al. (2016), where
$$\bar{\mathcal{G}}_i = (2\tilde{\mu}_i)^{-1} \binom{n}{2}^{-1} \sum_{1 \le j_1 < j_2 \le n_i} |X_{ij_1} - X_{ij_2}|$$
with $\tilde{\mu}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$ for $i = 0, 1$;

– *DRM:* $\hat{\mathcal{G}}_0$, $\hat{\mathcal{G}}_1$, and $\hat{\mathcal{G}}_0 - \hat{\mathcal{G}}_1$, where $\hat{\mathcal{G}}_i$ is the MELE given in (4.14) for $i = 0, 1$.

Three combinations of $\boldsymbol{\nu}$ are considered for the zero population proportions: $(0, 0)$, $(0.3, 0.3)$, $(0.7, 0.7)$. We evaluate the performance of a point estimator in terms of the bias and the MSE. Tables 4.1 and 4.2 present the simulated results for different settings.

Table 4.1: Bias ($\times$1000) and MSE ($\times$1000) for point estimators ($\chi^2$ distributions).

| | | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | | $\mathcal{G}_0 - \mathcal{G}_1$ | |
|---|---|---|---|---|---|---|---|---|
| $(n_0, n_1)$ | $\boldsymbol{\nu}$ | | Bias | MSE | Bias | MSE | Bias | MSE |
| (100,100) | (0,0) | EMP | 5.37 | 0.74 | 7.80 | 0.62 | -2.43 | 1.29 |
| | | JEL | -0.39 | 0.73 | 1.57 | 0.57 | -1.96 | 1.31 |
| | | DRM | 2.06 | 0.37 | 4.18 | 0.40 | -2.13 | 0.31 |
| | (0.3,0.3) | EMP | 6.17 | 1.19 | 6.19 | 1.21 | -0.02 | 2.35 |
| | | JEL | 2.16 | 1.18 | 1.83 | 1.20 | 0.33 | 2.40 |
| | | DRM | 2.60 | 0.95 | 3.04 | 1.06 | -0.44 | 1.71 |
| | (0.7,0.7) | EMP | 6.56 | 0.91 | 6.01 | 1.01 | 0.54 | 1.86 |
| | | JEL | 4.88 | 0.91 | 4.18 | 1.01 | 0.70 | 1.89 |
| | | DRM | 2.70 | 0.79 | 2.97 | 0.91 | -0.28 | 1.56 |
| (300,300) | (0,0) | EMP | 1.70 | 0.23 | 2.31 | 0.19 | -0.61 | 0.40 |
| | | JEL | -0.22 | 0.23 | 0.23 | 0.19 | -0.45 | 0.40 |
| | | DRM | 0.66 | 0.13 | 1.12 | 0.14 | -0.46 | 0.11 |
| | (0.3,0.3) | EMP | 2.52 | 0.39 | 1.75 | 0.41 | 0.78 | 0.79 |
| | | JEL | 1.18 | 0.39 | 0.29 | 0.41 | 0.90 | 0.80 |
| | | DRM | 0.94 | 0.32 | 0.89 | 0.37 | 0.06 | 0.57 |
| | (0.7,0.7) | EMP | 2.84 | 0.31 | 2.40 | 0.34 | 0.44 | 0.67 |
| | | JEL | 2.28 | 0.31 | 1.78 | 0.34 | 0.49 | 0.67 |
| | | DRM | 1.30 | 0.27 | 1.50 | 0.31 | -0.20 | 0.55 |

We observe from Tables 4.1 and 4.2 that the biases of the estimators of $\mathcal{G}_0$ and $\mathcal{G}_1$ are acceptable for all three methods under all scenarios. The EMP estimators $\tilde{\mathcal{G}}_0$ and $\tilde{\mathcal{G}}_1$ always give the largest biases. When the proportions of zero values are small, i.e., $\boldsymbol{\nu} = (0, 0)$ or $(0.3, 0.3)$, the biases of the JEL estimators $\bar{\mathcal{G}}_0$ and $\bar{\mathcal{G}}_1$ are the smallest. The DRM estimators $\hat{\mathcal{G}}_0$ and $\hat{\mathcal{G}}_1$ have a clear advantage in terms of bias when $\boldsymbol{\nu} = (0.7, 0.7)$. The performance of the EMP estimators $\tilde{\mathcal{G}}_0$ and $\tilde{\mathcal{G}}_1$ and the JEL estimators $\bar{\mathcal{G}}_0$ and $\bar{\mathcal{G}}_1$ is similar in terms of the MSE. The DRM estimators $\hat{\mathcal{G}}_0$ and $\hat{\mathcal{G}}_1$ give the smallest MSEs in all cases; this agrees with the result in Theorem 4.5. The MSEs of all the estimators decrease as $\boldsymbol{\nu}$ moves toward $(0, 0)$ or the sample size increases.

For the estimators of the difference $\mathcal{G}_0 - \mathcal{G}_1$, we find that the biases of all the estimators are relatively small in all cases. The biases of the DRM estimator $\hat{\mathcal{G}}_0 - \hat{\mathcal{G}}_1$ are usually the

Table 4.2: Bias (×1000) and MSE (×1000) for point estimators (exponential distributions).

| | | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | | $\mathcal{G}_0 - \mathcal{G}_1$ | |
|---|---|---|---|---|---|---|---|---|
| $(n_0, n_1)$ | $\boldsymbol{\nu}$ | | Bias | MSE | Bias | MSE | Bias | MSE |
| (100,100) | (0,0) | EMP | 4.59 | 0.81 | 5.94 | 0.86 | -1.35 | 1.59 |
| | | JEL | -0.42 | 0.81 | 0.95 | 0.84 | -1.37 | 1.63 |
| | | DRM | 1.55 | 0.66 | 2.82 | 0.41 | -1.27 | 0.58 |
| | (0.3,0.3) | EMP | 4.85 | 1.09 | 3.53 | 1.10 | 1.31 | 2.14 |
| | | JEL | 1.36 | 1.09 | 0.03 | 1.11 | 1.33 | 2.19 |
| | | DRM | 1.64 | 0.96 | 0.73 | 0.82 | 0.91 | 1.46 |
| | (0.7,0.7) | EMP | 5.17 | 0.81 | 3.62 | 0.80 | 1.55 | 1.55 |
| | | JEL | 3.71 | 0.81 | 2.14 | 0.80 | 1.57 | 1.58 |
| | | DRM | 1.81 | 0.73 | 1.24 | 0.65 | 0.56 | 1.20 |
| (300,300) | (0,0) | EMP | 1.78 | 0.28 | 1.97 | 0.27 | -0.18 | 0.57 |
| | | JEL | 0.11 | 0.28 | 0.30 | 0.27 | -0.19 | 0.58 |
| | | DRM | 0.76 | 0.22 | 0.96 | 0.13 | -0.20 | 0.21 |
| | (0.3,0.3) | EMP | 1.80 | 0.37 | 1.48 | 0.36 | 0.32 | 0.73 |
| | | JEL | 0.63 | 0.37 | 0.31 | 0.36 | 0.32 | 0.74 |
| | | DRM | 0.74 | 0.34 | 0.73 | 0.27 | 0.01 | 0.51 |
| | (0.7,0.7) | EMP | 1.81 | 0.26 | 1.98 | 0.26 | -0.16 | 0.53 |
| | | JEL | 1.32 | 0.26 | 1.48 | 0.26 | -0.16 | 0.54 |
| | | DRM | 0.80 | 0.24 | 0.90 | 0.22 | -0.10 | 0.43 |

smallest. The MSEs of the EMP estimator for $\mathcal{G}_0 - \mathcal{G}_1$ and JEL estimator for $\mathcal{G}_0 - \mathcal{G}_1$ are very close, whereas the MSEs of the DRM estimator are significantly smaller than those of the other two estimators. For instance, the MSE of $\hat{\mathcal{G}}_0 - \hat{\mathcal{G}}_1$ is less than 25% of the MSEs of $\tilde{\mathcal{G}}_0 - \tilde{\mathcal{G}}_1$ and $\bar{\mathcal{G}}_0 - \bar{\mathcal{G}}_1$ when the simulated samples come from $\chi^2$ distributions with $(n_0, n_1) = (100, 100)$ and $\boldsymbol{\nu} = (0, 0)$.

We conducted additional simulations with $\boldsymbol{\nu} = (0.1, 0.3)$ and $(0.6, 0.4)$; the results show similar patterns and are presented in the Section 4.5.2.

### 4.3.2 Performance of confidence intervals

We examine and compare the performance of the following CIs for the Gini indices in the simulation studies:

- *NA-EMP:* Wald-type CIs based on the normal approximation (Qin et al., 2010);
- *BT-EMP:* bootstrap-t CIs (Qin et al., 2010);

- *EL*: ELR-based CIs (Qin et al., 2010);
- *BT-EL*: bootstrap ELR-based CIs (Qin et al., 2010);
- *JEL:* jackknife ELR-based CIs (Wang et al., 2016; Wang and Zhao, 2016);
- *AJEL:* adjusted jackknife ELR-based CIs (Wang et al., 2016; Wang and Zhao, 2016);
- *NA-DRM:* Wald-type CIs based on the normal approximation under the DRM;
- *BT-DRM:* bootstrap-t CIs under the DRM.

The EL method, to our best knowledge, has not been used to construct CIs for the difference of two Gini indices in the existing literature. Hence, we consider all the methods, except for *EL* and *BT-EL*, in our comparisons of the CIs for the parameter $\mathcal{G}_0 - \mathcal{G}_1$. For those calibrated by the bootstrap method, we used 1,000 bootstrap samples drawn from the original samples with replacement.

Three combinations of $\boldsymbol{\nu}$ are considered for the zero population proportions: $(0, 0)$, $(0.3, 0.3)$, $(0.7, 0.7)$. We evaluate the performance of a CI in terms of the CP and the AL. Tables 4.3 and 4.4 contain the simulated results for the CIs of $\mathcal{G}_0$ and $\mathcal{G}_1$ under different settings. The simulated results for the CIs of $\mathcal{G}_0 - \mathcal{G}_1$ are shown in Table 4.5.

When the sample sizes are $(100, 100)$, we can see from Tables 4.3 and 4.4 that the NA-EMP and EL CIs for $\mathcal{G}_0$ and $\mathcal{G}_1$ tend to be narrow and have lower CPs, especially when the proportions of zero values are large, i.e., $\boldsymbol{\nu} = (0.7, 0, 7)$. With the help of bootstrap calibration, the BT-EMP and BT-EL CIs achieve better performance in terms of CP. However, when $\boldsymbol{\nu} = (0.7, 0.7)$, the BT-EMP CIs have slight overcoverage with inflated ALs. The AJEL CIs always have the longest ALs, and the JEL CIs are only slightly shorter. Moreover, when $\boldsymbol{\nu} = (0, 0)$ and $(0.3, 0.3)$, the CPs of the JEL and AJEL CIs are close to the nominal level of 95%. The JEL and AJEL CIs suffer from undercoverage when $\boldsymbol{\nu} = (0.7, 0.7)$. The NA-DRM CIs have the shortest ALs, and their CPs are very close to the 95% nominal level in all cases. This is strong evidence that using DRMs improves the performance of the CIs. The bootstrap calibration does little to improve the CIs: the performances of the NA-DRM and BT-DRM CIs are similar.

When the sample sizes increase to $(300, 300)$, the performance of all the CIs becomes satisfactory in terms of CP. The NA-DRM and BT-DRM CIs always have the shortest ALs, and there is little variation among the ALs of the other CIs.

Since the Gini index ranges from 0 to 1, a logit transformation may improve the performance of the CIs for $\mathcal{G}_0$ and $\mathcal{G}_1$ under the DRM. However, the results (reported in Section 4.5.2) show that the transformation does not provide any significant improvement.

Table 4.3: CP(%) and AL of CIs ($\chi^2$ distributions).

| | | (100,100) | | | | (300,300) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | |
| $\boldsymbol{\nu}$ | | CP | AL | CP | AL | CP | AL | CP | AL |
| (0,0) | NA-EMP | 93.85 | 0.100 | 94.20 | 0.092 | 94.60 | 0.059 | 94.80 | 0.054 |
| | BT-EMP | 94.10 | 0.103 | 94.75 | 0.094 | 94.85 | 0.059 | 95.05 | 0.054 |
| | EL | 93.85 | 0.100 | 94.20 | 0.091 | 94.55 | 0.059 | 94.80 | 0.054 |
| | BT-EL | 94.45 | 0.103 | 95.10 | 0.095 | 94.90 | 0.059 | 94.95 | 0.054 |
| | JEL | 94.45 | 0.102 | 94.85 | 0.094 | 94.70 | 0.059 | 95.15 | 0.054 |
| | AJEL | 94.80 | 0.105 | 95.50 | 0.096 | 94.90 | 0.060 | 95.30 | 0.055 |
| | NA-DRM | 95.25 | 0.074 | 94.65 | 0.078 | 94.70 | 0.043 | 94.70 | 0.045 |
| | BT-DRM | 95.55 | 0.075 | 95.00 | 0.079 | 94.55 | 0.043 | 94.55 | 0.046 |
| (0.3,0.3) | NA-EMP | 93.80 | 0.132 | 93.65 | 0.134 | 94.60 | 0.077 | 94.05 | 0.079 |
| | BT-EMP | 95.30 | 0.135 | 94.55 | 0.137 | 95.20 | 0.077 | 94.40 | 0.079 |
| | EL | 93.75 | 0.131 | 93.65 | 0.134 | 94.60 | 0.077 | 94.00 | 0.078 |
| | BT-EL | 94.50 | 0.136 | 94.85 | 0.139 | 94.65 | 0.078 | 94.55 | 0.079 |
| | JEL | 94.45 | 0.137 | 93.80 | 0.141 | 94.50 | 0.078 | 94.55 | 0.080 |
| | AJEL | 95.35 | 0.141 | 94.20 | 0.144 | 94.80 | 0.079 | 94.80 | 0.081 |
| | NA-DRM | 95.10 | 0.120 | 94.35 | 0.130 | 95.45 | 0.070 | 94.90 | 0.076 |
| | BT-DRM | 95.75 | 0.121 | 94.65 | 0.130 | 95.25 | 0.070 | 94.65 | 0.075 |
| (0.7,0.7) | NA-EMP | 92.20 | 0.113 | 92.95 | 0.119 | 94.90 | 0.067 | 93.90 | 0.070 |
| | BT-EMP | 96.75 | 0.122 | 96.55 | 0.128 | 96.30 | 0.068 | 95.40 | 0.072 |
| | EL | 92.35 | 0.111 | 92.90 | 0.117 | 95.15 | 0.067 | 93.75 | 0.070 |
| | BT-EL | 94.70 | 0.120 | 95.30 | 0.127 | 95.75 | 0.069 | 94.55 | 0.072 |
| | JEL | 90.75 | 0.123 | 90.80 | 0.129 | 94.00 | 0.069 | 93.00 | 0.072 |
| | AJEL | 91.35 | 0.127 | 91.55 | 0.133 | 94.25 | 0.070 | 93.10 | 0.073 |
| | NA-DRM | 94.50 | 0.111 | 94.85 | 0.121 | 95.10 | 0.065 | 95.20 | 0.071 |
| | BT-DRM | 95.40 | 0.113 | 95.90 | 0.123 | 95.45 | 0.064 | 95.60 | 0.070 |

We now discuss the simulation results for the CIs of the difference $\mathcal{G}_0 - \mathcal{G}_1$ presented in Table 4.5. We observe that the NA-EMP and BT-EMP CIs have similar performance; their performance is acceptable except when the simulated samples are from $\chi^2$ distributions with $(n_0, n_1) = (100, 100)$ and $\boldsymbol{\nu} = (0, 0)$. In this case, the CPs of the NA-EMP and BT-EMP CIs are below the 95% nominal level. The JEL and AJEL CIs always have the longest ALs. They experience overcoverage in some cases, especially when the proportions of zero values are high. The BT-DRM CIs have the shortest ALs, which leads to undercoverage in some cases. The performance of the NA-DRM CIs is consistently satisfactory in terms of CP and AL.

We also conduct additional simulations with $\boldsymbol{\nu} = (0.1, 0.3)$ and $(0.6, 0.4)$; the results display similar patterns and are presented in the Section 4.5.2.

Table 4.4: CP(%) and AL of CIs (exponential distributions).

| | | (100,100) | | | | (300,300) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | |
| $\boldsymbol{\nu}$ | | CP | AL | CP | AL | CP | AL | CP | AL |
| (0,0) | NA-EMP | 93.85 | 0.110 | 93.50 | 0.111 | 94.65 | 0.065 | 94.45 | 0.065 |
| | BT-EMP | 94.35 | 0.115 | 94.05 | 0.115 | 94.75 | 0.065 | 94.75 | 0.065 |
| | EL | 93.90 | 0.110 | 93.50 | 0.110 | 94.65 | 0.065 | 94.55 | 0.065 |
| | BT-EL | 94.50 | 0.113 | 94.00 | 0.113 | 94.80 | 0.065 | 94.60 | 0.065 |
| | JEL | 94.35 | 0.113 | 93.90 | 0.113 | 94.90 | 0.065 | 94.55 | 0.065 |
| | AJEL | 94.95 | 0.115 | 94.35 | 0.116 | 95.10 | 0.066 | 94.75 | 0.066 |
| | NA-DRM | 94.80 | 0.100 | 94.05 | 0.079 | 93.95 | 0.059 | 95.20 | 0.045 |
| | BT-DRM | 94.45 | 0.104 | 94.75 | 0.079 | 93.65 | 0.060 | 94.95 | 0.045 |
| (0.3,0.3) | NA-EMP | 93.55 | 0.127 | 94.25 | 0.128 | 94.55 | 0.075 | 93.45 | 0.075 |
| | BT-EMP | 95.35 | 0.132 | 94.90 | 0.132 | 94.70 | 0.075 | 93.80 | 0.075 |
| | EL | 93.60 | 0.126 | 94.10 | 0.127 | 94.70 | 0.075 | 93.30 | 0.075 |
| | BT-EL | 94.75 | 0.131 | 94.85 | 0.132 | 95.05 | 0.076 | 93.70 | 0.076 |
| | JEL | 93.80 | 0.132 | 94.55 | 0.133 | 94.65 | 0.076 | 93.65 | 0.076 |
| | AJEL | 94.50 | 0.136 | 95.15 | 0.136 | 95.00 | 0.076 | 93.80 | 0.076 |
| | NA-DRM | 95.55 | 0.124 | 94.95 | 0.112 | 95.15 | 0.073 | 94.60 | 0.065 |
| | BT-DRM | 95.45 | 0.125 | 95.30 | 0.112 | 94.60 | 0.072 | 94.60 | 0.064 |
| (0.7,0.7) | NA-EMP | 91.40 | 0.104 | 92.15 | 0.105 | 94.60 | 0.062 | 94.55 | 0.062 |
| | BT-EMP | 96.30 | 0.114 | 95.70 | 0.115 | 95.85 | 0.064 | 95.50 | 0.064 |
| | EL | 92.05 | 0.102 | 92.35 | 0.102 | 94.70 | 0.062 | 94.55 | 0.062 |
| | BT-EL | 95.00 | 0.110 | 94.20 | 0.111 | 95.40 | 0.064 | 95.25 | 0.064 |
| | JEL | 90.40 | 0.114 | 91.00 | 0.114 | 94.30 | 0.064 | 93.95 | 0.064 |
| | AJEL | 90.85 | 0.117 | 91.60 | 0.118 | 94.40 | 0.064 | 94.05 | 0.064 |
| | NA-DRM | 94.65 | 0.109 | 93.90 | 0.101 | 96.10 | 0.064 | 95.40 | 0.059 |
| | BT-DRM | 95.80 | 0.109 | 95.55 | 0.102 | 95.65 | 0.062 | 95.95 | 0.058 |

### 4.3.3 Performance of tests on the equality of two Gini indices

In this section, we examine the performance of our proposed semiparametric test for testing the equality of the two Gini indices, i.e., $H_0 : \mathcal{G}_0 = \mathcal{G}_1$, with comparisons to other existing methods. We consider the following tests:

- *NA-EMP:* Wald-type test based on the normal approximation of $\tilde{\mathcal{G}}_0 - \tilde{\mathcal{G}}_1$ (Qin et al., 2010);
- *NL-EMP:* Wald-type test based on the normal approximation of $\text{logit}(\tilde{\mathcal{G}}_0) - \text{logit}(\tilde{\mathcal{G}}_1)$;
- *JEL:* jackknife ELR test (Wang and Zhao, 2016);

Table 4.5: CP(%) and AL of CIs for $\mathcal{G}_0 - \mathcal{G}_1$.

| | | (100,100) | | | | (300,300) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\chi^2$ | | $Exp$ | | $\chi^2$ | | $Exp$ | |
| $\boldsymbol{\nu}$ | | CP | AL | CP | AL | CP | AL | CP | AL |
| (0,0) | NA-EMP | 92.69 | 0.137 | 94.90 | 0.157 | 94.65 | 0.079 | 95.15 | 0.092 |
| | BT-EMP | 92.89 | 0.139 | 94.80 | 0.161 | 94.35 | 0.080 | 95.15 | 0.092 |
| | JEL | 93.84 | 0.142 | 95.80 | 0.164 | 95.00 | 0.081 | 95.55 | 0.093 |
| | AJEL | 94.54 | 0.144 | 96.05 | 0.166 | 95.20 | 0.081 | 95.75 | 0.094 |
| | NA-DRM | 94.44 | 0.070 | 94.95 | 0.092 | 94.65 | 0.041 | 94.70 | 0.055 |
| | BT-DRM | 92.94 | 0.069 | 94.95 | 0.092 | 93.80 | 0.041 | 94.65 | 0.054 |
| (0.3,0.3) | NA-EMP | 94.19 | 0.188 | 93.05 | 0.181 | 95.00 | 0.110 | 94.65 | 0.106 |
| | BT-EMP | 94.54 | 0.191 | 93.45 | 0.184 | 95.00 | 0.110 | 94.65 | 0.106 |
| | JEL | 95.45 | 0.202 | 94.75 | 0.195 | 95.55 | 0.113 | 95.35 | 0.108 |
| | AJEL | 95.70 | 0.205 | 95.10 | 0.198 | 95.55 | 0.113 | 95.55 | 0.109 |
| | NA-DRM | 94.24 | 0.165 | 94.30 | 0.149 | 95.00 | 0.096 | 95.05 | 0.087 |
| | BT-DRM | 93.34 | 0.161 | 93.45 | 0.146 | 94.35 | 0.094 | 94.60 | 0.086 |
| (0.7,0.7) | NA-EMP | 93.20 | 0.164 | 94.29 | 0.148 | 94.05 | 0.097 | 94.90 | 0.088 |
| | BT-EMP | 93.65 | 0.170 | 94.14 | 0.153 | 94.05 | 0.098 | 94.90 | 0.089 |
| | JEL | 96.65 | 0.188 | 97.65 | 0.175 | 95.20 | 0.101 | 95.95 | 0.092 |
| | AJEL | 96.90 | 0.192 | 98.00 | 0.179 | 95.50 | 0.102 | 96.05 | 0.093 |
| | NA-DRM | 95.55 | 0.162 | 96.19 | 0.138 | 95.60 | 0.094 | 95.70 | 0.080 |
| | BT-DRM | 93.40 | 0.153 | 94.49 | 0.133 | 94.60 | 0.090 | 95.10 | 0.078 |

- *AJEL:* adjusted jackknife ELR test (Wang and Zhao, 2016);
- *NA-DRM:* Wald-type test based on the normal approximation of $\hat{\mathcal{G}}_0 - \hat{\mathcal{G}}_1$ under the DRM;
- *NL-DRM:* Wald-type test based on the normal approximation of $\text{logit}(\hat{\mathcal{G}}_0) - \text{logit}(\hat{\mathcal{G}}_1)$ under the DRM.

Several combinations of $\boldsymbol{\nu}$ are chosen to satisfy the null hypothesis $H_0$ or the alternative hypothesis $H_a$. The details are presented in Table 4.6. Tables 4.7 and 4.8 give the simulated type I error rate and simulated power of each test at the 5% significance level.

From Table 4.7, we observe that the type I error rates for NA-DRM are stable and close to the 5% significance level in all cases. The type I error rates for NL-DRM are similar to those for NA-DRM when the sample sizes are $(300, 300)$ and smaller when $(n_0, n_1) = (100, 100)$. This implies that the logit transformation of the Gini indices is unnecessary for the equality test. The type I error rates for NA-EMP and NL-EMP show similar trends. When the sample sizes are $(100, 100)$ and the proportions of zero values

Table 4.6: Choices of $\boldsymbol{\nu}$ in simulations of testing the equality of the two Gini indices.

| | Null hypothesis $H_0$ | | | | | |
|---|---|---|---|---|---|---|
| | $\chi^2$ | | | $Exp$ | | |
| $\boldsymbol{\nu}$ | (0,0.079) | (0.3,0.355) | (0.7,0.724) | (0,0) | (0.3,0.3) | (0.7,0.7) |
| | Alternative hypothesis $H_a$ | | | | | |
| | $\chi^2$ | | | $Exp$ | | |
| $\boldsymbol{\nu}$ | (0,0) | (0.1,0.3) | (0.4,0.65) | (0.1,0.3) | (0.3,0.45) | (0.5,0.4) |
| $\mathcal{G}_0 - \mathcal{G}_1$ | 0.049 | -0.081 | -0.127 | -0.100 | -0.075 | 0.050 |
| $\text{logit}(\mathcal{G}_0) - \text{logit}(\mathcal{G}_1)$ | 0.206 | -0.323 | -0.633 | -0.418 | -0.350 | 0.251 |

Table 4.7: Type I error rate (%) for testing $H_0 : \mathcal{G}_0 = \mathcal{G}_1$ at the 5% significance level.

| | | $\chi^2$ | | | $Exp$ | | |
|---|---|---|---|---|---|---|---|
| | $\boldsymbol{\nu}$ | (0,0.079) | (0.3,0.355) | (0.7,0.724) | (0,0) | (0.3,0.3) | (0.7,0.7) |
| (100,100) | NA-EMP | 5.50 | 5.00 | 7.10 | 5.55 | 4.85 | 6.90 |
| | NL-EMP | 5.45 | 4.85 | 6.40 | 5.50 | 4.70 | 6.30 |
| | JEL | 4.55 | 4.05 | 3.70 | 4.85 | 3.10 | 2.70 |
| | AJEL | 4.15 | 3.85 | 3.45 | 4.45 | 2.70 | 2.40 |
| | NA-DRM | 4.90 | 5.15 | 5.15 | 5.05 | 4.70 | 5.20 |
| | NL-DRM | 4.85 | 4.80 | 4.75 | 4.95 | 4.65 | 4.95 |
| (300,300) | NA-EMP | 5.10 | 5.70 | 5.70 | 6.15 | 5.35 | 5.55 |
| | NL-EMP | 5.10 | 5.70 | 5.70 | 6.15 | 5.35 | 5.55 |
| | JEL | 4.90 | 5.35 | 5.10 | 5.70 | 4.85 | 4.20 |
| | AJEL | 4.80 | 5.20 | 4.80 | 5.55 | 4.80 | 4.00 |
| | NA-DRM | 5.05 | 4.90 | 4.90 | 5.25 | 5.30 | 5.15 |
| | NL-DRM | 5.05 | 4.85 | 4.95 | 5.25 | 5.25 | 5.05 |

are high, NA-EMP, NL-EMP, JEL, and AJEL have either inflated or conservative type I error rates. Large sample sizes seem to improve their performance.

We observe from Table 4.8 that NA-DRM always gives the largest testing powers. The performance of NL-DRM is comparable to NA-DRM. When the true difference of the Gini indices is large, the testing powers of NA-DRM and NL-DRM are significantly larger than those of the other methods. For example, when the simulated samples are from the $\chi^2$ distributions with $(n_0, n_1) = (100, 100)$ and $\boldsymbol{\nu} = (0, 0)$, the testing powers of NA-DRM and NL-DRM are more than twice the others.

Table 4.8: Simulated testing power (%) of rejecting $H_0 : \mathcal{G}_0 = \mathcal{G}_1$ at the 5% significance level.

| | | $\chi^2$ | | | $Exp$ | | |
|---|---|---|---|---|---|---|---|
| | $\boldsymbol{\nu}$ | (0,0) | (0.1,0.3) | (0.4,0.65) | (0.1,0.3) | (0.3,0.45) | (0.5,0.4) |
| (100,100) | NA-EMP | 30.15 | 43.45 | 78.05 | 59.65 | 38.20 | 18.50 |
| | NL-EMP | 30.00 | 42.95 | 76.95 | 59.10 | 37.40 | 18.15 |
| | JEL | 28.85 | 42.90 | 75.70 | 58.00 | 33.75 | 14.90 |
| | AJEL | 28.00 | 42.05 | 74.85 | 56.95 | 32.75 | 14.00 |
| | NA-DRM | 82.60 | 58.35 | 83.20 | 80.75 | 50.05 | 23.20 |
| | NL-DRM | 82.45 | 58.05 | 82.25 | 80.45 | 49.95 | 22.05 |
| (300,300) | NA-EMP | 67.30 | 85.80 | 99.75 | 97.00 | 79.85 | 45.90 |
| | NL-EMP | 67.30 | 85.70 | 99.70 | 96.90 | 79.50 | 45.50 |
| | JEL | 66.90 | 86.10 | 99.65 | 97.10 | 79.05 | 44.70 |
| | AJEL | 66.50 | 85.85 | 99.65 | 96.85 | 78.60 | 44.05 |
| | NA-DRM | 99.95 | 95.70 | 99.85 | 99.90 | 90.75 | 56.90 |
| | NL-DRM | 99.95 | 95.70 | 99.80 | 99.90 | 90.80 | 55.75 |

## 4.4 Real Data Applications

In this section, we apply our proposed methods to analyze two real datasets. Each dataset can be viewed as consisting of two samples from two different populations, and we are interested in computing the point estimates as well as the construction of 95% CIs for the Gini indices and their difference. The populations for the first dataset contain a large proportions of zeros and the study variables for the second dataset are strictly positive.

The first dataset (Zhou and Cheng, 2008) is from a clinical drug utilization study of patients with uncomplicated hypertension, originally conducted by Murray et al. (2004). It consists of the inpatient charges of 483 patients by gender. We label the charges of the 282 male patients as sample 0 and those of the 201 female patients as sample 1. In most cases, uncomplicated hypertension can be controlled if the patients follow guidelines and take antihypertensive drugs regularly. If they do not need inpatient treatment, the corresponding charges are zero. There are 253 zero values (89.7%) in sample 0 and 171 (85.0%) in sample 1.

To analyze the dataset with our proposed method, we need to choose an appropriate $\boldsymbol{q}(x)$ in the DRM (4.3). The dataset is highly skewed to the right because of the high proportions of zero values and extra skewness in the positive inpatient charges. To balance model fit and model complexity, we choose $\boldsymbol{q}(x) = \log(x)$. The goodness-of-fit test of Qin and Zhang (1997) gives a p-value of 0.563, which indicates that this is a suitable choice. Figure 4.1(a) shows the fitted population distribution functions $\hat{F}_0$ and $\hat{F}_1$ under

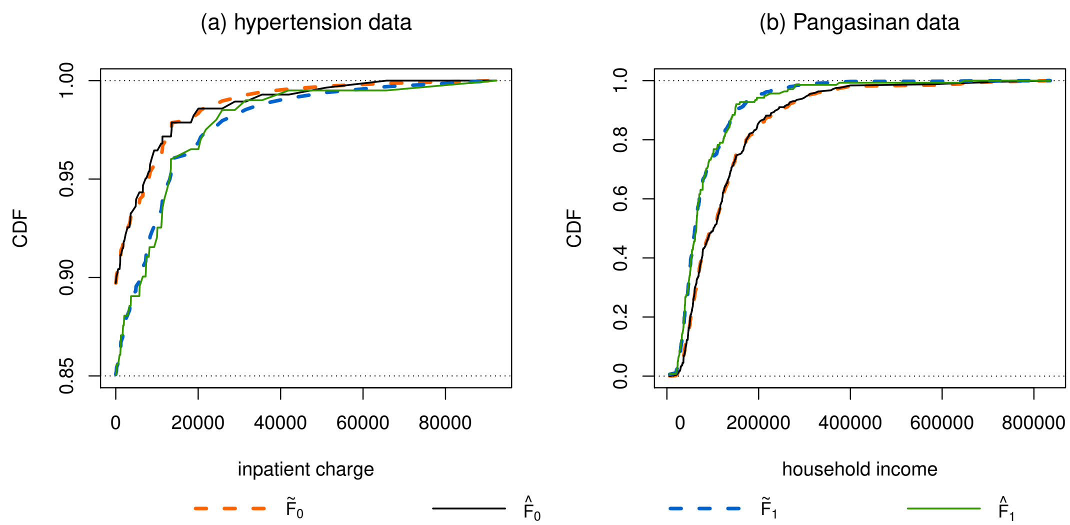


Figure 4.1: Fitted population distributions for real datasets. $\hat{F}_0$ and $\hat{F}_1$: fitted CDFs under the DRM; $\tilde{F}_0$ and $\tilde{F}_1$: empirical CDFs.

the DRM with $\boldsymbol{q}(x) = \log(x)$ together with the empirical CDFs $\tilde{F}_0$ and $\tilde{F}_1$. Clearly, the fit is adequate.

We apply the methods discussed in Sections 4.3.1 and 4.3.2 to this dataset. Table 4.9 presents the point estimates, and Table 4.10 shows the lower bound (LB), upper bound (UB), and length of the 95% CIs. The estimates of $\mathcal{G}_0$, $\mathcal{G}_1$, and $\mathcal{G}_0 - \mathcal{G}_1$ for all three methods are very close. In particular, the EMP and JEL estimates are almost the same. The estimates of $\mathcal{G}_0$ and $\mathcal{G}_1$ are greater than 0.93, indicating the large inequality of the inpatient charge for patients with uncomplicated hypertension; the high proportion of zero values contributes to this. All the methods give similar 95% CIs for $\mathcal{G}_0$. The 95% CIs for $\mathcal{G}_1$ and $\mathcal{G}_0 - \mathcal{G}_1$ for NA-DRM and BT-DRM are the shortest. All the CIs for $\mathcal{G}_0 - \mathcal{G}_1$ contain 0, which suggests no significant difference between the inequality of the inpatient charge for female and male patients at the 95% confidence level.

The second dataset comes from the 1997 Family and Income and Expenditure Survey conducted by the Philippine Statistics Authority; the metadata is available in the R package `ineq`. The province of Pangasinan is located in the Ilocos Region of Luzon. The dataset contains household incomes from different areas of Pangasinan: urban (Sample 0) and rural (Sample 1). Sample 0 has 245 observations and sample 1 has 138 observations. All the

Table 4.9: Point estimates of Gini indices and their difference (hypertension data).

| | $\mathcal{G}_0$ (male) | $\mathcal{G}_1$ (female) | $\mathcal{G}_0 - \mathcal{G}_1$ |
|---|---|---|---|
| EMP | 0.959 | 0.933 | 0.026 |
| JEL | 0.959 | 0.933 | 0.026 |
| DRM | 0.956 | 0.934 | 0.022 |

Table 4.10: 95% CIs for the two Gini indices and their difference (hypertension data).

| | $\mathcal{G}_0$ (male) | | | $\mathcal{G}_1$ (female) | | | $\mathcal{G}_0 - \mathcal{G}_1$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | LB | UB | Length | LB | UB | Length | LB | UB | Length |
| NA-EMP | 0.942 | 0.977 | 0.035 | 0.902 | 0.964 | 0.062 | -0.009 | 0.062 | 0.071 |
| BT-EMP | 0.936 | 0.974 | 0.039 | 0.888 | 0.959 | 0.071 | -0.013 | 0.066 | 0.078 |
| EL | 0.941 | 0.975 | 0.034 | 0.903 | 0.961 | 0.058 | – | – | – |
| BT-EL | 0.938 | 0.976 | 0.038 | 0.897 | 0.966 | 0.068 | – | – | – |
| JEL | 0.942 | 0.980 | 0.038 | 0.904 | 0.967 | 0.063 | -0.017 | 0.069 | 0.086 |
| AJEL | 0.942 | 0.981 | 0.039 | 0.904 | 0.967 | 0.064 | -0.018 | 0.069 | 0.087 |
| NA-DRM | 0.938 | 0.974 | 0.036 | 0.906 | 0.961 | 0.056 | -0.009 | 0.054 | 0.063 |
| BT-DRM | 0.934 | 0.972 | 0.038 | 0.901 | 0.957 | 0.056 | -0.007 | 0.048 | 0.055 |

incomes are positive.

The skewness of the dataset suggests setting $\boldsymbol{q}(x) = \log(x)$ in the DRM (4.3). The goodness-of-fit test of Qin and Zhang (1997) gives a p-value 0.607. Hence, there is no strong evidence to reject the choice of $\boldsymbol{q}(x) = \log(x)$. Figure 4.1(b) also shows that the DRM with $\boldsymbol{q}(x) = \log(x)$ fits the data well.

We use all the methods of Sections 4.3.1 and 4.3.2 to analyze the dataset and summarize the results in Tables 4.11 and 4.12. The EMP and JEL methods give similar estimates of $\mathcal{G}_0$, $\mathcal{G}_1$, and $\mathcal{G}_0 - \mathcal{G}_1$. The DRM estimate of $\mathcal{G}_0$ is comparable to the other estimates, while the DRM estimate of $\mathcal{G}_1$ is smaller than the others. Hence, the DRM estimate of $\mathcal{G}_0 - \mathcal{G}_1$ is larger. All the methods give similar results for the 95% CIs for $\mathcal{G}_0$. The 95% CIs for $\mathcal{G}_1$ and $\mathcal{G}_0 - \mathcal{G}_1$ by NA-DRM and BT-DRM are significantly shorter than the other CIs. This is strong evidence that our method helps to utilize information across the two samples and effectively improves inference when sample sizes are small or moderate. We do not reject the hypothesis that the income inequalities of urban and rural households are the same, since all the 95% CIs for $\mathcal{G}_0 - \mathcal{G}_1$ contain 0.

Table 4.11: Point estimates of Gini indices and their difference (Pangasinan data).

| | $\mathcal{G}_0$ (urban) | $\mathcal{G}_1$ (rural) | $\mathcal{G}_0 - \mathcal{G}_1$ |
|---|---|---|---|
| EMP | 0.393 | 0.394 | -0.001 |
| JEL | 0.391 | 0.389 | 0.002 |
| DRM | 0.399 | 0.371 | 0.028 |

Table 4.12: 95% CIs for the two Gini indices and their difference (Pangasinan data).

| | $\mathcal{G}_0$ (urban) | | | $\mathcal{G}_1$ (rural) | | | $\mathcal{G}_0 - \mathcal{G}_1$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | LB | UB | Length | LB | UB | Length | LB | UB | Length |
| NA-EMP | 0.354 | 0.433 | 0.079 | 0.332 | 0.455 | 0.123 | -0.074 | 0.073 | 0.146 |
| BT-EMP | 0.356 | 0.441 | 0.085 | 0.338 | 0.481 | 0.143 | -0.085 | 0.068 | 0.153 |
| EL | 0.354 | 0.433 | 0.079 | 0.333 | 0.456 | 0.123 | – | – | – |
| BT-EL | 0.353 | 0.434 | 0.080 | 0.335 | 0.455 | 0.120 | – | – | – |
| JEL | 0.356 | 0.436 | 0.081 | 0.339 | 0.466 | 0.127 | -0.083 | 0.070 | 0.153 |
| AJEL | 0.355 | 0.437 | 0.081 | 0.338 | 0.467 | 0.129 | -0.084 | 0.071 | 0.154 |
| NA-DRM | 0.361 | 0.436 | 0.075 | 0.343 | 0.399 | 0.055 | -0.003 | 0.059 | 0.062 |
| BT-DRM | 0.359 | 0.443 | 0.084 | 0.343 | 0.403 | 0.060 | -0.006 | 0.057 | 0.063 |

## 4.5 Technical Details and Additional Simulation Results

### 4.5.1 Proofs

#### Regularity conditions

The asymptotic results in this chapter are developed under some of the following regularity conditions:

C1: The true value $\nu_i^*$ satisfies $0 < \nu_i^* < 1$ for $i = 0, 1$.

C2: As the total sample size $n$ goes to infinity, $n_0/n \to w_0$ for some constant $w_0 \in (0, 1)$.

C3: The two CDFs $G_0$ and $G_1$ satisfy the DRM (4.3) with the true parameter $\boldsymbol{\theta}^*$, and $\int_0^\infty \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} dG_0(x) < \infty$ for all $\boldsymbol{\theta}$ in a neighborhood of the true value $\boldsymbol{\theta}^*$.

C4: The components of $\boldsymbol{Q}(x)$ are continuous and stochastically linearly independent.

C5: The moments $\int_0^\infty x^2 dG_0(x)$ and $\int_0^\infty x^2 \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} dG_0(x)$ exist for all $\boldsymbol{\theta}$ in a neighborhood of the true value $\boldsymbol{\theta}^*$.

Condition C1 ensures that the binomial log-likelihood function $\ell_0(\boldsymbol{\nu})$ has regular properties and the quadratic approximation is applicable. Condition C2 indicates that both $n_0$ and $n_1$ go to infinity at the same rate. For simplicity, and convenience of presentation, we write $w_0 = n_0/n$ and assume that it is a constant. This does not affect our technical development. Conditions C1 and C2 imply that $\boldsymbol{A}_{\boldsymbol{\nu}}$ is positive definite. Condition C3 guarantees the existence of the moment generating function of $\boldsymbol{Q}(X)$ in a neighborhood of $\boldsymbol{\theta}^*$ and therefore all its finite moments. Condition C4 is an identifiability condition. Conditions C3 and C4 together imply that $\boldsymbol{A}_{\boldsymbol{\theta}}$ is positive definite and the quadratic approximation of the dual empirical log-likelihood function $\ell_1(\boldsymbol{\theta})$ is applicable. Conditions C2–C5 guarantee that the linear approximations of $\hat{\mathcal{G}}_0$ and $\hat{\mathcal{G}}_1$ can be used.

**Alternative form of Gini index**

According to David (1968), the Gini's mean difference for sample $i$ can be equivalently expressed by

$$D_i = E|X_{i1} - X_{i2}| = 2\int_{-\infty}^{\infty} \{2xF_i(x) - x\}\, dF_i(x). \tag{4.17}$$

Under model (4.1), $F_i(x) = \nu_i I(x \geq 0) + (1-\nu_i)I(x > 0)G_i(x)$. Then $D_i$ can be further written as

$$\begin{aligned} D_i &= 2(1-\nu_i)\int_0^{\infty} [2x\{\nu_i + (1-\nu_i)G_i(x)\} - x]dG_i(x) \\ &= 2(1-\nu_i)\int_0^{\infty} x\{(2\nu_i - 1) + (1-\nu_i)2G_i(x)\}dG_i(x) \\ &= 2(2\nu_i - 1)\int_0^{\infty} x(1-\nu_i)dG_i(x) + 2(1-\nu_i)^2\int_0^{\infty}\{2xG_i(x)\}dG_i(x) \\ &= 2(2\nu_i - 1)\mu_i + 2(1-\nu_i)^2\int_0^{\infty}\{2xG_i(x)\}dG_i(x). \end{aligned}$$

Recall that $m_i = \int_0^{\infty} x dG_i(x)$ and $\psi_i = \int_0^{\infty}\{2xG_i(x)\}dG_i(x)$. We then have $\mu_i = (1-\nu_i)m_i$ and

$$\mathcal{G}_i = \frac{D_i}{2\mu_i} = (2\nu_i - 1) + (1-\nu_i)\frac{\psi_i}{m_i}.$$

**Proof of Theorem 4.1**

To derive the asymptotic properties, we define an expanded function:

$$\begin{aligned} H(\boldsymbol{\nu},\rho,\boldsymbol{\theta}) &= n_{00}\log(\nu_0)+n_{01}\log(1-\nu_0)+n_{10}\log(\nu_1)+n_{11}\log(1-\nu_1) \\ &\quad -\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\log\left\{1+\rho[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\}-1]\right\}+\sum_{j=1}^{n_{11}}\{\boldsymbol{\beta}^\top \boldsymbol{q}(X_{1j})\}. \end{aligned} \tag{4.18}$$

Since the MELEs are obtained by $\hat{\nu} = \arg\max_{\boldsymbol{\nu}} \ell_0(\boldsymbol{\nu})$ and $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell_1(\boldsymbol{\theta})$, we have

$$\frac{\partial H(\hat{\boldsymbol{\nu}},\hat{\rho},\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\nu}} = \mathbf{0} \text{ and } \frac{\partial H(\hat{\boldsymbol{\nu}},\hat{\rho},\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \mathbf{0}. \tag{4.19}$$

Note that

$$\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\frac{1}{n_{01}+n_{11}}\frac{1}{1+\hat{\rho}[\exp\{\hat{\alpha}+\hat{\boldsymbol{\beta}}^\top \boldsymbol{q}(X_{ij})\}-1]}=1,$$

which ensures that the MELE of $G_0(x)$ is a CDF. From this, we can verify that

$$\frac{\partial H(\hat{\boldsymbol{\nu}},\hat{\rho},\hat{\boldsymbol{\theta}})}{\partial \rho} = 0. \tag{4.20}$$

Then (4.19) and (4.20) together imply that $\hat{\boldsymbol{\eta}}$ satisfies

$$\frac{\partial H(\hat{\boldsymbol{\eta}})}{\partial \boldsymbol{\eta}} = \mathbf{0}, \tag{4.21}$$

which serves as the starting point of our proof for $\hat{\boldsymbol{\eta}}$.

Next, we apply the first-order Taylor expansion to $\partial H(\hat{\boldsymbol{\eta}})/\partial\boldsymbol{\eta}$ to find an approximation for $\hat{\boldsymbol{\eta}}$. In this process, the first and second derivatives of $H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})$ play important roles. Their detailed forms are given below.

- *First derivatives of* $H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})$

  After some calculation, we find the first derivatives of $H(\boldsymbol{\nu},\boldsymbol{\theta},\rho)$ as follows:

$$\frac{\partial H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})}{\partial \boldsymbol{\nu}} = \left(\frac{\partial H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})}{\partial \nu_0}, \frac{\partial H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})}{\partial \nu_1}\right)^\top = \left(\frac{n_{00}}{\nu_0}-\frac{n_{01}}{1-\nu_0}, \frac{n_{10}}{\nu_1}-\frac{n_{11}}{1-\nu_1}\right)^\top,$$

$$
\begin{aligned}
\frac{\partial H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})}{\partial \rho} &= -\sum_{ij} \frac{\omega(X_{ij};\boldsymbol{\theta})-1}{1+\rho\{\omega(X_{ij};\boldsymbol{\theta})-1\}} I(X_{ij}>0),\\
\frac{\partial H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \sum_{j=1}^{n_1} \boldsymbol{Q}(X_{1j}) I(X_{1j}>0) - \sum_{ij} \frac{\rho\omega(X_{ij};\boldsymbol{\theta})}{1+\rho\{\omega(X_{ij};\boldsymbol{\theta})-1\}} \boldsymbol{Q}(X_{ij}) I(X_{ij}>0),
\end{aligned}
$$

where $\sum_{ij}$ refers to summation over the full range of data.

We evaluate the above derivatives at $\boldsymbol{\eta}^*$ and define

$$
\boldsymbol{S}_n = \frac{\partial H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}} = \begin{pmatrix} \frac{\partial H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu}} \\ \frac{\partial H(\boldsymbol{\eta}^*)}{\partial \rho} \\ \frac{\partial H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{n,\boldsymbol{\nu}} \\ S_{n,\rho} \\ \boldsymbol{S}_{n,\boldsymbol{\theta}} \end{pmatrix}, \tag{4.22}
$$

where the corresponding entries are

$$
\begin{aligned}
\boldsymbol{S}_{n,\boldsymbol{\nu}} &= \left( \frac{n_{00}}{\nu_0^*} - \frac{n_{01}}{1-\nu_0^*}, \frac{n_{10}}{\nu_1^*} - \frac{n_{11}}{1-\nu_1^*} \right)^\top,\\
S_{n,\rho} &= -\sum_{ij} \frac{\omega(X_{ij})-1}{h(X_{ij})} I(X_{ij}>0),\\
\boldsymbol{S}_{n,\boldsymbol{\theta}} &= \sum_{j=1}^{n_1} \boldsymbol{Q}(X_{1j}) I(X_{1j}>0) - \sum_{ij} h_1(X_{ij}) \boldsymbol{Q}(X_{ij}) I(X_{ij}>0).
\end{aligned}
$$

- *Second derivatives of* $H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})$

We next calculate the second derivatives of $H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})$ and evaluate them at $\boldsymbol{\eta}^*$. This leads to

$$
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} = \begin{pmatrix} \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^\top} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu} \partial \rho} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu} \partial \boldsymbol{\theta}^\top} \\ \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \rho \partial \boldsymbol{\nu}^\top} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \rho^2} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \rho \partial \boldsymbol{\theta}^\top} \\ \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\nu}^\top} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta} \partial \rho} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \end{pmatrix}, \tag{4.23}
$$

where $h_0(x) = (1-\rho^*)/h(x) = 1 - h_1(x)$ and

$$
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^\top} = \operatorname{diag}\left\{ -\frac{n_{00}}{\nu_0^{*2}} - \frac{n_{01}}{(1-\nu_0^*)^2}, -\frac{n_{10}}{\nu_1^{*2}} - \frac{n_{11}}{(1-\nu_1^*)^2} \right\},
$$

$$\begin{aligned}
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \rho^2} &= -\sum_{ij} \frac{-\{\omega(X_{ij})-1\}^2}{h(X_{ij})^2} I(X_{ij}>0),\\
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu} \partial \rho} &= \left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \rho \partial \boldsymbol{\nu}^\top}\right\}^\top = \mathbf{0},\\
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta} \partial \rho} &= \left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \rho \partial \boldsymbol{\theta}^\top}\right\}^\top = -\sum_{ij} \frac{\omega(X_{ij})}{h(X_{ij})^2} \boldsymbol{Q}(X_{ij}) I(X_{ij}>0),\\
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= -\sum_{ij} h_0(X_{ij}) h_1(X_{ij}) \{\boldsymbol{Q}(X_{ij}) \boldsymbol{Q}(X_{ij})^\top\} I(X_{ij}>0),\\
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu} \partial \boldsymbol{\theta}^\top} &= \left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\nu}^\top}\right\}^\top = \mathbf{0}.
\end{aligned}$$

- *Some useful lemmas*

In the proof of Theorem 4.1, we need the expectation of $\partial^2 H(\boldsymbol{\eta}^*)/(\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top)$ and the asymptotic property of $\boldsymbol{S}_n$. The following lemma is used to ease the calculation burden in our main proofs.

**Lemma 4.1.** *Suppose that $f$ is an arbitrary vector-valued function. Let $E_0(\cdot)$ represent the expectation with respect to $G_0$ and $X$ refer to a random variable from $G_0$. Then*

$$E\left\{\sum_{ij} f(X_{ij}) I(X_{ij}>0)\right\} = n\Delta^* E_0\{h(X) f(X)\}.$$

*Proof.* Note that

$$\begin{aligned}
E\left\{\sum_{ij} f(X_{ij}) I(X_{ij}>0)\right\} &= \sum_{i=0}^{1} n_i E\{f(X_{i1}) I(X_{i1}>0)\}\\
&= n_0(1-\nu_0^*) E_0\{f(X)\} + n_1(1-\nu_1^*) E_0\{\omega(X) f(X)\},
\end{aligned}$$

where we use the DRM (4.3) in the last step. Using the facts that $w_i = n_i/n$, we further have

$$E\left\{\sum_{ij} f(X_{ij}) I(X_{ij}>0)\right\} = n w_0(1-\nu_0^*) E_0\{f(X)\} + n w_1(1-\nu_1^*) E_0\{\omega(X) f(X)\}.$$

Recall the definitions of $\Delta^*$ and $\rho^*$. We then have

$$\begin{aligned} E\left\{\sum_{ij} f(X_{ij}) I(X_{ij} > 0)\right\} &= n\Delta^* E_0\{(1-\rho^*)f(X)\} + n\Delta^* E_0[\rho^*\omega(X)f(X)] \\ &= n\Delta^* E_0\{h(X)f(X)\}. \end{aligned}$$

This completes the proof. □

With the help of Lemma 4.1, we calculate the expectation of $\partial^2 H(\boldsymbol{\eta}^*)/(\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top)$.

**Lemma 4.2.** *With the form of $\partial^2 H(\boldsymbol{\eta}^*)/(\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top)$ given in (4.23), we have*

$$-\frac{1}{n}E\left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top}\right\} = \boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\nu}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -A_\rho & \boldsymbol{A}_{\rho,\boldsymbol{\theta}} \\ \boldsymbol{0} & \boldsymbol{A}_{\boldsymbol{\theta},\rho} & \boldsymbol{A}_{\boldsymbol{\theta}} \end{pmatrix},$$

*where*

$$\begin{aligned} \boldsymbol{A}_{\boldsymbol{\nu}} &= \mathit{diag}\left\{\frac{w_0}{\nu_0^*(1-\nu_0^*)}, \frac{w_1}{\nu_1^*(1-\nu_1^*)}\right\},\ \boldsymbol{A}_{\boldsymbol{\theta}} = \Delta^*(1-\rho^*)E_0\left[h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}^\top(X)\right], \\ A_\rho &= \Delta^* E_0\left\{\frac{\{\omega(X)-1\}^2}{h(X)}\right\} = \{\rho^*(1-\rho^*)\}^{-1}\left[\Delta^* - \{\rho^*(1-\rho^*)\}^{-1}\boldsymbol{e}^\top\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e}\right], \\ \boldsymbol{A}_{\boldsymbol{\theta},\rho} &= \boldsymbol{A}_{\rho,\boldsymbol{\theta}}^\top = \Delta^* E_0\left\{\frac{\omega(X)}{h(X)}\boldsymbol{Q}(X)\right\} = \{\rho^*(1-\rho^*)\}^{-1}\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e} \end{aligned}$$

*with $\boldsymbol{e} = (1, \boldsymbol{0}_{d\times 1}^\top)^\top$.*

*Proof.* Note that $n_{00} \sim \mathrm{Bin}(n_0, \nu_0)$ and $n_{10} \sim \mathrm{Bin}(n_1, \nu_1)$, where "Bin" denotes the binomial distribution. Since $w_i = n_i/n$, we can easily show that

$$-\frac{1}{n}E\left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\nu\partial\nu^\top}\right\} = \boldsymbol{A}_{\boldsymbol{\nu}}.$$

Next, we apply Lemma 4.1 to find the remaining entries of $E\left\{\partial^2 H(\boldsymbol{\eta}^*)/(\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top)\right\}$. We use

$$E\left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\}$$

as an illustration. For the other entries, the idea is similar and we omit the details.

Note that

$$
\begin{aligned}
-\frac{1}{n}E\left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\} &= \frac{1}{n}E\left\{\sum_{ij} h_0(X_{ij})h_1(X_{ij})\boldsymbol{Q}(X_{ij})\boldsymbol{Q}(X_{ij})^\top I(X_{ij}>0)\right\} \\
&= \Delta^* E_0\left\{h(X)h_0(X)h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}(X)^\top\right\} \\
&= \Delta^*(1-\rho^*)E_0\left\{h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}(X)^\top\right\},
\end{aligned}
$$

where we have used Lemma 4.1 in the second step and the fact that $h(x)h_0(x) = 1-\rho^*$ in the third step. This completes the proof. □

We now study the asymptotic properties of $\boldsymbol{S}_n$ defined in (4.22). Recall that $\boldsymbol{W} = ((1-\nu_0^*)^{-1}, -(1-\nu_1^*)^{-1})$ and define $S = w_0^{-1} + w_1^{-1}$.

**Lemma 4.3.** *With the form of $\boldsymbol{S}_n$ in (4.22), as $n \to \infty$*

$$
n^{-1/2}\boldsymbol{S}_n \to N(\boldsymbol{0}, \boldsymbol{B}),
$$

*in distribution, where*

$$
\begin{aligned}
\boldsymbol{B} &= \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\nu}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & A_\rho & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\boldsymbol{\theta}} \end{pmatrix} \\
&\quad + \begin{pmatrix} \boldsymbol{0} & -\rho^*(1-\rho^*)A_\rho\boldsymbol{W}^\top & \boldsymbol{W}^\top\boldsymbol{e}^\top\boldsymbol{A}_{\boldsymbol{\theta}} \\ -\rho^*(1-\rho^*)A_\rho\boldsymbol{W} & -S\{\rho^*(1-\rho^*)\}^2A_\rho^2 & S\rho^*(1-\rho^*)A_\rho\boldsymbol{e}^\top\boldsymbol{A}_{\boldsymbol{\theta}} \\ \boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e}\boldsymbol{W} & S\rho^*(1-\rho^*)A_\rho\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e} & -S\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e}(\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e})^\top \end{pmatrix}.
\end{aligned}
$$

*Proof.* Using the results in Lemma 4.1, it is easy to show that $E(\boldsymbol{S}_n) = \boldsymbol{0}$; we omit the details.

Next, we verify that $Var(\boldsymbol{S}_n) = \boldsymbol{B}$. For convenience, we write $\boldsymbol{B}$ as

$$
\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \boldsymbol{B}_{13} \\ \boldsymbol{B}_{21} & B_{22} & \boldsymbol{B}_{23} \\ \boldsymbol{B}_{31} & \boldsymbol{B}_{32} & \boldsymbol{B}_{33} \end{pmatrix}.
$$

We concentrate on deriving $\boldsymbol{B}_{13}$; the other entries can be similarly obtained and we omit the details.

Note that $\boldsymbol{S}_{n,\boldsymbol{\nu}}$ and $\boldsymbol{S}_{n,\boldsymbol{\theta}}$ can be rewritten as

$$\begin{aligned}
\boldsymbol{S}_{n,\nu_0} &= \frac{n_{00}}{\nu_0^*} - \frac{n_{01}}{1-\nu_0^*} = -\frac{n_{01}}{\nu_0^*(1-\nu_0^*)} = -\frac{1}{\nu_0^*(1-\nu_0^*)}\sum_{j=1}^{n_0} I(X_{0j}>0),\\
\boldsymbol{S}_{n,\nu_1} &= \frac{n_{10}}{\nu_1^*} - \frac{n_{11}}{1-\nu_1^*} = -\frac{n_{11}}{\nu_1^*(1-\nu_1^*)} = -\frac{1}{\nu_1^*(1-\nu_1^*)}\sum_{j=1}^{n_1} I(X_{1j}>0),\\
\boldsymbol{S}_{n,\boldsymbol{\theta}} &= \sum_{j=1}^{n_1}\boldsymbol{Q}(X_{1j})I(X_{1j}>0) - \sum_{ij} h_1(X_{ij})\boldsymbol{Q}(X_{ij})I(X_{ij}>0)\\
&= \sum_{j=1}^{n_1} h_0(X_{1j})\boldsymbol{Q}(X_{1j})I(X_{1j}>0) - \sum_{j=1}^{n_0} h_1(X_{0j})\boldsymbol{Q}(X_{0j})I(X_{0j}>0).
\end{aligned}$$

Then we have

$$\begin{aligned}
&\frac{1}{n}Cov(\boldsymbol{S}_{n,\nu_0}, \boldsymbol{S}_{n,\boldsymbol{\theta}}^\top)\\
=\ & \frac{1}{n\nu_0^*(1-\nu_0^*)}Cov\left\{\sum_{j=1}^{n_0} I(X_{0j}>0), \sum_{j=1}^{n_0} h_1(X_{0j})\boldsymbol{Q}(X_{0j})^\top I(X_{0j}>0)\right\}\\
=\ & \frac{n_0}{n\nu_0^*(1-\nu_0^*)}\left[(1-\nu_0^*)E_0\left\{h_1(X)\boldsymbol{Q}(X)^\top\right\} - (1-\nu_0^*)^2E_0\left\{h_1(X)\boldsymbol{Q}(X)^\top\right\}\right]\\
=\ & w_0E_0\left\{h_1(X)\boldsymbol{Q}(X)^\top\right\}\\
=\ & (1-\nu_0^*)^{-1}(\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e})^\top.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\frac{1}{n}Cov(\boldsymbol{S}_{n,\nu_1}, \boldsymbol{S}_{n,\boldsymbol{\theta}}^\top)\\
=\ & \frac{-1}{n\nu_1^*(1-\nu_1^*)}Cov\left\{\sum_{j=1}^{n_1} I(X_{1j}>0), \sum_{j=1}^{n_1} h_0(X_{1j})\boldsymbol{Q}(X_{1j})^\top I(X_{1j}>0)\right\}\\
=\ & \frac{-n_1}{n\nu_1^*(1-\nu_1^*)}\left[(1-\nu_1^*)E_0\left\{h_0(X)\omega(X)\boldsymbol{Q}(X)^\top\right\} - (1-\nu_1^*)^2E_0\left\{h_0(X)\omega(X)\boldsymbol{Q}(X)^\top\right\}\right]\\
=\ & -w_1\cdot\frac{1-\rho^*}{\rho^*}E_0\left\{h_1(X)\boldsymbol{Q}(X)^\top\right\}\\
=\ & -(1-\nu_1^*)^{-1}(\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e})^\top.
\end{aligned}$$

Recall that $\boldsymbol{W} = ((1-\nu_0^*)^{-1}, -(1-\nu_1^*)^{-1})$. Then $\boldsymbol{B}_{13} = \boldsymbol{W}^\top \boldsymbol{e}^\top \boldsymbol{A}_{\boldsymbol{\theta}}$.

Note that $\boldsymbol{S}_n$ in (4.22) is a sum of independent random vectors. Therefore, by the classical central limit theorem, we have as $n \to \infty$

$$n^{-1/2}\boldsymbol{S}_n \to N(\mathbf{0}, \boldsymbol{B}),$$

in distribution, which completes the proof. □

**Proof of Theorem 4.1**

With the above preparation, we now move to the asymptotic property of $\hat{\boldsymbol{\eta}}$.

Recall that $\hat{\boldsymbol{\eta}}$ satisfies

$$\frac{\partial H(\hat{\boldsymbol{\eta}})}{\partial \boldsymbol{\eta}} = \mathbf{0}.$$

Applying the first-order Taylor expansion to $\partial H(\hat{\boldsymbol{\eta}})/\partial \boldsymbol{\eta}$, and using (4.22) and Lemma 4.2, we have

$$\begin{aligned}
\mathbf{0} &= \frac{\partial H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}} + \left(\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top}\right)(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + o_p(n^{1/2}) \\
&= \boldsymbol{S}_n - n\boldsymbol{A}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + o_p(n^{1/2}).
\end{aligned}$$

Conditions C1–C4 ensure that the matrix $\boldsymbol{A}$ is positive definite. Hence, we obtain an approximation for $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*$ as

$$\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = \begin{pmatrix} \hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^* \\ \hat{\rho} - \rho^* \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \end{pmatrix} = \frac{1}{n}\boldsymbol{A}^{-1}\boldsymbol{S}_n + o_p(n^{-1/2}). \tag{4.24}$$

This together with the asymptotic property of $\boldsymbol{S}_n$ in Lemma 4.3 and Slutsky's theorem gives

$$n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \to N(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}),$$

in distribution, as $n \to \infty$.

To find the explicit form of $\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}$, we first identify the structure of $\boldsymbol{A}^{-1}$. We write

$$\begin{pmatrix} -A_\rho & \boldsymbol{A}_{\rho,\boldsymbol{\theta}} \\ \boldsymbol{A}_{\boldsymbol{\theta},\rho} & \boldsymbol{A}_{\boldsymbol{\theta}} \end{pmatrix}^{-1} = \begin{pmatrix} A^{11} & \boldsymbol{A}^{12} \\ \boldsymbol{A}^{21} & \boldsymbol{A}^{22} \end{pmatrix}.$$

Using the formula for the inverse of a $2 \times 2$ block matrix, we have

$$\begin{aligned}
A^{11} &= \{-A_\rho - (\boldsymbol{A}_{\rho,\boldsymbol{\theta}})\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{A}_{\boldsymbol{\theta},\rho})\}^{-1} \\
&= \left[\{\rho^*(1-\rho^*)\}^{-2}\boldsymbol{e}^\top \boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e} - \Delta^*\{\rho^*(1-\rho^*)\}^{-1} - \{\rho^*(1-\rho^*)\}^{-2}\boldsymbol{e}^\top \boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e}\right]^{-1} \\
&= -\frac{\rho^*(1-\rho^*)}{\Delta}, \\
\boldsymbol{A}^{12} &= (\boldsymbol{A}^{21})^\top = -A^{11}(\boldsymbol{A}_{\rho,\boldsymbol{\theta}})\boldsymbol{A}_{\boldsymbol{\theta}}^{-1} = \frac{\boldsymbol{e}^\top}{\Delta^*}, \\
\boldsymbol{A}^{22} &= \boldsymbol{A}_{\boldsymbol{\theta}}^{-1} + \boldsymbol{A}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{A}_{\boldsymbol{\theta},\rho})A^{11}(\boldsymbol{A}_{\rho,\boldsymbol{\theta}})\boldsymbol{A}_{\boldsymbol{\theta}}^{-1} = \boldsymbol{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)}.
\end{aligned}$$

Hence, $\boldsymbol{A}^{-1}$ is given by

$$\boldsymbol{A}^{-1} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\nu}}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\frac{\rho^*(1-\rho^*)}{\Delta^*} & \frac{\boldsymbol{e}^\top}{\Delta^*} \\ \mathbf{0} & \frac{\boldsymbol{e}}{\Delta^*} & \boldsymbol{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)} \end{pmatrix}. \tag{4.25}$$

With the form of $\boldsymbol{A}^{-1}$ in (4.25) and the form of $\boldsymbol{B}$ in Lemma 4.3, after some tedious algebra, we find that

$$\begin{aligned}
\boldsymbol{\Lambda} &= \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} \\
&= \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\nu}}^{-1} & \rho^*(1-\rho^*)\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^\top & \mathbf{0} \\ \rho^*(1-\rho^*)\boldsymbol{W}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1} & \rho^*(1-\rho^*)\{\frac{1}{\Delta^*} - S\rho^*(1-\rho^*)\} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)} \end{pmatrix}.
\end{aligned}$$

Recall that $S = w_0^{-1} + w_1^{-1}$. Some algebra leads to

$$\frac{1}{\Delta^*} - S\rho^*(1-\rho^*) = \frac{1}{\Delta^*}\{\rho^*\nu_0^* + (1-\rho^*)\nu_1^*\}.$$

This completes the proof of Theorem 4.1.

**Proof of Theorem 4.2**

Recall that

$$\hat{p}_{ij} = \frac{1}{n_{01}+n_{11}}\left\{1+\hat{\rho}[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}-1]\right\}^{-1}$$
$$= \frac{1}{nw_0(1-\hat{\nu}_0)+nw_1(1-\hat{\nu}_1)}\left\{1+\hat{\rho}[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}-1]\right\}^{-1}. \quad (4.26)$$

The MELE of $\boldsymbol{\gamma}$ is then given by

$$\hat{\boldsymbol{\gamma}} = \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \hat{p}_{ij}\boldsymbol{u}(X_{ij};\hat{\boldsymbol{\nu}},\hat{\boldsymbol{\theta}})$$
$$= \frac{1}{nw(1-\hat{\nu}_0)+nw_1(1-\hat{\nu}_1)}\sum_{ij}\frac{\boldsymbol{u}(X_{ij};\hat{\boldsymbol{\nu}},\hat{\boldsymbol{\theta}})}{1+\hat{\rho}[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\}-1]}I(X_{ij}>0).$$

Note that $\hat{\boldsymbol{\gamma}}$ is a function of $\hat{\boldsymbol{\eta}}$, so we define

$$\boldsymbol{\gamma}(\boldsymbol{\eta}) = \frac{1}{nw(1-\nu_0)+nw_1(1-\nu_1)}\sum_{ij}\frac{\boldsymbol{u}(X_{ij};\boldsymbol{\nu},\boldsymbol{\theta})}{1+\rho[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\}-1]}I(X_{ij}>0).$$

We then have $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}(\hat{\boldsymbol{\eta}})$. From Theorem 4.1, we have $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}^* + O_p(n^{-1/2})$. Applying the first-order Taylor expansion to $\boldsymbol{\gamma}(\hat{\boldsymbol{\eta}})$, we get

$$\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}(\boldsymbol{\eta}^*) + \left(\frac{\partial \boldsymbol{\gamma}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right)(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}^*) + o_p(n^{-1/2}). \quad (4.27)$$

For convenience, we write $\boldsymbol{u}(x) = \boldsymbol{u}(x;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)$. Note that

$$\frac{\partial \boldsymbol{\gamma}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}} = \left(\frac{\partial \boldsymbol{\gamma}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu}}, \frac{\partial \boldsymbol{\gamma}(\boldsymbol{\eta}^*)}{\partial \rho}, \frac{\partial \boldsymbol{\gamma}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta}}\right)$$

where

$$\frac{\partial \boldsymbol{\gamma}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu}} = \frac{1}{n\Delta^{*2}}\sum_{ij}\left\{\frac{\partial \boldsymbol{u}(X_{ij};\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)/\partial\boldsymbol{\nu}}{h(X_{ij})}\Delta^* + (w_0,w_1)\otimes\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}\right\}I(X_{ij}>0),$$
$$\frac{\partial \boldsymbol{\gamma}(\boldsymbol{\eta}^*)}{\partial \rho} = -\frac{1}{n\Delta^*}\sum_{ij}\frac{\boldsymbol{u}(X_{ij})\{\omega(X_{ij})-1\}}{h(X_{ij})^2}I(X_{ij}>0),$$

$$\frac{\partial \boldsymbol{\gamma}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta}} = \frac{1}{n\Delta^*}\sum_{ij}\frac{\{\partial \boldsymbol{u}(X_{ij};\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}\}\cdot h(X_{ij}) - \boldsymbol{u}(X_{ij})\rho^*\omega(X_{ij})\boldsymbol{Q}(X)^\top}{h(X_{ij})^2} I(X_{ij}>0),$$

and $\otimes$ indicates the Kronecker product. By the law of large numbers and Lemma 4.1, we have

$$\frac{\partial \boldsymbol{\gamma}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}} \to \boldsymbol{C}$$

in probability, as $n\to\infty$, where $\boldsymbol{C} = (\boldsymbol{C}_{\boldsymbol{\nu}}, \boldsymbol{C}_\rho, \boldsymbol{C}_{\boldsymbol{\theta}})$ with

$$\begin{aligned}
\boldsymbol{C}_{\boldsymbol{\nu}} &= E_0\left\{\frac{\partial \boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)}{\partial\boldsymbol{\nu}}\right\} + (w_0,w_1)\otimes\frac{\boldsymbol{\gamma}^*}{\Delta^*},\\
\boldsymbol{C}_\rho &= -E_0\left\{\frac{\boldsymbol{u}(X)\{\omega(X)-1\}}{h(X)}\right\} = \frac{\rho^*\boldsymbol{\gamma}^* - E_0\{h_1(X)\boldsymbol{u}(X)\}}{\rho^*(1-\rho^*)},\\
\boldsymbol{C}_{\boldsymbol{\theta}} &= E_0\left[\frac{\{\partial \boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}\}\cdot h(X) - \boldsymbol{u}(X)\rho^*\omega(X)\boldsymbol{Q}(X)^\top}{h(X)}\right]\\
&= E_0\left\{\frac{\partial \boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}}\right\} - E_0\left\{h_1(X)\boldsymbol{u}(X)\boldsymbol{Q}(X)^\top\right\}.
\end{aligned}$$

For convenience, we let

$$\mathbf{E}_{0\boldsymbol{u}} = E_0\{h_0(X)\boldsymbol{u}(X)\} \quad \text{and} \quad \mathbf{E}_{1\boldsymbol{u}} = E_0\{h_1(X)\boldsymbol{u}(X)\}.$$

Then $\mathbf{E}_{0\boldsymbol{u}} + \mathbf{E}_{1\boldsymbol{u}} = \boldsymbol{\gamma}^*$ and

$$\boldsymbol{C}_\rho = \frac{\rho^*\boldsymbol{\gamma}^* - \mathbf{E}_{1\boldsymbol{u}}}{\rho^*(1-\rho^*)}.$$

Note that the first term of (4.27) involves

$$\hat{\boldsymbol{\gamma}}(\boldsymbol{\eta}^*) = \frac{1}{n\Delta^*}\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0).$$

Then Equation (4.27) can be written as

$$\hat{\boldsymbol{\gamma}} = \frac{1}{n\Delta^*}\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0) + \boldsymbol{C}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + o_p(n^{-1/2}). \tag{4.28}$$

Recall from (4.24) that $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = n^{-1}\boldsymbol{A}^{-1}\boldsymbol{S}_n + o_p(n^{-1/2})$. Therefore, as $n \to \infty$, $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*)$ has the same limiting distribution as

$$n^{1/2}\left[\left\{\frac{1}{n\Delta^*}\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0) - \boldsymbol{\gamma}^*\right\} + \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{S}_n/n\right]. \tag{4.29}$$

It can easily be verified that (4.29) has expectation zero. We will now decompose its asymptotic variance into three parts.

The variance of the first term in (4.29) is

$$\begin{aligned}
\boldsymbol{\Gamma}_1 &= Var\{\boldsymbol{\gamma}(\boldsymbol{\eta}^*)\} \\
&= \frac{1}{\Delta^*}E_0\left\{\frac{\boldsymbol{u}(X)\boldsymbol{u}(X)^\top}{h(X)}\right\} - \frac{1}{w_0}E_0\{h_0(X)\boldsymbol{u}(X)\}E_0\{h_0(X)\boldsymbol{u}(X)^\top\} \\
&\quad -\frac{1}{w_1}E_0\{h_1(X)\boldsymbol{u}(X)\}E_0\{h_1(X)\boldsymbol{u}(X)^\top\} \\
&= \frac{1}{\Delta^*}E_0\left\{\frac{\boldsymbol{u}(X)\boldsymbol{u}(X)^\top}{h(X)}\right\} - \frac{1}{w_0}\mathbf{E}_{0\boldsymbol{u}}\mathbf{E}_{0\boldsymbol{u}}^\top - \frac{1}{w_1}\mathbf{E}_{1\boldsymbol{u}}\mathbf{E}_{1\boldsymbol{u}}^\top,
\end{aligned} \tag{4.30}$$

where in the first step we have used the results in Lemma 4.1, and in the second step we have used the definitions of $\mathbf{E}_{0\boldsymbol{u}}$ and $\mathbf{E}_{1\boldsymbol{u}}$.

Next, we derive the variance of the second term in (4.29):

$$\boldsymbol{\Gamma}_2 = nVar(\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{S}_n/n) = \boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}^\top.$$

Together with the form of $\boldsymbol{\Lambda}$ in Theorem 1, we have

$$\begin{aligned}
\boldsymbol{\Gamma}_2 &= \boldsymbol{C}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^\top + \rho^*(1-\rho^*)\boldsymbol{C}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^\top\boldsymbol{C}_\rho^\top + \rho^*(1-\rho^*)\boldsymbol{C}_\rho\boldsymbol{W}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^\top \\
&\quad +(\Delta^*)^{-1}\rho^*(1-\rho^*)\{\rho^*\nu_0^* + (1-\rho^*)\nu_1^*\}\boldsymbol{C}_\rho\boldsymbol{C}_\rho^\top + \boldsymbol{C}_{\boldsymbol{\theta}}\left\{\boldsymbol{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)}\right\}\boldsymbol{C}_{\boldsymbol{\theta}}^\top.
\end{aligned}$$

Note that

$$\boldsymbol{W}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^\top = \frac{\rho^*\nu_0^* + (1-\rho^*)\nu_1^*}{\Delta^*\rho^*(1-\rho^*)}.$$

Then

$$\begin{aligned}\boldsymbol{\Gamma}_2 &= \{\boldsymbol{C}_{\boldsymbol{\nu}} + \rho^*(1-\rho^*)\boldsymbol{C}_\rho \boldsymbol{W}\}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\{\boldsymbol{C}_{\boldsymbol{\nu}} + \rho^*(1-\rho^*)\boldsymbol{C}_\rho \boldsymbol{W}\}^\top \\ &\quad - \frac{1}{\Delta^*\rho^*(1-\rho^*)}(\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{e})(\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{e})^\top + \boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{C}_{\boldsymbol{\theta}}^\top. \end{aligned} \tag{4.31}$$

Lastly, we derive the covariance of the two terms in (4.29). That is,

$$\boldsymbol{\Gamma}_3 = nCov[\boldsymbol{\gamma}(\boldsymbol{\eta}^*), n^{-1}\{\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{S}_n\}^\top] = Cov\{\boldsymbol{\gamma}(\boldsymbol{\eta}^*), \boldsymbol{S}_n^\top\}\boldsymbol{A}^{-1}\boldsymbol{C}^\top.$$

For convenience, we write $Cov\{\boldsymbol{\gamma}(\boldsymbol{\eta}^*), \boldsymbol{S}_n^\top\} = (\boldsymbol{D}_{\boldsymbol{\nu}}, \boldsymbol{D}_\rho, \boldsymbol{D}_{\boldsymbol{\theta}})$.

We first look at

$$\begin{aligned} & Cov\{\boldsymbol{\gamma}(\boldsymbol{\eta}^*), \boldsymbol{S}_{n,\nu_1}\} \\ = \; & Cov\left\{\frac{1}{n\Delta^*}\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0), \frac{-n_{11}}{\nu_1^*(1-\nu_1^*)}\right\} \\ = \; & \frac{-1}{n\Delta^*\nu_1^*(1-\nu_1^*)}Cov\left\{\sum_{j=1}^{n_1}\frac{\boldsymbol{u}(X_{1j})}{h(X_{1j})}I(X_{1j}>0), \sum_{j=1}^{n_1}I(X_{1j}>0)\right\} \\ = \; & \frac{-n_1}{n\Delta^*\nu_1^*(1-\nu_1^*)}\left[(1-\nu_1^*)E_0\left\{\frac{\boldsymbol{u}(X)\omega(X)}{h(X)}\right\} - (1-\nu_1^*)^2E_0\left\{\frac{\boldsymbol{u}(X)\omega(X)}{h(X)}\right\}\right] \\ = \; & \frac{-w_1}{\Delta^*}E_0\left\{\frac{\boldsymbol{u}(X)\omega(X)}{h(X)}\right\} \\ = \; & -(1-\nu_1^*)^{-1}\mathbf{E}_{1\boldsymbol{u}}. \end{aligned}$$

Similarly, we find

$$Cov\{\boldsymbol{\gamma}(\boldsymbol{\eta}^*), \boldsymbol{S}_{n,\nu_0}\} = -(1-\nu_0^*)^{-1}\mathbf{E}_{0\boldsymbol{u}}.$$

Hence,

$$\boldsymbol{D}_{\boldsymbol{\nu}} = \left(-(1-\nu_0^*)^{-1}\mathbf{E}_{0\boldsymbol{u}}, -(1-\nu_1^*)^{-1}\mathbf{E}_{1\boldsymbol{u}}\right).$$

We can find $\boldsymbol{D}_\rho$ and $\boldsymbol{D}_{\boldsymbol{\theta}}$ in a similar manner. For $\boldsymbol{D}_\rho$,

$$\begin{aligned}
\boldsymbol{D}_\rho &= Cov\{\boldsymbol{\gamma}(\boldsymbol{\eta}^*), S_{n,\rho}\} \\
&= -\frac{1}{n\Delta^*}Cov\left\{\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0), \sum_{ij}\frac{\omega(X_{ij})-1}{h(X_{ij})}I(X_{ij}>0)\right\} \\
&= C_\rho + \frac{\Delta^*}{w_0}E_0\{h_0(X)\boldsymbol{u}(X)\}E_0[h_0(X)\{\omega(X)-1\}] \\
&\quad +\frac{\Delta^*}{w_1}E_0\{h_1(X)\boldsymbol{u}(X)\}E_0[h_1(X)\{\omega(X)-1\}] \\
&= C_\rho - \Delta^*\boldsymbol{m}E_0[h_1(X)\{\omega(X)-1\}],
\end{aligned}$$

where $\boldsymbol{m} = \boldsymbol{\gamma}^*/w_0 - SE_0\{h_1(X)\boldsymbol{u}(X)\} = \boldsymbol{\gamma}^*/w_0 - \mathbf{E}_{1\boldsymbol{u}}/\{w_0w_1\}$.

For $\boldsymbol{D}_{\boldsymbol{\theta}}$,

$$\begin{aligned}
\boldsymbol{D}_{\boldsymbol{\theta}} &= Cov\{\boldsymbol{\gamma}(\boldsymbol{\eta}^*), \boldsymbol{S}_{n,\boldsymbol{\theta}}^\top\} \\
&= \frac{1}{n\Delta^*}Cov\Bigg\{\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0), \sum_{j=1}^{n_1}\boldsymbol{Q}(X_{1j})^\top I(X_{1j}>0) \\
&\qquad\qquad -\sum_{ij}h_1(X_{ij})\boldsymbol{Q}(X_{ij})^\top I(X_{ij}>0)\Bigg\} \\
&= \frac{1}{n\Delta^*}Cov\left\{\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0), \sum_{j=1}^{n_1}h_0(X_{1j})\boldsymbol{Q}(X_{1j})^\top I(X_{1j}>0)\right\} \\
&\quad -\frac{1}{n\Delta^*}Cov\left\{\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0), \sum_{j=1}^{n_0}h_1(X_{0j})\boldsymbol{Q}(X_{0j})^\top I(X_{0j}>0)\right\} \\
&= (1-\rho^*)\Delta^*\boldsymbol{m}E_0\{h_1(X)\boldsymbol{Q}(X)^\top\} \\
&= \boldsymbol{m}(\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e})^\top.
\end{aligned}$$

With the form of $(\boldsymbol{D}_{\boldsymbol{\nu}}, \boldsymbol{D}_\rho, \boldsymbol{D}_{\boldsymbol{\theta}})$ and the form of $\boldsymbol{A}^{-1}$ in (4.25), $\boldsymbol{\Gamma}_3$ is given as

$$\begin{aligned}
\boldsymbol{\Gamma}_3 &= (\boldsymbol{D}_{\boldsymbol{\nu}}, \boldsymbol{D}_\rho, \boldsymbol{D}_{\boldsymbol{\theta}})\boldsymbol{A}^{-1}\boldsymbol{C}^\top \\
&= \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^\top - \frac{\rho^*(1-\rho^*)}{\Delta^*}\boldsymbol{D}_\rho\boldsymbol{C}_\rho^\top + \boldsymbol{D}_\rho\frac{\boldsymbol{e}^\top}{\Delta^*}\boldsymbol{C}_{\boldsymbol{\theta}}^\top + \boldsymbol{D}_{\boldsymbol{\theta}}\frac{\boldsymbol{e}}{\Delta^*}\boldsymbol{C}_\rho^\top \\
&\quad + \boldsymbol{D}_{\boldsymbol{\theta}}\left\{\boldsymbol{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)}\right\}\boldsymbol{C}_{\boldsymbol{\theta}}^\top \\
&= \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^\top + \boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{C}_{\boldsymbol{\theta}}^\top + \frac{1}{\Delta^*}\left\{\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{e} - \rho^*(1-\rho^*)\boldsymbol{D}_\rho\right\}\boldsymbol{C}_\rho^\top \\
&\quad + \left\{\frac{\boldsymbol{D}_\rho}{\Delta^*} - \frac{\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{e}}{\Delta^*\rho^*(1-\rho^*)}\right\}\boldsymbol{e}^\top\boldsymbol{C}_{\boldsymbol{\theta}}^\top.
\end{aligned}$$

With the forms of $\boldsymbol{D}_\rho$ and $\boldsymbol{D}_{\boldsymbol{\theta}}$, we have

$$\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1} = \boldsymbol{m}\boldsymbol{e}^\top \text{ and } \frac{1}{\Delta^*}\left\{\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{e} - \rho^*(1-\rho^*)\boldsymbol{D}_\rho\right\} = \rho^*(1-\rho^*)\boldsymbol{m} - \rho^*(1-\rho^*)\boldsymbol{C}_\rho/\Delta^*.$$

Hence,

$$\begin{aligned}
\boldsymbol{\Gamma}_3 &= \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^\top + \boldsymbol{m}\boldsymbol{e}^\top\boldsymbol{C}_{\boldsymbol{\theta}}^\top + \left(\boldsymbol{m} - \frac{\boldsymbol{C}_\rho}{\Delta^*}\right)\left\{\rho^*(1-\rho^*)\boldsymbol{C}_\rho^\top - \boldsymbol{e}^\top\boldsymbol{C}_{\boldsymbol{\theta}}^\top\right\} \\
&= \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^\top + \left(\boldsymbol{m} - \frac{\boldsymbol{C}_\rho}{\Delta^*}\right)\rho^*(1-\rho^*)\boldsymbol{C}_\rho^\top + \frac{1}{\Delta^*}\boldsymbol{C}_\rho\boldsymbol{e}^\top\boldsymbol{C}_{\boldsymbol{\theta}}^\top. \qquad (4.32)
\end{aligned}$$

Substituting $\boldsymbol{\Gamma}_2$ into (4.31) and $\boldsymbol{\Gamma}_3$ into (4.32) and using the facts that $\boldsymbol{C}_{\boldsymbol{\theta}} = \mathcal{M}_3$,

$$\boldsymbol{C}_{\boldsymbol{\nu}} + \rho^*(1-\rho^*)\boldsymbol{C}_\rho\boldsymbol{W} + \boldsymbol{D}_{\boldsymbol{\nu}} = \mathcal{M}_1, \qquad (4.33)$$

and

$$-\frac{(\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{e})(\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{e})^\top}{\Delta^*\rho^*(1-\rho^*)} + \frac{1}{\Delta^*}\boldsymbol{C}_\rho\boldsymbol{e}^\top\boldsymbol{C}_{\boldsymbol{\theta}}^\top + \frac{1}{\Delta^*}\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{e}\boldsymbol{C}_\rho^\top - \frac{\boldsymbol{C}_\rho}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_\rho^\top = -\frac{\mathcal{M}_2\mathcal{M}_2^\top}{\Delta^*\rho^*(1-\rho^*)},$$

we have

$$
\begin{aligned}
& \boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3 + \boldsymbol{\Gamma}_3^\top \\
= \; & (\mathcal{M}_1 - \boldsymbol{D}_{\boldsymbol{\nu}})\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}(\mathcal{M}_1 - \boldsymbol{D}_{\boldsymbol{\nu}})^\top - \frac{\mathcal{M}_2\mathcal{M}_2^\top}{\Delta^*\rho^*(1-\rho^*)} + \mathcal{M}_3\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\mathcal{M}_3^\top \\
& + \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^\top + \boldsymbol{C}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^\top + \rho^*(1-\rho^*)(\boldsymbol{m}\boldsymbol{C}_\rho^\top + \boldsymbol{C}_\rho\boldsymbol{m}^\top) - \frac{\boldsymbol{C}_\rho}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_\rho^\top \\
= \; & \mathcal{M}_1\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\mathcal{M}_1^\top - \frac{\mathcal{M}_2\mathcal{M}_2^\top}{\Delta^*\rho^*(1-\rho^*)} + \mathcal{M}_3\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\mathcal{M}_3^\top \\
& + \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}(\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1)^\top + (\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1)\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^\top + \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^\top \\
& + \rho^*(1-\rho^*)(\boldsymbol{m}\boldsymbol{C}_\rho^\top + \boldsymbol{C}_\rho\boldsymbol{m}^\top) - \frac{\boldsymbol{C}_\rho}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_\rho^\top. \qquad (4.34)
\end{aligned}
$$

Next we further simplify the form of $\boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3 + \boldsymbol{\Gamma}_3^\top$. Note that with (4.33), we have

$$
\begin{aligned}
& \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}(\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1)^\top + \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^\top + \rho^*(1-\rho^*)\boldsymbol{m}\boldsymbol{C}_\rho^\top - \frac{\boldsymbol{C}_\rho}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_\rho^\top \\
= \; & -\rho^*(1-\rho^*)\boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^\top\boldsymbol{C}_\rho^\top + \rho^*(1-\rho^*)\boldsymbol{m}\boldsymbol{C}_\rho^\top - \frac{\boldsymbol{C}_\rho}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_\rho^\top \\
= \; & \rho^*(1-\rho^*)\left(-\boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^\top + \boldsymbol{m} - \frac{\boldsymbol{C}_\rho}{\Delta^*}\right)\boldsymbol{C}_\rho^\top.
\end{aligned}
$$

With the forms of $\boldsymbol{D}_{\boldsymbol{\nu}}$, $\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}$, and $\boldsymbol{W}$, we have

$$
\begin{aligned}
\boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^\top &= -\frac{\nu_0}{\Delta^*(1-\rho^*)}\boldsymbol{\gamma}^* + \frac{\rho^*\nu_0 + (1-\rho^*)\nu_1}{\Delta^*\rho^*(1-\rho^*)}\mathbf{E}_{1\boldsymbol{u}} \\
&= -\frac{\nu_0}{\Delta^*(1-\rho^*)}\boldsymbol{\gamma}^* + \left\{\frac{1}{\Delta^*\rho^*(1-\rho^*)} - S\right\}\mathbf{E}_{1\boldsymbol{u}} \\
&= \frac{1-\nu_0}{\Delta^*(1-\rho^*)}\boldsymbol{\gamma}^* - S\mathbf{E}_{1\boldsymbol{u}} - \frac{1}{\Delta^*\rho^*(1-\rho^*)}\{\rho^*\boldsymbol{\gamma}^* - \mathbf{E}_{1\boldsymbol{u}}\} \\
&= \boldsymbol{m} - \frac{\boldsymbol{C}_\rho}{\Delta^*}.
\end{aligned}
$$

Hence,

$$
\boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}(\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1)^\top + \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^\top + \rho^*(1-\rho^*)\boldsymbol{m}\boldsymbol{C}_\rho^\top - \frac{\boldsymbol{C}_\rho}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_\rho^\top = \mathbf{0}
$$

and $\boldsymbol{\Gamma}_2+\boldsymbol{\Gamma}_3+\boldsymbol{\Gamma}_3^\top$ in (4.34) becomes

$$\begin{aligned}\boldsymbol{\Gamma}_2+\boldsymbol{\Gamma}_3+\boldsymbol{\Gamma}_3^\top &= \mathcal{M}_1\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\mathcal{M}_1^\top-\frac{\mathcal{M}_2\mathcal{M}_2^\top}{\Delta^*\rho^*(1-\rho^*)}+\mathcal{M}_3\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\mathcal{M}_3^\top\\ &\quad+(\boldsymbol{C}_{\boldsymbol{\nu}}-\mathcal{M}_1)\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^\top+\rho^*(1-\rho^*)\boldsymbol{C}_\rho\boldsymbol{m}^\top. \qquad (4.35)\end{aligned}$$

With $\boldsymbol{\Gamma}_1$ in (4.30) and $\boldsymbol{\Gamma}_2+\boldsymbol{\Gamma}_3+\boldsymbol{\Gamma}_3^\top$ in (4.35), to show that $\boldsymbol{\Gamma}=\boldsymbol{\Gamma}_1+\boldsymbol{\Gamma}_2+\boldsymbol{\Gamma}_3+\boldsymbol{\Gamma}_3^\top$, we need to argue that

$$-\frac{\boldsymbol{\gamma}^*\boldsymbol{\gamma}^{*\top}}{\Delta^*}=(\boldsymbol{C}_{\boldsymbol{\nu}}-\mathcal{M}_1)\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^\top+\rho^*(1-\rho^*)\boldsymbol{C}_\rho\boldsymbol{m}^\top-\frac{1}{w_0}\mathbf{E}_{0\boldsymbol{u}}\mathbf{E}_{0\boldsymbol{u}}^\top-\frac{1}{w_1}\mathbf{E}_{1\boldsymbol{u}}\mathbf{E}_{1\boldsymbol{u}}^\top. \quad (4.36)$$

Note that

$$\boldsymbol{C}_{\boldsymbol{\nu}}=\mathcal{M}_1+(w_0,w_1)\otimes\frac{\boldsymbol{\gamma}^*}{\Delta^*}.$$

Then

$$(\boldsymbol{C}_{\boldsymbol{\nu}}-\mathcal{M}_1)\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^\top = -\frac{\nu_0}{\Delta^*}\boldsymbol{\gamma}\mathbf{E}_{0\boldsymbol{u}}^\top-\frac{\nu_1}{\Delta^*}\boldsymbol{\gamma}\mathbf{E}_{1\boldsymbol{u}}^\top=-\boldsymbol{\gamma}\left(\frac{\nu_0}{\Delta^*}\mathbf{E}_{0\boldsymbol{u}}^\top+\frac{\nu_1}{\Delta^*}\mathbf{E}_{1\boldsymbol{u}}\right)^\top. \quad (4.37)$$

Recall that

$$\rho^*(1-\rho^*)\boldsymbol{C}_\rho=\rho^*\boldsymbol{\gamma}^*-\mathbf{E}_{1\boldsymbol{u}}=\rho^*\mathbf{E}_{0\boldsymbol{u}}-(1-\rho^*)\mathbf{E}_{1\boldsymbol{u}}$$

and

$$\boldsymbol{m}=\boldsymbol{\gamma}^*/w_0-\mathbf{E}_{1\boldsymbol{u}}/\{w_0w_1\}=\mathbf{E}_{0\boldsymbol{u}}/w_0-\mathbf{E}_{1\boldsymbol{u}}/w_1.$$

Then

$$\begin{aligned}&\rho^*(1-\rho^*)\boldsymbol{C}_\rho\boldsymbol{m}^\top-\frac{1}{w_0}\mathbf{E}_{0\boldsymbol{u}}\mathbf{E}_{0\boldsymbol{u}}^\top-\frac{1}{w_1}\mathbf{E}_{1\boldsymbol{u}}\mathbf{E}_{1\boldsymbol{u}}^\top\\ =\;&\{\rho^*\mathbf{E}_{0\boldsymbol{u}}-(1-\rho^*)\mathbf{E}_{1\boldsymbol{u}}\}\{\mathbf{E}_{0\boldsymbol{u}}/w_0-\mathbf{E}_{1\boldsymbol{u}}/w_1\}^\top-\frac{1}{w_0}\mathbf{E}_{0\boldsymbol{u}}\mathbf{E}_{0\boldsymbol{u}}^\top-\frac{1}{w_1}\mathbf{E}_{1\boldsymbol{u}}\mathbf{E}_{1\boldsymbol{u}}^\top\\ =\;&-\frac{1-\rho^*}{w_0}\mathbf{E}_{0\boldsymbol{u}}\mathbf{E}_{0\boldsymbol{u}}^\top-\frac{1-\rho^*}{w_0}\mathbf{E}_{1\boldsymbol{u}}\mathbf{E}_{0\boldsymbol{u}}^\top-\frac{\rho^*}{w_1}\mathbf{E}_{0\boldsymbol{u}}\mathbf{E}_{1\boldsymbol{u}}^\top-\frac{\rho^*}{w_1}\mathbf{E}_{1\boldsymbol{u}}\mathbf{E}_{1\boldsymbol{u}}^\top\\ =\;&-(\mathbf{E}_{0\boldsymbol{u}}+\mathbf{E}_{1\boldsymbol{u}})\left(\frac{1-\rho^*}{w_0}\mathbf{E}_{0\boldsymbol{u}}+\frac{\rho^*}{w_1}\mathbf{E}_{1\boldsymbol{u}}\right)^\top\\ =\;&-\boldsymbol{\gamma}^*\left(\frac{1-\nu_0}{\Delta^*}\mathbf{E}_{0\boldsymbol{u}}+\frac{1-\nu_1}{\Delta^*}\mathbf{E}_{1\boldsymbol{u}}\right)^\top. \qquad (4.38)\end{aligned}$$

Since $\mathbf{E}_{0\boldsymbol{u}} + \mathbf{E}_{1\boldsymbol{u}} = \boldsymbol{\gamma}$, combining (4.37) and (4.38) gives

$$(\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1)\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^{\top} + \rho^*(1-\rho^*)\boldsymbol{C}_{\rho}\boldsymbol{m}^{\top} - \frac{1}{w_0}\mathbf{E}_{0\boldsymbol{u}}\mathbf{E}_{0\boldsymbol{u}}^{\top} - \frac{1}{w_1}\mathbf{E}_{1\boldsymbol{u}}\mathbf{E}_{1\boldsymbol{u}}^{\top} = -\frac{1}{\Delta}\boldsymbol{\gamma}\boldsymbol{\gamma}^{*\top},$$

which verifies (4.36). Hence,

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3 + \boldsymbol{\Gamma}_3^{\top}.$$

Applying Slutsky's theorem and the central limit theorem to (4.29), we get as $n \to \infty$

$$n^{1/2}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\right) \to N\left(\mathbf{0}, \boldsymbol{\Gamma}\right),$$

in distribution. This completes the proof of Theorem 4.2.

### Proof of Theorem 4.3

- *Approximations of $\hat{\psi}_0$ and $\hat{\psi}_1$*

To develop the asymptotic properties of $(\hat{\mathcal{G}}_0, \hat{\mathcal{G}}_1)$, we first find the linear approximations of $\hat{\psi}_0$ and $\hat{\psi}_1$. We start with $\hat{\psi}_0$.

With the form of $\hat{p}_{ij}$ in (4.26), the MELE $\hat{\psi}_0$ is then given by

$$\begin{aligned}
\hat{\psi}_0 &= \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \hat{p}_{ij}\{2X_{ij}\hat{G}_0(X_{ij})\} \\
&= \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \hat{p}_{ij}X_{ij}\left\{2\sum_{l=0}^{1}\sum_{s=1}^{n_{l1}} \hat{p}_{ls} I(X_{ls} \le X_{ij})\right\} \\
&= \{nw_0(1-\hat{\nu}_0) + nw_1(1-\hat{\nu}_1)\}^{-2} \times \\
&\quad \sum_{ij}\sum_{ls} \frac{2I(X_{ls} \le X_{ij})X_{ij} \cdot I(X_{ij} > 0)I(X_{ls} > 0)}{\left\{1 + \hat{\rho}[\exp\{\hat{\boldsymbol{\theta}}^{\top}\boldsymbol{Q}(X_{ij})\} - 1]\right\}\left\{1 + \hat{\rho}[\exp\{\hat{\boldsymbol{\theta}}^{\top}\boldsymbol{Q}(X_{ls})\} - 1]\right\}}.
\end{aligned}$$

Note that $\hat{\psi}_0$ is a function of $\boldsymbol{\eta}$, and hence we define

$$\begin{aligned}
\psi_0(\boldsymbol{\eta}) &= \{nw_0(1-\nu_0) + nw_1(1-\nu_1)\}^{-2} \times \\
&\quad \sum_{ij}\sum_{ls} \frac{2I(X_{ls} \le X_{ij})X_{ij} \cdot I(X_{ij} > 0)I(X_{ls} > 0)}{\{1 + \rho[\exp\{\boldsymbol{\theta}^{\top}\boldsymbol{Q}(X_{ij})\} - 1]\}\{1 + \rho[\exp\{\boldsymbol{\theta}^{\top}\boldsymbol{Q}(X_{ls})\} - 1]\}}.
\end{aligned}$$

We then have $\hat{\psi}_0 = \psi_0(\hat{\boldsymbol{\eta}})$. With the definition of $\Delta^*$ and $h(x)$, we have

$$\psi_0(\boldsymbol{\eta}^*) = \frac{1}{(n\Delta^*)^2}\sum_{ij}\sum_{ls}\frac{2I(X_{ls} \le X_{ij})X_{ij}}{h(X_{ij})h(X_{ls})}I(X_{ij} > 0)I(X_{ls} > 0)$$

and $E_0\{\psi_0(\boldsymbol{\eta}^*)\} = \psi_0$.

By Theorem 4.1, we have $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}^* + O_p(n^{-1/2})$. Applying the first-order Taylor expansion gives

$$\psi_0(\hat{\boldsymbol{\eta}}) = \psi_0(\boldsymbol{\eta}^*) + \left\{\frac{\partial\psi_0(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\eta}}\right\}^\top (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + o_p(n^{-1/2}). \tag{4.39}$$

Define

$$\begin{aligned} U(a,b) &= \frac{I(b \le a)a}{h(a)h(b)}I(a > 0)I(b > 0), \quad V_{nil} = \frac{1}{n_i}\frac{1}{n_l}\sum_{j=1}^{n_i}\sum_{s=1}^{n_l}U(X_{ij}, X_{ls}), \\ V_{ni} &= \frac{1}{n_i^2}\sum_{j=1}^{n_i}\sum_{s=1}^{n_i}2U(X_{ij}, X_{is}) = \frac{1}{n_i^2}\sum_{j=1}^{n_i}\sum_{s=1}^{n_i}\left\{U(X_{ij}, X_{is}) + U(X_{is}, X_{ij})\right\}, \end{aligned}$$

for $i, l \in \{0, 1\}$ and $i \ne l$. We then rewrite $\psi_0(\boldsymbol{\eta}^*)$ as

$$\begin{aligned} \psi_0(\boldsymbol{\eta}^*) &= \frac{1}{(\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1}w_i w_l \frac{1}{n_i}\frac{1}{n_l}\sum_{j=1}^{n_i}\sum_{s=1}^{n_l}2U(X_{ij}, X_{ls}) \\ &= \frac{1}{(\Delta^*)^2}\left\{\sum_{i=0}^{1}w_i^2 V_{ni} + \sum_{i=0}^{1}\sum_{l\ne i}w_i w_l 2V_{nil}\right\}. \end{aligned}$$

Note that $V_{ni}$ is a von Mises statistic (Mises, 1947). We denote the associated U-statistic by

$$U_{ni} = \binom{n_i}{2}^{-1}\sum_{1\le j<s\le n_i}\left\{U(X_{ij}, X_{is}) + U(X_{is}, X_{ij})\right\}.$$

According to Serfling (1980), the projection of $U_{ni}$ is defined as

$$\hat{U}_{ni} = E\{U_i(X_{i1})\} + \frac{2}{n_i}\sum_{j=1}^{n_i}\left[U_i(X_{ij}) - E\{U_i(X_{i1})\}\right],$$

where $U_i(a) = E\{U(a, X_{i1}) + U(X_{i1}, a)\}$. It follows from Serfling (1980, p. 190 & p. 206) that under Condition C5,

$$\sqrt{n_i}(\hat{U}_{ni} - U_{ni}) = o_p(1) \text{ and} \sqrt{n_i}(V_{ni} - U_{ni}) = o_p(1).$$

This leads to

$$V_{ni} = E\{U_i(X_{i1})\} + \frac{2}{n_i}\sum_{j=1}^{n_i}[U_i(X_{ij}) - E\{U_i(X_{i1})\}] + o_p(n^{-1/2}).$$

When $l \neq i$, $V_{nil}$ is a two-sample U-statistic. Define $U_{il} = E\{U(X_{i1}, X_{l1})\}$, $U_{il10}(a) = E\{U(a, X_{l1})\} - U_{il}$, and $U_{il01}(a) = E\{U(X_{i1}, a)\} - U_{il}$. From Theorem 12.6 in Van der Vaart (2000), we have

$$V_{nil} = U_{il} + \frac{1}{n_i}\sum_{j=1}^{n_i} U_{il10}(X_{ij}) + \frac{1}{n_l}\sum_{s=1}^{n_l} U_{il01}(X_{ls}) + o_p(n^{-1/2}).$$

Since

$$\begin{aligned} E\{U_i(X_{i1})\} &= 2E\{U(X_{i1}, X_{i1})\} = 2U_{ii} \\ U_i(a) - E\{U_i(X_{i1})\} &= U_{ii10}(a) + U_{ii01}(a), \end{aligned}$$

we have

$$V_{ni} = 2\left\{U_{ii} + \frac{1}{n_i}\sum_{j=1}^{n_i} U_{ii10}(X_{ij}) + \frac{1}{n_i}\sum_{j=1}^{n_i} U_{ii01}(X_{ij})\right\} + o_p(n^{-1/2}).$$

Hence,

$$\begin{aligned} \psi_0(\boldsymbol{\eta}^*) &= \frac{2}{(\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} w_i w_l U_{il} + \frac{2}{(\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} w_i w_l \frac{1}{n_i}\sum_{j=1}^{n_i} U_{il10}(X_{ij}) \\ &\quad + \frac{2}{(\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} w_i w_l \frac{1}{n_l}\sum_{s=1}^{n_l} U_{il01}(X_{ls}) + o_p(n^{-1/2}) \\ &= \frac{2}{(n\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} n_i n_l U_{il} + \frac{2}{(n\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} n_l \sum_{j=1}^{n_i} U_{il10}(X_{ij}) \\ &\quad + \frac{2}{(n\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} n_i \sum_{s=1}^{n_l} U_{il01}(X_{ls}) + o_p(n^{-1/2}). \end{aligned} \tag{4.40}$$

We now simplify each term in (4.40). With Lemma 4.1 and the definition of $U_{il}$, we have

$$\begin{aligned}\sum_{l=0}^{1} n_l E\{U(X_{i1}, X_{l1})|X_{i1}\} &= \sum_{l=0}^{1} n_l E\left\{\frac{I(X_{l1} \le X_{i1})X_{i1}}{h(X_{i1})h(X_{l1})} I(X_{i1} > 0)I(X_{l1} > 0)|X_{i1}\right\} \\ &= n\Delta^* \frac{X_{i1}}{h(X_{i1})} I(X_{i1} > 0)E_0\{I(X \le X_{i1})\} \\ &= n\Delta^* \frac{X_{i1}G_0(X_{i1})}{h(X_{i1})} I(X_{i1} > 0).\end{aligned}$$

Hence,

$$\begin{aligned}\frac{2}{(n\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} n_i n_l U_{il} &= \frac{2}{(n\Delta^*)^2}\sum_{i=0}^{1} n_i E\left[\sum_{l=0}^{1} n_l E\{U(X_{i1}, X_{l1})|X_i\}\right] \\ &= \frac{2}{n\Delta^*}\sum_{i=0}^{1} n_i E\left\{\frac{X_{i1}G_0(X_{i1})}{h(X_{i1})} I(X_{i1} > 0)\right\}.\end{aligned}$$

Using the result in Lemma 4.1, we have

$$\frac{2}{(n\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} n_i n_l U_{il} = 2E_0\{XG_0(X)\} = \psi_0.$$

We move to the second term of $\psi_0(\boldsymbol{\eta}^*)$ in (4.40). Recall that

$$U_{il10}(a) = E\{U(a, X_{l1})\} - E\{U(X_{i1}, X_{l1})\}.$$

We then have

$$\begin{aligned}\sum_{l=0}^{1} n_l U_{il10}(X_{ij}) &= \sum_{l=0}^{1} n_l \{E\{U(X_{ij}, X_{l1})|X_{ij}\} - E[E\{U(X_{ij}, X_{l1})|X_{ij}\}]\} \\ &= n\Delta^* \left[\frac{X_{ij}G_0(X_{ij})}{h(X_{ij})} I(X_{ij} > 0) - E\left\{\frac{X_{ij}G_0(X_{ij})}{h(X_{ij})} I(X_{ij} > 0)\right\}\right].\end{aligned}$$

This leads to

$$
\begin{aligned}
& \frac{2}{(n\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} n_l \sum_{j=1}^{n_i} U_{il10}(X_{ij}) \\
= \ & \frac{2}{n\Delta^*}\sum_{ij}\left[\frac{X_{ij}G_0(X_{ij})}{h(X_{ij})}I(X_{ij}>0) - E\left\{\frac{X_{ij}G_0(X_{ij})}{h(X_{ij})}I(X_{ij}>0)\right\}\right] \\
= \ & \frac{2}{n\Delta^*}\sum_{ij}\frac{I(X_{ij}>0)}{h(X_{ij})}X_{ij}G_0(X_{ij}) - \psi_0.
\end{aligned}
$$

Similarly, with the definition of $U_{il01}(a)$, we have

$$
\sum_{i=0}^{1} n_i U_{il01}(X_{ls}) \ = \ \sum_{i=0}^{1} n_i \left\{E\{U(X_{i1}, X_{ls})|X_{ls}\} - E[E\{U(X_{i1}, X_{ls})|X_{ls}\}]\right\}.
$$

Note that

$$
\begin{aligned}
\sum_{i=0}^{1} n_i E\{U(X_{i1}, X_{ls})|X_{ls}\} \ &= \ \sum_{i=0}^{1} n_i E\left\{\frac{I(X_{ls}\le X_{i1})X_{i1}}{h(X_{i1})h(X_{ls})}I(X_{i1}>0)I(X_{ls}>0)|X_{ls}\right\} \\
&= \ n\Delta^*\frac{I(X_{ls}>0)}{h(X_{ls})}E\left\{XI(X_{ls}\le X)\right\} \\
&= \ n\Delta^*\frac{I(X_{ls}>0)}{h(X_{ls})}\int_{X_{ls}}^{\infty} x dG_0(x).
\end{aligned}
$$

Together with the result of Lemma 4.1, we have

$$
\begin{aligned}
& \frac{2}{(n\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} n_i \sum_{s=1}^{n_l} U_{il01}(X_{ls}) \\
= \ & \frac{2}{n\Delta^*}\sum_{ls}\left[\frac{I(X_{ls}>0)}{h(X_{ls})}\int_{X_{ls}}^{\infty} x dG_0(x) - E\left\{\frac{I(X_{ls}>0)}{h(X_{ls})}\int_{X_{ls}}^{\infty} x dG_0(x)\right\}\right] \\
= \ & \frac{2}{n\Delta^*}\sum_{ls}\frac{I(X_{ls}>0)}{h(X_{ls})}\int_{X_{ls}}^{\infty} x dG_0(x) - \psi_0.
\end{aligned}
$$

For $a > 0$, we define the function

$$H_0(a) = \left\{ aG_0(a) + \int_a^\infty x dG_0(x) \right\}.$$

The approximation of $\psi_0(\boldsymbol{\eta}^*)$ is then given by

$$\psi_0(\boldsymbol{\eta}^*) = \frac{1}{n\Delta^*} \sum_{ij} \frac{I(X_{ij} > 0)}{h(X_{ij})} \cdot \{2H_0(X_{ij})\} - \psi_0 + o_p(n^{-1/2}).$$

We also need the first derivative of $\psi_0(\boldsymbol{\eta})$ when finding the approximation of $\psi_0(\hat{\boldsymbol{\eta}})$. We take the first derivative of $\psi_0(\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$ and evaluate the derivative at the true value $\boldsymbol{\eta}^*$. This leads to

$$\begin{aligned}
\frac{\partial \psi_0(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu}} &= \frac{2}{\Delta^*} \psi_0(\boldsymbol{\eta}^*) \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}, \\
\frac{\partial \psi_0(\boldsymbol{\eta}^*)}{\partial \rho} &= -\frac{2}{(n\Delta^*)^2} \sum_{ij} \sum_{ls} \left\{ \frac{\omega(X_{ij}) - 1}{h(X_{ij})^2 h(X_{ls})} + \frac{\omega(X_{ls}) - 1}{h(X_{ij}) h(X_{ls})^2} \right\} \\
&\quad \times I(X_{ls} \leq X_{ij}) X_{ij} I(X_{ij} > 0) I(X_{ls} > 0), \\
\frac{\partial \psi_0(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta}} &= -\frac{2}{(n\Delta^*)^2} \sum_{ij} \sum_{ls} \left\{ \frac{\rho^* \omega(X_{ij}) \boldsymbol{Q}(X_{ij})}{h(X_{ij})^2 h(X_{ls})} + \frac{\rho^* \omega(X_{ls}) \boldsymbol{Q}(X_{ls})}{h(X_{ij}) h(X_{ls})^2} \right\} \\
&\quad \times I(X_{ls} \leq X_{ij}) X_{ij} I(X_{ij} > 0) I(X_{ls} > 0), \\
&= -\frac{2}{(n\Delta^*)^2} \sum_{ij} \sum_{ls} \{ h_1(X_{ij}) \boldsymbol{Q}(X_{ij}) + h_1(X_{ls}) \boldsymbol{Q}(X_{ls}) \} \\
&\quad \times \frac{I(X_{ls} \leq X_{ij}) X_{ij} I(X_{ij} > 0) I(X_{ls} > 0)}{h(X_{ij}) h(X_{ls})}.
\end{aligned}$$

By the law of large numbers, we have

$$\frac{\partial \psi_0(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}} = E\left\{ \frac{\partial \psi_0(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}} \right\} + o_p(1) = \boldsymbol{C}_0 + o_p(1),$$

with $\boldsymbol{C}_0 = (\boldsymbol{C}_{0\boldsymbol{\nu}}^\top, \boldsymbol{C}_{0\rho}, \boldsymbol{C}_{0\boldsymbol{\theta}}^\top)^\top$.

Since $E\{\psi_0(\boldsymbol{\eta}^*)\} = \psi_0$, we have

$$\boldsymbol{C}_{0\boldsymbol{\nu}} = \frac{2\psi_0}{\Delta^*} (w_0, w_1)^\top.$$

For $\boldsymbol{C}_{0\rho} = E\{\partial\psi_0(\boldsymbol{\eta}^*)/\partial\rho\}$,

$$\begin{aligned}
& E\left\{\frac{\partial\psi_0(\boldsymbol{\eta}^*)}{\partial\rho}\right\} \\
= & -\frac{2}{(n\Delta^*)^2}E\left[\sum_{ij}E\left\{\sum_{ls}\frac{\omega(X_{ij})-1}{h(X_{ij})^2h(X_{ls})}I(X_{ls}\le X_{ij})X_{ij}I(X_{ij}>0)I(X_{ls}>0)|X_{ij}\right\}\right] \\
& -\frac{2}{(n\Delta^*)^2}E\left[\sum_{ls}E\left\{\sum_{ij}\frac{\omega(X_{ls})-1}{h(X_{ij})h(X_{ls})^2}I(X_{ls}\le X_{ij})X_{ij}I(X_{ij}>0)I(X_{ls}>0)|X_{ls}\right\}\right] \\
= & -\frac{2}{n\Delta^*}E\left[\sum_{ij}\frac{\omega(X_{ij})-1}{h(X_{ij})^2}X_{ij}I(X_{ij}>0)G_0(x)\right] \\
& -\frac{2}{n\Delta^*}E\left[\sum_{ls}\frac{\omega(X_{ls})-1}{h(X_{ls})^2}I(X_{ls}>0)E_0\left\{I(X_{ls}\le X)X\right\}\right] \\
= & -\frac{2}{n\Delta^*}E\left[\sum_{ij}\frac{\omega(X_{ij})-1}{h(X_{ij})^2}I(X_{ij}>0)H_0(X_{ij})\right] \\
= & -2E_0\left\{\frac{H_0(X)\{\omega(X)-1\}}{h(X)}\right\}.
\end{aligned}$$

The expression for $\boldsymbol{C}_{0\boldsymbol{\theta}}$ can be found in a similar manner:

$$\boldsymbol{C}_{0\boldsymbol{\theta}} = -2E_0\left\{h_1(X)H_0(X)\boldsymbol{Q}(X)\right\}.$$

The details are omitted here.

It can be verified that the matrix $\boldsymbol{C}_0^\top$ is the same as the matrix $\boldsymbol{C}$ in (4.28) when we set $\boldsymbol{u}(x;\boldsymbol{\nu},\boldsymbol{\theta}) = 2H_0(x)$ in the definition of $\boldsymbol{\gamma}$ in (4.10). Hence, the expression in (4.39) can be further written as

$$\begin{aligned}
\hat{\psi}_0 &= \frac{1}{n\Delta^*}\sum_{ij}\frac{I(X_{ij}>0)}{h(X_{ij})}\cdot\{2H_0(X_{ij})\} - \psi_0 + \boldsymbol{C}_0^\top(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}^*) + o_p(n^{-1/2}) \\
&= \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\hat{p}_{ij}\{2H_0(X_{ij})\} - \psi_0 + o_p(n^{-1/2}). \qquad (4.41)
\end{aligned}$$

The remaining term $o_p(n^{-1/2})$ is introduced by the projection of the von Mises statistic and the U-statistic when approximating $\psi_0(\boldsymbol{\eta}^*)$.

Define

$$\mathcal{H}_0(a) = 2H_0(a) - \psi_0.$$

With the natural constraint $\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} p_{ij} = 1$, Equation (4.41) implies

$$\hat{\psi}_0 = \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \hat{p}_{ij}\mathcal{H}_0(X_{ij}) + o_p(n^{-1/2}). \tag{4.42}$$

Next, we consider the approximation of MELE $\hat{\psi}_1$. Recall that

$$\hat{\psi}_1 = \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \hat{p}_{ij}\omega(X_{ij};\hat{\boldsymbol{\theta}})X_{ij}\left\{2\sum_{l=0}^{1}\sum_{s=1}^{n_{l1}} \hat{p}_{ls}\omega(X_{ls};\hat{\boldsymbol{\theta}})I(X_{ls} \le X_{ij})\right\}.$$

With the definition of $\hat{p}_{ij}$ in (4.26), the MELE $\hat{\psi}_i$ can be written as

$$\begin{aligned}\hat{\psi}_1 &= \{nw_0(1-\hat{\nu}_0) + nw_1(1-\hat{\nu}_1)\}^{-2}\\ &\quad\times\sum_{ij}\sum_{ls}\frac{2I(X_{ls}\le X_{ij})X_{ij}\omega(X_{ij};\hat{\boldsymbol{\theta}})\omega(X_{ls};\hat{\boldsymbol{\theta}})I(X_{ij}>0)I(X_{ls}>0)}{\left\{1+\hat{\rho}[\exp\{\hat{\boldsymbol{\theta}}^{\top}\boldsymbol{Q}(X_{ij})\}-1]\right\}\left\{1+\hat{\rho}[\exp\{\hat{\boldsymbol{\theta}}^{\top}\boldsymbol{Q}(X_{ls})\}-1]\right\}}.\end{aligned} \tag{4.43}$$

Define

$$\begin{aligned}\tilde{U}(a,b) &= \frac{I(b\le a)a}{h(a)h(b)}\omega(a)\omega(b)I(a>0)I(b>0), \quad \tilde{V}_{nil} = \frac{1}{n_i}\frac{1}{n_l}\sum_{j=1}^{n_i}\sum_{s=1}^{n_l}\tilde{U}(X_{ij},X_{ls}),\\ \tilde{V}_{ni} &= \frac{1}{n_i^2}\sum_{j=1}^{n_i}\sum_{s=1}^{n_i}2\tilde{U}(X_{ij},X_{is}) = \frac{1}{n_i^2}\sum_{j=1}^{n_i}\sum_{s=1}^{n_i}\left\{\tilde{U}(X_{ij},X_{is})+\tilde{U}(X_{is},X_{ij})\right\},\end{aligned}$$

for $i, l \in \{0,1\}$. We use $\psi_1(\hat{\boldsymbol{\eta}})$ to denote $\hat{\psi}_1$ and have

$$\begin{aligned}
\psi_1(\boldsymbol{\eta}^*) &= \frac{1}{(n\Delta^*)^2}\sum_{ij}\sum_{ls} 2\tilde{U}(X_{ij}, X_{ls}) \\
&= \frac{1}{(\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} w_i w_l \frac{1}{n_i}\frac{1}{n_l}\sum_{j=1}^{n_i}\sum_{s=1}^{n_l} 2\tilde{U}(X_{ij}, X_{ls}) \\
&= \frac{1}{(\Delta^*)^2}\left\{\sum_{i=0}^{1} w_i^2 \tilde{V}_{ni} + \sum_{i=0}^{1}\sum_{l\neq i} w_i w_l 2\tilde{V}_{nil}\right\}.
\end{aligned}$$

Note that $\tilde{V}_{ni}$ is a von Mises statistic and $\tilde{V}_{nil}$ is a two-sample U-statistic. Using the technique used to obtain the approximation of $\psi_0(\boldsymbol{\eta}^*)$ in (4.40), we have

$$\begin{aligned}
\psi_1(\boldsymbol{\eta}^*) &= \frac{2}{(n\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} n_i n_l \tilde{U}_{il} + \frac{2}{(n\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} n_l \sum_{j=1}^{n_i}\tilde{U}_{il10}(X_{ij}) \\
&\quad + \frac{2}{(n\Delta^*)^2}\sum_{i=0}^{1}\sum_{l=0}^{1} n_i \sum_{s=1}^{n_l}\tilde{U}_{il01}(X_{ls}) + o_p(n^{-1/2}),
\end{aligned}$$

where $\tilde{U}_{il} = E\{\tilde{U}(X_{i1}, X_{l1})\}$, $\tilde{U}_{il10}(a) = E\{\tilde{U}(a, X_{l1})\} - \tilde{U}_{il}$, and $\tilde{U}_{il01}(a) = E\{\tilde{U}(X_{i1}, a)\} - \tilde{U}_{il}$.

With Lemma 4.1 and the definition of $\tilde{U}(a,b)$, we have

$$\begin{aligned}
&\sum_{l=0}^{1} n_l E\{\tilde{U}(X_{ij}, X_{l1})|X_{ij}\} \\
&= \sum_{l=0}^{1} n_l E\left\{\frac{I(X_{l1} \leq X_{ij})X_{ij}}{h(X_{ij})h(X_{l1})}\omega(X_{ij})\omega(X_{l1})I(X_{ij} > 0)I(X_{l1} > 0)|X_{ij}\right\} \\
&= n\Delta^* \frac{X_{ij}\omega(X_{ij})}{h(X_{ij})} I(X_{ij} > 0) E_0\left\{\omega(X)I(X \leq X_{ij})\right\} \\
&= n\Delta^* \frac{X_{ij}\omega(X_{ij})G_1(X_{ij})}{h(X_{ij})} I(X_{ij} > 0)
\end{aligned}$$

and

$$
\begin{aligned}
& \sum_{i=0}^{1} n_i E\{\tilde{U}(X_{i1}, X_{ls})|X_{ls}\} \\
= \ & \sum_{i=0}^{1} n_i E\left\{\frac{I(X_{ls} \le X_{i1})X_{i1}}{h(X_{i1})h(X_{ls})}\omega(X_{i1})\omega(X_{ls})I(X_{i1} > 0)I(X_{ls} > 0)|X_{ls}\right\} \\
= \ & n\Delta^* \frac{\omega(X_{ls})I(X_{ls} > 0)}{h(X_{ls})} E_0\left\{X\omega(X)I(X_{ls} \le X)\right\} \\
= \ & n\Delta^* \frac{\omega(X_{ls})I(X_{ls} > 0)}{h(X_{ls})} \int_{X_{ls}}^{\infty} x dG_1(x).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\frac{2}{(n\Delta^*)^2} \sum_{i=0}^{1}\sum_{l=0}^{1} n_i n_l \tilde{U}_{il} &= \frac{2}{n\Delta^*}\sum_{i=0}^{1} n_i E\left\{\frac{X_{i1}\omega(X_{i1})G_1(X_{i1})}{h(X_{i1})} I(X_{i1} > 0)\right\} \\
&= 2E_0\{X\omega(X)G_1(X)\} \\
&= \psi_1, \\
\frac{2}{(n\Delta^*)^2} \sum_{i=0}^{1}\sum_{l=0}^{1} n_l \sum_{j=1}^{n_i} \tilde{U}_{il10}(X_{ij}) &= \frac{2}{n\Delta^*}\sum_{ij} \frac{\omega(X_{ij})I(X_{ij} > 0)}{h(X_{ij})} X_{ij} G_1(X_{ij}) - \psi_1, \\
\frac{2}{(n\Delta^*)^2} \sum_{i=0}^{1}\sum_{l=0}^{1} n_i \sum_{s=1}^{n_l} \tilde{U}_{il01}(X_{ls}) &= \frac{2}{n\Delta^*}\sum_{ls} \frac{\omega(X_{ls})I(X_{ls} > 0)}{h(X_{ls})} \int_{X_{ls}}^{\infty} x dG_1(x) - \psi_1.
\end{aligned}
$$

Hence, $\psi_1(\boldsymbol{\eta}^*)$ is given by

$$
\psi_1(\boldsymbol{\eta}^*) = \frac{1}{n\Delta^*}\sum_{ij} \frac{\omega(X_{ij})I(X_{ij} > 0)}{h(X_{ij})} \cdot \{2H_1(X_{ij})\} - \psi_1 + o_p(n^{-1/2}),
$$

where $H_1(a) = aG_1(a) + \int_a^{\infty} x dG_1(x)$ for $a > 0$.

Applying the first-order Taylor expansion to $\hat{\psi}_1$ in (4.43) yields

$$
\psi_1(\hat{\boldsymbol{\eta}}) = \psi_1(\boldsymbol{\eta}^*) + \left\{\frac{\partial \psi_0(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\}^{\top} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + o_p(n^{-1/2}).
$$

With the law of large numbers, we have

$$\frac{\partial \psi_1(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}} = E\left\{\frac{\partial \psi_1(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\} + o_p(1) = \boldsymbol{C}_1 + o_p(1)$$

with $\boldsymbol{C}_1 = (\boldsymbol{C}_{1\boldsymbol{\nu}}^\top, \boldsymbol{C}_{1\rho}, \boldsymbol{C}_{1\boldsymbol{\theta}}^\top)^\top$ and

$$\boldsymbol{C}_{1\boldsymbol{\nu}} = \frac{2\psi_1}{\Delta^*}\begin{pmatrix} w_0 \\ w_1 \end{pmatrix}, \boldsymbol{C}_{1\rho} = -2E_0\left\{\frac{H_1(X)\{\omega(X)-1\}}{h(X)}\right\}, \boldsymbol{C}_{1\boldsymbol{\theta}} = 2E_0\left\{h_0(X)H_1(X)\boldsymbol{Q}(X)\right\}.$$

The expression of each element in $\boldsymbol{C}_1$ can be found similarly to the derivation of $\boldsymbol{C}_0$; we omit the details. By setting $\boldsymbol{u}(x;\boldsymbol{\nu},\boldsymbol{\theta}) = 2\omega(x;\boldsymbol{\theta})H_1(x)$ in (4.10), we can verify that the matrix $\boldsymbol{C}_1^\top$ is the same as the matrix $\boldsymbol{C}$ in (4.28). Hence, the approximation is given by

$$\psi_1(\hat{\boldsymbol{\eta}}) = \frac{1}{n\Delta^*}\sum_{ij}\frac{\omega(X_{ij})I(X_{ij}>0)}{h(X_{ij})}\cdot\{2H_1(X_{ij})\} - \psi_1 + \boldsymbol{C}_1^\top(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + o_p(n^{-1/2}).$$

With the natural constraint $\sum_{i=0}^{1}\sum_{j=1}^{n_i}\hat{p}_{ij}\omega(X_{ij};\hat{\boldsymbol{\theta}}) = 1$, the above approximation equation implies

$$\hat{\psi}_1 = \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\hat{p}_{ij}\omega(X_{ij};\hat{\boldsymbol{\theta}})\mathcal{H}_1(X_{ij}) + o_p(n^{-1/2}), \tag{4.44}$$

where we define $\mathcal{H}_1(a) = 2H_1(a) - \psi_1$ for $a > 0$.

- *Asymptotic properties of $\hat{\mathcal{G}}_0$ and $\hat{\mathcal{G}}_1$*

  We now proceed to derive the asymptotic properties of $\hat{\mathcal{G}}_0$ and $\hat{\mathcal{G}}_1$. Recall that

$$\hat{\mathcal{G}}_0 = (2\hat{\nu}_0 - 1) + (1-\hat{\nu}_0)\frac{\hat{\psi}_0}{\hat{m}_0},$$

where $\hat{m}_0 = \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\hat{p}_{ij}X_{ij}$ and the approximation of $\hat{\psi}_0$ is in (4.42). Then

$$\begin{aligned}\hat{\mathcal{G}}_0 &= \frac{(2\hat{\nu}_0-1)\hat{m}_0 + (1-\hat{\nu}_0)\hat{\psi}_0}{\hat{m}_0} \\ &= \frac{\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\hat{p}_{ij}\left\{(2\hat{\nu}_0-1)X_{ij} + (1-\hat{\nu}_0)\mathcal{H}_0(X_{ij})\right\}}{\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\hat{p}_{ij}X_{ij}} + o_p(n^{-1/2}). \end{aligned} \tag{4.45}$$

Similarly,

$$\hat{\mathcal{G}}_1 = \frac{\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \hat{p}_{ij}\left\{(2\hat{\nu}_1 - 1)\omega(X_{ij};\hat{\boldsymbol{\theta}})X_{ij} + (1-\hat{\nu}_1)\mathcal{H}_1(X_{ij};\hat{\boldsymbol{\theta}})\right\}}{\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \hat{p}_{ij}\omega(X_{ij};\hat{\boldsymbol{\theta}})X_{ij}} + o_p(n^{-1/2}). \quad (4.46)$$

Note that the numerators and denominators of the leading terms in (4.45) and (4.46) all have the forms in (4.11) with $\boldsymbol{u}(\cdot;\cdot)$ taking some specific forms. We define these specific $\boldsymbol{u}(x;\boldsymbol{\nu},\boldsymbol{\theta})$ as

$$\boldsymbol{u}(x;\boldsymbol{\nu},\boldsymbol{\theta}) \quad = \quad (x, u_0(x;\boldsymbol{\nu}), \omega(x;\boldsymbol{\theta})x, \omega(x;\boldsymbol{\theta})u_1(x;\boldsymbol{\nu}))^\top \quad (4.47)$$

with

$$u_0(x;\boldsymbol{\nu}) = (2\nu_0 - 1)x + (1-\nu_0)\mathcal{H}_0(x) \quad \text{and} \quad u_1(x;\boldsymbol{\nu}) = (2\nu_1 - 1)x + (1-\nu_1)\mathcal{H}_1(x). (4.48)$$

Further, we define

$$\boldsymbol{\gamma} = \int_0^\infty \boldsymbol{u}(x;\boldsymbol{\nu},\boldsymbol{\theta})dG_0(x) = (\gamma_1,\gamma_2,\gamma_3,\gamma_4)^\top \quad (4.49)$$

and

$$\hat{\boldsymbol{\gamma}} = \int_0^\infty \boldsymbol{u}(x;\hat{\boldsymbol{\nu}},\hat{\boldsymbol{\theta}})d\hat{G}_0(x) = (\hat{\gamma}_1,\hat{\gamma}_2,\hat{\gamma}_3,\hat{\gamma}_4)^\top. \quad (4.50)$$

Then we have

$$\hat{\mathcal{G}}_0 = \hat{\gamma}_1/\hat{\gamma}_2 + o_p(n^{-1/2}) \quad \text{and} \quad \hat{\mathcal{G}}_1 = \hat{\gamma}_3/\hat{\gamma}_4 + o_p(n^{-1/2}). \quad (4.51)$$

Hence, the joint limiting distribution of $\sqrt{n}(\hat{\mathcal{G}}_0 - \mathcal{G}_0^*, \hat{\mathcal{G}}_1 - \mathcal{G}_1^*)$ is determined by that of $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*)$, where the $\mathcal{G}_i^*$ are the true values of $\mathcal{G}_i$ for $i = 0, 1$, and

$$\boldsymbol{\gamma}^* \quad = \quad \int_0^\infty \boldsymbol{u}(x;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)dG_0(x) = (m_0, m_0\mathcal{G}_0^*, m_1, m_1\mathcal{G}_1^*)^\top$$

is the true value of $\boldsymbol{\gamma}$.

Let

$$\tilde{\boldsymbol{u}}(x) = (\tilde{\boldsymbol{u}}_0(x)^\top, \tilde{\boldsymbol{u}}_1(x)^\top)^\top = (-\rho^*(x, u_0(x;\boldsymbol{\nu}^*)), (1-\rho^*)(x, u_1(x;\boldsymbol{\nu}^*)))^\top.$$

By using the form of $\boldsymbol{u}(x;\boldsymbol{\nu},\boldsymbol{\theta})$ in (4.47), we obtain the simplified forms of $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ in Theorem 4.2 as follow:

$$\mathcal{M}_1 = \begin{pmatrix} 0 & 0 \\ 2m_0 - \psi_0 & 0 \\ 0 & 0 \\ 0 & 2m_1 - \psi_1 \end{pmatrix}, \quad \mathcal{M}_2 = \begin{pmatrix} -\rho^* m_0 \\ -\rho^*(m_0 \mathcal{G}_0^*) \\ (1-\rho^*)m_1 \\ (1-\rho^*)(m_1\mathcal{G}_1^*) \end{pmatrix} = E_0\{\tilde{\boldsymbol{u}}(X)\},$$

$$\mathcal{M}_3 = \begin{pmatrix} -E_0\{h_1(X)X\boldsymbol{Q}(X)^\top\} \\ -E_0\{h_1(X)u_0(X;\boldsymbol{\nu}^*)\boldsymbol{Q}(X)^\top\} \\ E_0\{h_0(X)\omega(X)X\boldsymbol{Q}(X)^\top\} \\ E_0\{h_0(X)\omega(X)u_1(X;\boldsymbol{\nu}^*)\boldsymbol{Q}(X)^\top\} \end{pmatrix} = \frac{1}{\rho^*}E_0\{h_1(X)\tilde{\boldsymbol{u}}(X)\boldsymbol{Q}(X)^\top\}.$$

Let $\boldsymbol{g}(\boldsymbol{\gamma}) = (\gamma_2/\gamma_1, \gamma_4/\gamma_3)^\top = (\mathcal{G}_0, \mathcal{G}_1)^\top$. Then $\boldsymbol{g}(\hat{\boldsymbol{\gamma}}) = (\hat{\mathcal{G}}_0, \hat{\mathcal{G}}_1)^\top$ and $\boldsymbol{g}(\boldsymbol{\gamma}^*) = (\mathcal{G}_0^*, \mathcal{G}_1^*)^\top$. Applying the Delta method to the limiting distribution in Theorem 4.2, we have, as $n \to \infty$,

$$\sqrt{n}\left(\hat{\mathcal{G}}_0 - \mathcal{G}_0^*\right) \to N(\mathbf{0}, \boldsymbol{\Sigma})$$

in distribution, where $\boldsymbol{\Sigma} = \boldsymbol{J}\boldsymbol{\Gamma}\boldsymbol{J}^\top$ and

$$\boldsymbol{J} = \frac{\partial \boldsymbol{g}(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{\gamma}} = \begin{pmatrix} \boldsymbol{J}_0^\top & \mathbf{0} \\ \mathbf{0} & \boldsymbol{J}_1^\top \end{pmatrix} = \begin{pmatrix} -\frac{\mathcal{G}_0^*}{m_0} & \frac{1}{m_0} & 0 & 0 \\ 0 & 0 & -\frac{\mathcal{G}_1^*}{m_1} & \frac{1}{m_1} \end{pmatrix}. \tag{4.52}$$

To finish the proof of Theorem 4.3, we use the forms of $\boldsymbol{\Gamma}$ in (4.12) and $\boldsymbol{J}$ in (4.52) to simplify $\boldsymbol{\Sigma}$. Note that

$$\boldsymbol{J}_0^\top \begin{pmatrix} m_0 \\ m_0\mathcal{G}_0^* \end{pmatrix} = -\frac{\mathcal{G}_0^*}{m_0}\cdot m_0 + \frac{1}{m_0}\cdot(m_0\mathcal{G}_0^*) = 0,$$
$$\boldsymbol{J}_1^\top \begin{pmatrix} m_1 \\ m_1\mathcal{G}_1^* \end{pmatrix} = -\frac{\mathcal{G}_1^*}{m_1}\cdot m_1 + \frac{1}{m_1}\cdot(m_1\mathcal{G}_1^*) = 0.$$

This leads to

$$\boldsymbol{J}\left\{\frac{\boldsymbol{\gamma}^*\boldsymbol{\gamma}^{*\top}}{\Delta^*}\right\}\boldsymbol{J}^\top = \boldsymbol{J}\left\{-\frac{\mathcal{M}_2\mathcal{M}_2^\top}{\Delta^*\rho^*(1-\rho^*)}\right\}\boldsymbol{J}^\top = \mathbf{0}. \tag{4.53}$$

With the fact that

$$\boldsymbol{J}_0^\top \begin{pmatrix} 0 \\ 2m_0 - \psi_0 \end{pmatrix} = \frac{1-\mathcal{G}_0^*}{1-\nu_0} \quad \text{and} \quad \boldsymbol{J}_1^\top \begin{pmatrix} 0 \\ 2m_1 - \psi_1 \end{pmatrix} = \frac{1-\mathcal{G}_1^*}{1-\nu_1},$$

we have

$$\boldsymbol{J}\mathcal{M}_1 = \operatorname{diag}\left\{\frac{1-\mathcal{G}_0^*}{1-\nu_0}, \frac{1-\mathcal{G}_1^*}{1-\nu_1}\right\}.$$

Hence,

$$\boldsymbol{J}(\mathcal{M}_1\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\mathcal{M}_1)\boldsymbol{J}^\top = \operatorname{diag}\left\{\frac{\nu_0^*(1-\mathcal{G}_0^*)^2}{\Delta^*(1-\rho^*)}, \frac{\nu_1^*(1-\mathcal{G}_1^*)^2}{\Delta^*\rho^*}\right\}. \tag{4.54}$$

Substituting (4.12) and (4.52) into $\boldsymbol{\Sigma}$ and using (4.53)–(4.54), $\boldsymbol{\Sigma}$ has the following simplified form:

$$\begin{aligned}\boldsymbol{\Sigma} &= \frac{1}{\Delta^*}\boldsymbol{J}\left[E_0\left\{\frac{\boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)\boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)^\top}{h(X)}\right\} + \frac{1}{(\rho^*)^2}\boldsymbol{B}\right]\boldsymbol{J}^\top \\ &\quad + \operatorname{diag}\left\{\frac{\nu_0^*(1-\mathcal{G}_0^*)^2}{\Delta^*(1-\rho^*)}, \frac{\nu_1^*(1-\mathcal{G}_1^*)^2}{\Delta^*\rho^*}\right\},\end{aligned} \tag{4.55}$$

where

$$\boldsymbol{B} = E_0\{h_1(X)\tilde{\boldsymbol{u}}(X)\boldsymbol{Q}(X)^\top\}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}E_0\{h_1(X)\boldsymbol{Q}(X)\tilde{\boldsymbol{u}}(X)^\top\},$$

as claimed in Theorem 4.3. This completes the proof.

### Proof of Theorem 4.4

The proof of Theorem 4.4 is similar to that of Theorem 1. The results of Li et al. (2018) are helpful for this proof.

### Proof of Theorem 4.5

We start with (a). Recall that the nonparametric estimator of the Gini index for sample $i = 0, 1$ is defined as

$$\tilde{\mathcal{G}}_i = (2\hat{\nu}_i - 1) + (1-\hat{\nu}_i)\tilde{\psi}_i/\tilde{m}_i,$$

where

$$\tilde{m}_i = {n_{i1}}^{-1}\sum_{j=1}^{n_{i1}} X_{ij} \quad \text{and} \quad \tilde{\psi}_i = \int_0^\infty \{2x\tilde{G}_i(x)\}d\tilde{G}_i(x).$$

After some algebra, we have

$$\tilde{\mathcal{G}}_i = \frac{n_i^{-1}\sum_{j=1}^{n_i}\{2\tilde{F}_i(X_{ij})-1\}X_{ij}}{\tilde{\mu}_i}, \tag{4.56}$$

where $\tilde{\mu}_i = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}$ and $\tilde{F}_i(x) = n_i^{-1} \sum_{j=1}^{n_i} I(X_{ij} \leq x)$ are the sample mean and the empirical CDF based on sample $i$.

Applying Theorem 1 of Qin et al. (2010), we have

$$\sqrt{n} \begin{pmatrix} \tilde{\mathcal{G}}_0 - \mathcal{G}_0^* \\ \tilde{\mathcal{G}}_1 - \mathcal{G}_1^* \end{pmatrix} \to N(\mathbf{0}, \mathbf{\Sigma}_{non})$$

in distribution with

$$\mathbf{\Sigma}_{non} = \begin{pmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}, \qquad (4.57)$$

where

$$\sigma_i^2 = \frac{Var_{F_i}\{u_i(X) - \mathcal{G}_i^* X\}}{w_i \mu_i^2}, \qquad (4.58)$$

where $Var_{F_i}$ means the variance is taken with respect to $F_i$ and $u_i(x) = u_i(x; \boldsymbol{\nu}^*)$ with $u_i(x; \boldsymbol{\nu})$ defined in (4.48).

We now show that $\mathbf{\Sigma}_{non}$ has the form claimed in (a). Note that

$$Var_{F_i}\{u_i(X) - \mathcal{G}_i^* X\} = E_{F_i}[\{u_i(X) - \mathcal{G}_i^* X\}^2] - [E_{F_i}\{u_i(X) - \mathcal{G}_i^* X\}]^2, \quad (4.59)$$

where $E_{F_i}$ indicates that the expectation is taken with respect to $F_i$. After some calculus work, we can show that

$$E_{F_i}\{u_i(X) - \mathcal{G}_i^* X\} = \nu_i^*(1 - \nu_i^*)(2m_i - \psi_i). \qquad (4.60)$$

For $E_{F_i}[\{u_i(X) - \mathcal{G}_i^* X\}^2]$, we have under model (4.1)

$$E_{F_i}[\{u_i(X) - \mathcal{G}_i^* X\}^2] = \nu_i^* u_i^2(0) + (1 - \nu_i^*) E_{G_i}[\{u_i(X) - \mathcal{G}_i^* X\}^2],$$

where $E_{G_i}$ indicates that the expectation is taken with respect to $G_i$. Then

$$\begin{aligned} & E_{F_i}[\{u_i(X) - \mathcal{G}_i^* X\}^2] \\ = \ & \nu_i^* u_i^2(0) + (1 - \nu_i^*) E_{G_i}\{u_i^2(X) - 2\mathcal{G}_i^* X u_i(X) + \mathcal{G}_i^{*2} X^2\}. \end{aligned} \qquad (4.61)$$

With the form of $u_i(x; \boldsymbol{\nu})$ in (4.48), we have

$$u_i(0) = u_i(0; \boldsymbol{\nu}^*) = 2(1 - \nu_i^*) m_i - (1 - \nu_i^*) \psi_i. \qquad (4.62)$$

Combining (4.59)–(4.62) gives

$$\begin{aligned} & Var_{F_i}\{u_i(X) - \mathcal{G}_i X\} \\ = \; & \nu_i^*(1-\nu_i^*)^3(2m_i - \psi_i)^2 + (1-\nu_i^*)E_{G_i}\{u_i^2(X) - 2\mathcal{G}_i^* X u_i(X) + \mathcal{G}_i^{*2}X^2\}. \quad (4.63) \end{aligned}$$

The fact that $\mu_i = (1-\nu_i^*)m_i$ and (4.63) together imply that $\sigma_i^2$ in (4.58) has the following form:

$$\begin{aligned} \sigma_i^2 &= \frac{\nu_i^*(1-\nu_i^*)^3(2m_i-\psi_i)^2 + (1-\nu_i^*)E_{G_i}\{u_i^2(X) - 2\mathcal{G}_i^* X u_i(X) + \mathcal{G}_i^{*2}X^2\}}{w_i(1-\nu_i^*)^2 m_i^2} \\ &= \frac{\nu_i^*(1-\nu_i^*)^3(2m_i-\psi_i)^2 + (1-\nu_i^*)E_{G_i}\{u_i^2(X) - 2\mathcal{G}_i^* X u_i(X) + \mathcal{G}_i^{*2}X^2\}}{w_i(1-\nu_i^*)^2 m_i^2} \\ &= \frac{E_{G_i}\{u_i^2(X) - 2\mathcal{G}_i^* X u_i(X) + \mathcal{G}_i^{*2}X^2\}}{w_i(1-\nu_i^*)m_i^2} + \frac{\nu_i^*(1-\mathcal{G}_i^*)^2}{w_i(1-\nu_i^*)}, \end{aligned}$$

where in the last step, we have used the fact that $\mathcal{G}_i^* = (2\nu_i^* - 1) + (1-\nu_i^*)\psi_i/m_i$.

Under the DRM (4.3) and since $\Delta^*\rho^* = w_1(1-\nu_1^*)$ and $\Delta^*(1-\rho^*) = w_0(1-\nu_0^*)$, we further have

$$\sigma_0^2 = \frac{E_0\{u_0^2(X) - 2\mathcal{G}_0^* X u_0(X) + \mathcal{G}_0^{*2}X^2\}}{\Delta^*(1-\rho^*)m_0^2} + \frac{\nu_0^*(1-\mathcal{G}_0^*)^2}{\Delta^*(1-\rho^*)}$$

and

$$\sigma_1^2 = \frac{E_0[\omega(X)\{u_1^2(X) - 2\mathcal{G}_1^* X u_1(X) + \mathcal{G}_1^{*2}X^2\}]}{\Delta^*\rho^* m_1^2} + \frac{\nu_1^*(1-\mathcal{G}_1^*)^2}{\Delta^*\rho^*}.$$

Recall that

$$\boldsymbol{J}_0 = \left(-\frac{\mathcal{G}_0^*}{m_0}, \frac{1}{m_0}\right)^\top, \quad \boldsymbol{J}_1 = \left(-\frac{\mathcal{G}_1^*}{m_1}, \frac{1}{m_1}\right)^\top$$

and

$$\tilde{\boldsymbol{u}}_0(x) = -\rho^*\big(x, u_0(x)\big)^\top, \quad \tilde{\boldsymbol{u}}_1(x) = (1-\rho^*)\big(x, u_1(x)\big)^\top.$$

After some algebra work, we get

$$\sigma_0^2 = \frac{1}{\Delta^*(\rho^*)^2(1-\rho^*)}\boldsymbol{J}_0^\top E_0\{\tilde{\boldsymbol{u}}_0(X)\tilde{\boldsymbol{u}}_0(X)^\top\}\boldsymbol{J}_0 + \frac{\nu_0^*(\mathcal{G}_0^*-1)^2}{\Delta^*(1-\rho^*)} \quad (4.64)$$

and

$$\sigma_1^2 = \frac{1}{\Delta^*\rho^*(1-\rho^*)^2}\boldsymbol{J}_1^\top E_0\{\omega(X)\tilde{\boldsymbol{u}}_1(X)\tilde{\boldsymbol{u}}_1(X)^\top\}\boldsymbol{J}_1 + \frac{\nu_1^*(\mathcal{G}_1^*-1)^2}{\Delta^*\rho^*}. \tag{4.65}$$

Substituting (4.64) and (4.65) into (4.57) gives the asymptotic variance $\boldsymbol{\Sigma}_{non}$ as

$$\boldsymbol{\Sigma}_{non} = \boldsymbol{J}\boldsymbol{\Sigma}_{np1}\boldsymbol{J}^\top + \mathrm{diag}\left\{\frac{\nu_0^*(1-\mathcal{G}_0^*)^2}{\Delta^*(1-\rho^*)}, \frac{\nu_1^*(1-\mathcal{G}_1^*)^2}{\Delta^*\rho^*}\right\}, \tag{4.66}$$

with

$$\boldsymbol{\Sigma}_{np1} = \frac{1}{\Delta^*\rho^*(1-\rho)}\mathrm{diag}\left\{\frac{E_0\{\tilde{\boldsymbol{u}}_0(X)\tilde{\boldsymbol{u}}_0(X)^\top\}}{\rho^*}, \frac{E_0\{\omega(X)\tilde{\boldsymbol{u}}_1(X)\tilde{\boldsymbol{u}}_1(X)^\top\}}{1-\rho^*}\right\}.$$

Hence, $\boldsymbol{\Sigma}_{non}$ has the form claimed in (a).

We now move to (b). Since $\boldsymbol{u}(x;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*) = (-(\rho^*)^{-1}\tilde{\boldsymbol{u}}_0(x)^\top, (1-\rho^*)^{-1}\tilde{\boldsymbol{u}}_1(x)^\top)^\top$, after some algebra, we find that

$$\begin{aligned}
& \frac{1}{\Delta^*}E_0\left\{\frac{\boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)\boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)^\top}{h(X)}\right\} \\
= & \frac{1}{\Delta^*(\rho^*)^2(1-\rho^*)}\begin{pmatrix} E_0\{h_0(X)\tilde{\boldsymbol{u}}_0(X)\tilde{\boldsymbol{u}}_0(X)^\top\} & -E_0\{h_1\tilde{\boldsymbol{u}}_0(X)\tilde{\boldsymbol{u}}_1(X)^\top\} \\ -E_0\{h_1(X)\tilde{\boldsymbol{u}}_1(X)\tilde{\boldsymbol{u}}_0(X)^\top\} & \frac{\rho^*}{1-\rho^*}E_0\{h_1(X)\omega(X)\tilde{\boldsymbol{u}}_1(X)\tilde{\boldsymbol{u}}_1(X)^\top\}\end{pmatrix} \\
= & \boldsymbol{\Sigma}_{np1} - \frac{1}{\Delta^*(\rho^*)^2(1-\rho)}E_0\{h_1(X)\tilde{\boldsymbol{u}}(X)\tilde{\boldsymbol{u}}(X)^\top\}.
\end{aligned}$$

Together with the expression for $\boldsymbol{\Sigma}$ in (4.55), it follows that

$$\boldsymbol{\Sigma}_{non} - \boldsymbol{\Sigma} = \frac{1}{\Delta^*(\rho^*)^2(1-\rho^*)}\boldsymbol{J}\left[E_0\{h_1(X)\tilde{\boldsymbol{u}}(X)\tilde{\boldsymbol{u}}(X)^\top\} - \Delta^*(1-\rho^*)\boldsymbol{B}\right]\boldsymbol{J}^\top,$$

where

$$\boldsymbol{B} = E_0\{h_1(X)\tilde{\boldsymbol{u}}(X)\boldsymbol{Q}(X)^\top\}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}E_0\{h_1(X)\boldsymbol{Q}(X)\tilde{\boldsymbol{u}}(X)^\top\}. \tag{4.67}$$

Let $\boldsymbol{D}(a) = \left(\boldsymbol{D}_0(a)^\top, \boldsymbol{D}_1(a)^\top\right)^\top$ for $a > 0$ with

$$\boldsymbol{D}_i(a) = \tilde{\boldsymbol{u}}_i(x) - \Delta^*(1-\rho^*)E_0\left\{h_1(X)\tilde{\boldsymbol{u}}_i(X)\boldsymbol{Q}(X)^\top\right\}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{Q}(a),\ i = 0, 1.$$

Recall that

$$\boldsymbol{A}_{\boldsymbol{\theta}} = \Delta^*(1-\rho^*)E_0\left[h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}^\top(X)\right].$$

It can be verified that for $i, j \in \{0, 1\}$,

$$\begin{aligned} & E_0\{h_1(X)\boldsymbol{D}_i(X)\boldsymbol{D}_j(X)^\top\} \\ = \; & E_0\{h_1(X)\tilde{\boldsymbol{u}}_i(X)\tilde{\boldsymbol{u}}_j(X)^\top\} \\ & -\Delta^*(1-\rho^*)E_0\{h_1(X)\tilde{\boldsymbol{u}}_i(X)\boldsymbol{Q}(X)^\top\}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}E_0\{h_1(X)\boldsymbol{Q}(X)\tilde{\boldsymbol{u}}_j(X)^\top\}. \end{aligned}$$

Recall that

$$\tilde{\boldsymbol{u}}(X) = (\tilde{\boldsymbol{u}}_0(X)^\top, \tilde{\boldsymbol{u}}_1(X)^\top)^\top$$

and $\boldsymbol{B}$ is given in (4.67). Then,

$$\boldsymbol{\Sigma}_{non} - \boldsymbol{\Sigma} = \frac{1}{\Delta^*(\rho^*)^2(1-\rho^*)}\boldsymbol{J}E_0\{h_1(X)\boldsymbol{D}(X)\boldsymbol{D}(X)^\top\}\boldsymbol{J}^\top,$$

as claimed in (b). This completes the proof.

**Proof of Theorem 4.6**

The result in Theorem 4.6 is a direct consequence of applying the Delta method and the results in Theorems 4.3 and 4.4.

### 4.5.2 Additional simulation results

**Results for point estimator**

Tables 4.13 and 4.14 present the additional simulated results for the point estimators of the Gini indices $\mathcal{G}_0$, $\mathcal{G}_1$, and their difference $\mathcal{G}_0 - \mathcal{G}_1$ under different distributional settings. The general trends are similar to those in Section 4.3.1. The DRM method always gives the smallest MSEs.

**Results for CIs**

Tables 4.15 and 4.16 contain the the complete results for the CIs of $\mathcal{G}_0$ and $\mathcal{G}_1$ under different distributional settings. NL-DRM and BL-DRM refer to the Wald-type CIs for $\mathcal{G}_0$ or $\mathcal{G}_1$ using the logit transformation under the DRM and the corresponding bootstrap-t

Table 4.13: Bias (×1000) and MSE (×1000) for point estimators ($\chi^2$ distributions).

| | | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | | $\mathcal{G}_0 - \mathcal{G}_1$ | |
|---|---|---|---|---|---|---|---|---|
| $(n_0, n_1)$ | $\boldsymbol{\nu}$ | | Bias | MSE | Bias | MSE | Bias | MSE |
| (100,100) | (0.1,0.3) | EMP | 6.13 | 0.98 | 7.43 | 1.30 | -1.30 | 2.23 |
| | | JEL | 0.96 | 0.96 | 3.08 | 1.28 | -2.13 | 2.28 |
| | | DRM | 2.51 | 0.67 | 3.79 | 1.10 | -1.28 | 1.40 |
| | (0.6,0.4) | EMP | 7.15 | 1.14 | 6.00 | 1.27 | 1.15 | 2.44 |
| | | JEL | 4.90 | 1.14 | 2.27 | 1.27 | 2.63 | 2.49 |
| | | DRM | 2.71 | 0.94 | 3.48 | 1.18 | -0.77 | 1.98 |
| (300,300) | (0.1,0.3) | EMP | 1.40 | 0.31 | 2.96 | 0.41 | -1.56 | 0.72 |
| | | JEL | -0.33 | 0.31 | 1.51 | 0.40 | -1.83 | 0.72 |
| | | DRM | 0.75 | 0.23 | 1.38 | 0.35 | -0.63 | 0.46 |
| | (0.6,0.4) | EMP | 2.63 | 0.38 | 1.58 | 0.43 | 1.05 | 0.80 |
| | | JEL | 1.87 | 0.38 | 0.33 | 0.43 | 1.54 | 0.81 |
| | | DRM | 1.02 | 0.32 | 0.78 | 0.41 | 0.25 | 0.66 |

Table 4.14: Bias (×1000) and MSE (×1000) for point estimators (exponential distributions).

| | | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | | $\mathcal{G}_0 - \mathcal{G}_1$ | |
|---|---|---|---|---|---|---|---|---|
| $(n_0, n_1)$ | $\boldsymbol{\nu}$ | | Bias | MSE | Bias | MSE | Bias | MSE |
| (100,100) | (0.1,0.3) | EMP | 4.83 | 0.99 | 4.12 | 1.08 | 0.70 | 1.95 |
| | | JEL | 0.33 | 0.99 | 0.63 | 1.09 | -0.30 | 1.98 |
| | | DRM | 1.63 | 0.88 | 1.61 | 0.77 | 0.02 | 1.22 |
| | (0.6,0.4) | EMP | 6.12 | 0.95 | 3.94 | 1.12 | 2.19 | 2.04 |
| | | JEL | 4.17 | 0.95 | 0.95 | 1.13 | 3.22 | 2.09 |
| | | DRM | 2.34 | 0.83 | 2.09 | 0.92 | 0.25 | 1.51 |
| (300,300) | (0.1,0.3) | EMP | 1.52 | 0.33 | 2.26 | 0.37 | -0.75 | 0.66 |
| | | JEL | 0.02 | 0.33 | 1.10 | 0.36 | -1.08 | 0.67 |
| | | DRM | 0.70 | 0.29 | 0.96 | 0.25 | -0.25 | 0.40 |
| | (0.6,0.4) | EMP | 2.03 | 0.31 | 1.26 | 0.38 | 0.77 | 0.69 |
| | | JEL | 1.37 | 0.31 | 0.27 | 0.38 | 1.11 | 0.70 |
| | | DRM | 0.83 | 0.29 | 0.45 | 0.30 | 0.39 | 0.51 |

CIs. The additional results for the CIs of $\mathcal{G}_0 - \mathcal{G}_1$ are shown in Table 4.17. Again, the general patterns are similar to those in 4.3.2. The NA-DRM CIs provide accurate CPs in all situations and have shorter ALs than the existing nonparametric methods. Further, the bootstrap method and logit transformation do not help to improve the coverage accuracy. Hence, we recommend using the NA-DRM CI.

Table 4.15: CP(%) and AL of CIs ($\chi^2$ distributions).

| | | (100,100) | | | | (300,300) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | |
| $\boldsymbol{\nu}$ | | CP | AL | CP | AL | CP | AL | CP | AL |
| (0,0) | NA-EMP | 93.85 | 0.100 | 94.20 | 0.092 | 94.60 | 0.059 | 94.80 | 0.054 |
| | BT-EMP | 94.10 | 0.103 | 94.75 | 0.094 | 94.85 | 0.059 | 95.05 | 0.054 |
| | EL | 93.85 | 0.100 | 94.20 | 0.091 | 94.55 | 0.059 | 94.80 | 0.054 |
| | BT-EL | 94.45 | 0.103 | 95.10 | 0.095 | 94.90 | 0.059 | 94.95 | 0.054 |
| | JEL | 94.45 | 0.102 | 94.85 | 0.094 | 94.70 | 0.059 | 95.15 | 0.054 |
| | AJEL | 94.80 | 0.105 | 95.50 | 0.096 | 94.90 | 0.060 | 95.30 | 0.055 |
| | NA-DRM | 95.25 | 0.074 | 94.65 | 0.078 | 94.70 | 0.043 | 94.70 | 0.045 |
| | BT-DRM | 95.55 | 0.075 | 95.00 | 0.079 | 94.55 | 0.043 | 94.55 | 0.046 |
| | NL-DRM | 95.35 | 0.074 | 94.50 | 0.077 | 94.75 | 0.043 | 94.75 | 0.045 |
| | BL-DRM | 95.45 | 0.075 | 94.80 | 0.079 | 94.50 | 0.043 | 94.55 | 0.045 |
| (0.1,0.3) | NA-EMP | 94.00 | 0.116 | 93.85 | 0.134 | 95.00 | 0.068 | 95.10 | 0.079 |
| | BT-EMP | 94.80 | 0.119 | 95.05 | 0.137 | 95.35 | 0.068 | 95.10 | 0.079 |
| | EL | 93.90 | 0.116 | 93.95 | 0.133 | 95.00 | 0.068 | 95.05 | 0.078 |
| | BT-EL | 95.25 | 0.119 | 94.65 | 0.139 | 95.20 | 0.068 | 95.10 | 0.080 |
| | JEL | 94.70 | 0.120 | 94.00 | 0.140 | 95.25 | 0.069 | 94.65 | 0.080 |
| | AJEL | 95.25 | 0.123 | 94.60 | 0.144 | 95.40 | 0.069 | 94.80 | 0.081 |
| | NA-DRM | 93.60 | 0.099 | 94.80 | 0.128 | 94.60 | 0.058 | 95.25 | 0.075 |
| | BT-DRM | 93.95 | 0.100 | 95.25 | 0.129 | 94.55 | 0.058 | 94.95 | 0.074 |
| | NL-DRM | 93.85 | 0.099 | 95.00 | 0.128 | 94.65 | 0.058 | 95.20 | 0.075 |
| | BL-DRM | 93.65 | 0.099 | 94.95 | 0.127 | 94.55 | 0.058 | 94.85 | 0.074 |
| (0.3,0.3) | NA-EMP | 93.80 | 0.132 | 93.65 | 0.134 | 94.60 | 0.077 | 94.05 | 0.079 |
| | BT-EMP | 95.30 | 0.135 | 94.55 | 0.137 | 95.20 | 0.077 | 94.40 | 0.079 |
| | EL | 93.75 | 0.131 | 93.65 | 0.134 | 94.60 | 0.077 | 94.00 | 0.078 |
| | BT-EL | 94.50 | 0.136 | 94.85 | 0.139 | 94.65 | 0.078 | 94.55 | 0.079 |
| | JEL | 94.45 | 0.137 | 93.80 | 0.141 | 94.50 | 0.078 | 94.55 | 0.080 |
| | AJEL | 95.35 | 0.141 | 94.20 | 0.144 | 94.80 | 0.079 | 94.80 | 0.081 |
| | NA-DRM | 95.10 | 0.120 | 94.35 | 0.130 | 95.45 | 0.070 | 94.90 | 0.076 |
| | BT-DRM | 95.75 | 0.121 | 94.65 | 0.130 | 95.25 | 0.070 | 94.65 | 0.075 |
| | NL-DRM | 95.60 | 0.120 | 94.70 | 0.129 | 95.50 | 0.070 | 95.00 | 0.076 |
| | BL-DRM | 95.30 | 0.119 | 94.60 | 0.128 | 95.10 | 0.069 | 94.60 | 0.075 |
| (0.6,0.4) | NA-EMP | 93.45 | 0.124 | 94.10 | 0.138 | 94.30 | 0.073 | 95.10 | 0.080 |
| | BT-EMP | 95.85 | 0.131 | 95.00 | 0.142 | 95.55 | 0.074 | 95.10 | 0.081 |
| | EL | 94.00 | 0.123 | 94.10 | 0.137 | 94.45 | 0.073 | 95.05 | 0.080 |
| | BT-EL | 95.35 | 0.130 | 94.95 | 0.143 | 94.90 | 0.075 | 95.30 | 0.082 |
| | JEL | 92.90 | 0.133 | 94.15 | 0.145 | 93.35 | 0.075 | 94.90 | 0.082 |
| | AJEL | 93.40 | 0.137 | 94.90 | 0.149 | 93.60 | 0.075 | 95.05 | 0.083 |
| | NA-DRM | 94.60 | 0.119 | 95.05 | 0.137 | 95.30 | 0.069 | 95.15 | 0.080 |
| | BT-DRM | 94.95 | 0.120 | 95.45 | 0.137 | 95.55 | 0.069 | 94.95 | 0.078 |
| | NL-DRM | 95.00 | 0.119 | 95.25 | 0.136 | 95.60 | 0.069 | 95.15 | 0.080 |
| | BL-DRM | 94.45 | 0.116 | 94.85 | 0.134 | 95.10 | 0.068 | 94.80 | 0.078 |
| (0.7,0.7) | NA-EMP | 92.20 | 0.113 | 92.95 | 0.119 | 94.90 | 0.067 | 93.90 | 0.070 |
| | BT-EMP | 96.75 | 0.122 | 96.55 | 0.128 | 96.30 | 0.068 | 95.40 | 0.072 |
| | EL | 92.35 | 0.111 | 92.90 | 0.117 | 95.15 | 0.067 | 93.75 | 0.070 |
| | BT-EL | 94.70 | 0.120 | 95.30 | 0.127 | 95.75 | 0.069 | 94.55 | 0.072 |
| | JEL | 90.75 | 0.123 | 90.80 | 0.129 | 94.00 | 0.069 | 93.00 | 0.072 |
| | AJEL | 91.35 | 0.127 | 91.55 | 0.133 | 94.25 | 0.070 | 93.10 | 0.073 |
| | NA-DRM | 94.50 | 0.111 | 94.85 | 0.121 | 95.10 | 0.065 | 95.20 | 0.071 |
| | BT-DRM | 95.40 | 0.113 | 95.90 | 0.123 | 95.45 | 0.064 | 95.60 | 0.070 |
| | NL-DRM | 95.70 | 0.111 | 96.05 | 0.121 | 96.05 | 0.065 | 96.10 | 0.071 |
| | BL-DRM | 94.60 | 0.108 | 95.05 | 0.118 | 95.30 | 0.063 | 95.20 | 0.069 |

Table 4.16: CP(%) and AL of CIs (exponential distributions).

| | | (100,100) | | | | (300,300) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | | $\mathcal{G}_0$ | | $\mathcal{G}_1$ | |
| $\boldsymbol{\nu}$ | | CP | AL | CP | AL | CP | AL | CP | AL |
| (0,0) | NA-EMP | 93.85 | 0.110 | 93.50 | 0.111 | 94.65 | 0.065 | 94.45 | 0.065 |
| | BT-EMP | 94.35 | 0.115 | 94.05 | 0.115 | 94.75 | 0.065 | 94.75 | 0.065 |
| | EL | 93.90 | 0.110 | 93.50 | 0.110 | 94.65 | 0.065 | 94.55 | 0.065 |
| | BT-EL | 94.50 | 0.113 | 94.00 | 0.113 | 94.80 | 0.065 | 94.60 | 0.065 |
| | JEL | 94.35 | 0.113 | 93.90 | 0.113 | 94.90 | 0.065 | 94.55 | 0.065 |
| | AJEL | 94.95 | 0.115 | 94.35 | 0.116 | 95.10 | 0.066 | 94.75 | 0.066 |
| | NA-DRM | 94.80 | 0.100 | 94.05 | 0.079 | 93.95 | 0.059 | 95.20 | 0.045 |
| | BT-DRM | 94.45 | 0.104 | 94.75 | 0.079 | 93.65 | 0.060 | 94.95 | 0.045 |
| | NL-DRM | 94.90 | 0.100 | 94.10 | 0.078 | 93.95 | 0.059 | 95.25 | 0.045 |
| | BL-DRM | 94.20 | 0.103 | 94.50 | 0.079 | 93.55 | 0.059 | 94.85 | 0.045 |
| (0.1,0.3) | NA-EMP | 93.45 | 0.119 | 94.30 | 0.128 | 95.25 | 0.070 | 94.50 | 0.075 |
| | BT-EMP | 94.25 | 0.123 | 95.50 | 0.132 | 95.00 | 0.071 | 94.80 | 0.075 |
| | EL | 93.45 | 0.119 | 94.45 | 0.127 | 95.25 | 0.070 | 94.40 | 0.075 |
| | BT-EL | 94.40 | 0.122 | 95.20 | 0.131 | 95.35 | 0.071 | 94.60 | 0.076 |
| | JEL | 94.30 | 0.122 | 94.85 | 0.133 | 95.35 | 0.071 | 94.75 | 0.076 |
| | AJEL | 94.80 | 0.126 | 95.20 | 0.136 | 95.40 | 0.071 | 95.05 | 0.076 |
| | NA-DRM | 94.70 | 0.114 | 94.75 | 0.109 | 95.35 | 0.067 | 94.95 | 0.063 |
| | BT-DRM | 94.35 | 0.116 | 95.15 | 0.109 | 95.10 | 0.066 | 94.85 | 0.063 |
| | NL-DRM | 94.90 | 0.113 | 95.10 | 0.108 | 95.45 | 0.067 | 95.10 | 0.063 |
| | BL-DRM | 93.90 | 0.114 | 94.70 | 0.108 | 94.85 | 0.066 | 94.85 | 0.062 |
| (0.3,0.3) | NA-EMP | 93.55 | 0.127 | 94.25 | 0.128 | 94.55 | 0.075 | 93.45 | 0.075 |
| | BT-EMP | 95.35 | 0.132 | 94.90 | 0.132 | 94.70 | 0.075 | 93.80 | 0.075 |
| | EL | 93.60 | 0.126 | 94.10 | 0.127 | 94.70 | 0.075 | 93.30 | 0.075 |
| | BT-EL | 94.75 | 0.131 | 94.85 | 0.132 | 95.05 | 0.076 | 93.70 | 0.076 |
| | JEL | 93.80 | 0.132 | 94.55 | 0.133 | 94.65 | 0.076 | 93.65 | 0.076 |
| | AJEL | 94.50 | 0.136 | 95.15 | 0.136 | 95.00 | 0.076 | 93.80 | 0.076 |
| | NA-DRM | 95.55 | 0.124 | 94.95 | 0.112 | 95.15 | 0.073 | 94.60 | 0.065 |
| | BT-DRM | 95.45 | 0.125 | 95.30 | 0.112 | 94.60 | 0.072 | 94.60 | 0.064 |
| | NL-DRM | 95.65 | 0.123 | 95.00 | 0.111 | 95.25 | 0.073 | 94.90 | 0.064 |
| | BL-DRM | 95.25 | 0.122 | 95.15 | 0.110 | 94.45 | 0.071 | 94.55 | 0.064 |
| (0.6,0.4) | NA-EMP | 92.70 | 0.115 | 94.05 | 0.127 | 93.65 | 0.068 | 94.30 | 0.075 |
| | BT-EMP | 95.50 | 0.123 | 95.35 | 0.132 | 94.90 | 0.070 | 94.80 | 0.075 |
| | EL | 93.10 | 0.113 | 94.05 | 0.126 | 93.75 | 0.068 | 94.40 | 0.074 |
| | BT-EL | 94.30 | 0.121 | 95.20 | 0.132 | 94.15 | 0.070 | 94.75 | 0.075 |
| | JEL | 92.45 | 0.124 | 94.55 | 0.133 | 93.45 | 0.070 | 94.50 | 0.076 |
| | AJEL | 92.75 | 0.127 | 95.15 | 0.137 | 93.85 | 0.070 | 94.50 | 0.076 |
| | NA-DRM | 94.95 | 0.116 | 95.30 | 0.118 | 94.70 | 0.068 | 94.60 | 0.068 |
| | BT-DRM | 95.45 | 0.116 | 95.50 | 0.118 | 94.30 | 0.066 | 94.50 | 0.067 |
| | NL-DRM | 96.05 | 0.116 | 95.80 | 0.118 | 95.10 | 0.068 | 94.85 | 0.068 |
| | BL-DRM | 94.75 | 0.112 | 95.05 | 0.116 | 94.05 | 0.066 | 94.35 | 0.067 |
| (0.7,0.7) | NA-EMP | 91.40 | 0.104 | 92.15 | 0.105 | 94.60 | 0.062 | 94.55 | 0.062 |
| | BT-EMP | 96.30 | 0.114 | 95.70 | 0.115 | 95.85 | 0.064 | 95.50 | 0.064 |
| | EL | 92.05 | 0.102 | 92.35 | 0.102 | 94.70 | 0.062 | 94.55 | 0.062 |
| | BT-EL | 95.00 | 0.110 | 94.20 | 0.111 | 95.40 | 0.064 | 95.25 | 0.064 |
| | JEL | 90.40 | 0.114 | 91.00 | 0.114 | 94.30 | 0.064 | 93.95 | 0.064 |
| | AJEL | 90.85 | 0.117 | 91.60 | 0.118 | 94.40 | 0.064 | 94.05 | 0.064 |
| | NA-DRM | 94.65 | 0.109 | 93.90 | 0.101 | 96.10 | 0.064 | 95.40 | 0.059 |
| | BT-DRM | 95.80 | 0.109 | 95.55 | 0.102 | 95.65 | 0.062 | 95.95 | 0.058 |
| | NL-DRM | 97.05 | 0.109 | 95.75 | 0.102 | 96.75 | 0.065 | 95.95 | 0.059 |
| | BL-DRM | 94.40 | 0.104 | 94.50 | 0.098 | 95.35 | 0.061 | 95.50 | 0.057 |

Table 4.17: CP(%) and AL of CIs for the difference $\mathcal{G}_0 - \mathcal{G}_1$.

| | | (100,100) | | | | (300,300) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\chi^2$ | | $Exp$ | | $\chi^2$ | | $Exp$ | |
| | | CP | AL | CP | AL | CP | AL | CP | AL |
| (0.1,0.3) | NA-EMP | 94.45 | 0.178 | 94.55 | 0.104 | 95.20 | 0.175 | 94.69 | 0.103 |
| | BT-EMP | 95.10 | 0.181 | 94.55 | 0.104 | 95.15 | 0.179 | 94.74 | 0.103 |
| | JEL | 95.45 | 0.190 | 95.15 | 0.106 | 96.40 | 0.187 | 95.15 | 0.105 |
| | AJEL | 95.60 | 0.193 | 95.25 | 0.107 | 96.60 | 0.190 | 95.25 | 0.106 |
| | NA-DRM | 94.40 | 0.146 | 95.65 | 0.085 | 94.85 | 0.138 | 94.89 | 0.080 |
| | BT-DRM | 93.45 | 0.143 | 95.10 | 0.083 | 93.80 | 0.135 | 94.34 | 0.080 |
| (0.6,0.4) | NA-EMP | 94.60 | 0.186 | 94.30 | 0.109 | 93.84 | 0.172 | 95.00 | 0.101 |
| | BT-EMP | 94.90 | 0.190 | 94.70 | 0.109 | 94.39 | 0.177 | 95.10 | 0.102 |
| | JEL | 96.25 | 0.205 | 94.90 | 0.112 | 96.85 | 0.192 | 96.30 | 0.105 |
| | AJEL | 96.40 | 0.208 | 95.00 | 0.113 | 97.10 | 0.196 | 96.40 | 0.105 |
| | NA-DRM | 95.35 | 0.175 | 95.25 | 0.102 | 95.40 | 0.150 | 95.10 | 0.088 |
| | BT-DRM | 93.95 | 0.168 | 94.05 | 0.098 | 94.59 | 0.147 | 94.75 | 0.086 |

# Chapter 5

# Discussion and Future work

In this chapter, we present a summary of our contributions in previous chapters along with some possible extensions and future research problems.

## 5.1 Summary of Current Achievements

In this thesis, we studied several important inference problems under two-sample DRMs by using the empirical likelihood method.

In Chapter 2, we developed empirical likelihood inference procedures for unknown parameters and distribution functions along with their quantiles under two-sample DRMs with estimating equations. We also proposed a testing procedure on the validity of estimating equations under DRMs, which leads to a practical validation method on the usefulness of the auxiliary information. Our inferential framework and theoretical results are very general. The results in Qin et al. (2015) and Chatterjee et al. (2016) for case-control studies are special cases of our theory for an appropriate choice of estimating functions. We also generalized the inference under DRMs to incorporate auxiliary information and covered more interesting parameters. Our results on the two-sample DRMs contain more advanced development than those in Qin and Lawless (1994) for the one-sample case. Our proposed ELR test, to our best knowledge, is the first formal procedure to test the validity of auxiliary information under the DRM or for case-control studies. Our proposed inference procedures for distribution functions and quantiles in the presence of auxiliary information are also new to the literature. The work in this chapter has been prepared as a research paper (Yuan et al., 2021a) submitted to a journal for publication.

In Chapter 3, we constructed the MELEs of the Youden index and optimal cutoff point under the DRM based on the two-sample data with or without a LLOD. Our method provide a simple solution to the estimation of optimal cutoff point. Our asymptotic results show that when there is no LLOD the proposed estimator of optimal cutoff point has a faster convergence rate than the existing nonparametric estimators, and the proposed estimator of the Youden index is asymptotically more efficient than the existing nonparametric estimators. When there is a LLOD, the proposed method is the first semiparametric or nonparametric method with rigorous theoretical justifications. Simulation experiments show that the proposed estimator for the optimal cutoff point has clear advantages over existing ones for all scenarios considered in the simulation. The work in this chapter has been wrapped up as a research paper (Yuan et al., 2021b), which has been published by *The Canadian Journal of Statistics*.

In Chapter 4, we proposed the MELEs of Gini indices of two semicontinuous distributions under the DRM. In order to establish the asymptotic normality of proposed MELEs, we first investigated the asymptotic properties of the estimators of model parameters and a special class of statistical functionals. Using techniques from U-statistics and V-statistics, we derived the asymptotic normality of the MELEs of Gini indices and showed that their asymptotic variances are smaller than those of nonparametric estimators. We also explored the asymptotic properties of a general function of two Gini indices, and used the difference of the two Gini indices as an illustrating example. We used the asymptotic results to construct CIs and perform hypothesis tests for single Gini index and difference of two Gini indices. Our methods are applicable whether or not there are excess zero values. Extensive simulation studies and applications to two real datasets demonstrate the advantages of our proposed methods over existing ones. The work of Section 4.2.1 and 4.2.2 in Chapter 4 serves as the main results in a research paper (Yuan et al., 2021d), which has been accepted for publication by *Annals of the Institute of Statistical Mathematics*. The rest of the work in this chapter has been prepared as a research paper (Yuan et al., 2021c) submitted to a journal for publication.

## 5.2 Future Work

In this section, we highlight some potential topics for future research.

### Extension to multi-sample DRMs

In practice, many statistical problems involve multiple samples. Suppose we have $k+1$

independent samples

$$\{X_{i1}, \cdots, X_{in_i}\} \sim F_i, \quad \text{for} \quad i = 0, \cdots, k,$$

where $n_i$ is the size of sample $i$ and $F_i$'s are CDFs. Let $dF_i$ denote the density of $F_i$. The multi-sample DRM is defined as

$$dF_i(x) = \exp\{\alpha_i + \boldsymbol{\beta}_i^\top \boldsymbol{q}(x)\} dF_0(x), \quad i = 1, \cdots, k,$$

for unknown parameters $\alpha_i$'s and $\boldsymbol{\beta}_i$'s, and a pre-specified, non-trivial basis function $\boldsymbol{q}(x)$. The multi-sample DRM is a desirable tool for combing information when the samples share certain characteristics; see for example Cai et al. (2017); Chen and Liu (2013) and Wang et al. (2017a). In this thesis, we focused on two-sample data under the DRM. We may generalize our framework and results in Chapter 2 to multi-sample DRM. Comparison for multiple Gini indices under multi-sample DRM would also be an interesting topic.

**Inference on Gini indices based on paired data**

Paired data are frequently seen in many field. For example, in biomedical studies, there is always interest in finding new promising substitutes for the conventional biomarkers or approaches. Different diagnostic tests are administered to the same or identical subjects. In economics, some indices of each country such as gross domestic product are collected every year. The pairs are correlated and do share some common characteristics. Simply ignoring the correlation would result in efficiency loss. Chen et al. (2021) proposed a composite empirical likelihood method for the inference problems for clustered data under DRMs, and demonstrated the advantages of this approach numerically. This motivates us to adopt the composite empirical likelihood for paired data under the DRM.

Suppose we have a paired sample $\{(X_{01}, X_{11}), \cdots, (X_{0n}, X_{1n})\}$ from the a population with joint distribution function $F$. Let $F_0(x) = F(x, \infty)$ and $F_1(x) = F(\infty, x)$ denote two marginal CDFs. We propose to link $F_0$ and $F_1$ by a DRM

$$dF_1(x) = \exp\{\alpha + \boldsymbol{\beta}^\top \boldsymbol{q}(x)\} dF_0(x),$$

where $dF_i$ is the density of $F_i$; $\alpha$ and $\boldsymbol{\beta}$ are the unknown parameters; $\boldsymbol{q}(x)$ is the basis function. Instead of directly modelling the correlation structure between $X_{0j}$ and $X_{1j}$, we propose to use the composite empirical likelihood of Chen et al. (2021) as the inference tool. We plan to study the asymptotic properties of the corresponding maximum composite empirical likelihood estimators of two Gini indices under the DRM based on the paired data. We leave this as a future research problem.

**ELR-based CIs for Gini indices**

As we discussed in Section 1.2.1, the ELR-based CIs have many nice properties such as range-preserving. In Chapter 4, we concentrated on the Wald-type CIs for two Gini indices $\mathcal{G}_0$ and $\mathcal{G}_1$, and their difference $\mathcal{G}_0 - \mathcal{G}_1$. It would be interesting to consider the ELR-based CIs for $\mathcal{G}_0$, $\mathcal{G}_1$, and $\mathcal{G}_0 - \mathcal{G}_1$. For convenience of presentation, we assume that there is no excessive zero in the two samples and we focus on the ELR-based CI for $\mathcal{G}_0$.

Note that the definition of Gini index in (4.2) and the equivalent form of $D_i$ in (4.17) imply that $\mathcal{G}_0$ can be expressed as

$$\mathcal{G}_0 = \frac{E_0\{2XF_0(X) - X\}}{E_0(X)},$$

where $E_0$ means the expectation is taken with respect to $F_0$. Hence, $\mathcal{G}_0$ satisfies the following estimating equation:

$$E_0\{g(X; \mathcal{G}_0, F_0)\} = 0 \tag{5.1}$$

with $g(x; \mathcal{G}_0, F_0) = 2xF_0(x) - (\mathcal{G}_0 + 1)x$. The above estimating equation is different from the estimating equations we considered in Chapter 2 since it involves the unknown distribution function $F_0(x)$.

Following Qin et al. (2010), we can replace $F_0$ in $g(x; \mathcal{G}_0, F_0)$ by the empirical CDF $\bar{F}_0(x) = n_0^{-1} \sum_{j=1}^{n_0} I(X_{0j} \leq x)$ or $\tilde{F}_0(x)$ in (1.17), the MELE of $F_0(x)$ under two-sample DRMs without auxiliary information. We then use $g(x; \mathcal{G}_0, \bar{F}_0)$ or $g(x; \mathcal{G}_0, \tilde{F}_0)$ as one of the estimating equations in Chapter 2 and construct the ELR-based CI for $\mathcal{G}_0$. We plan to examine the asymptotic properties of the corresponding ELR statistic for $\mathcal{G}_0$ and compare the ELR-based CI with the Wald-type CI.

Following a similar approach, we could embed the nuisance functionals into the estimating equations (2.2) in Chapter 2; consequently, the framework of Chapter 2 will include more complex quantities such as Lorenz curve and ROC curve. We leave this for future research.

# References

Alonzo, T. A. and Pepe, M. S. (2002). Distribution-free ROC analysis using binary regression techniques. *Biostatistics*, 3:421–432.

Anand, S. (1983). Inequality and poverty in Malaysia: Measurement and decomposition. Report 9268, The World Bank.

Anderson, J. A. (1979). Multivariate logistic compounds. *Biometrika*, 66:17–26.

Andrews, D. F. and Herzberg, A. M. (2012). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer, New York.

Bantis, L. E., Nakas, C. T., and Reiser, B. (2019). Construction of confidence intervals for the maximum of the Youden index and the corresponding cutoff point of a continuous biomarker. *Biometrical Journal*, 61:138–156.

Bantis, L. E., Tsimikas, J. V., and Georgiou, S. D. (2012). Survival estimation through the cumulative hazard function with monotone natural cubic splines. *Lifetime Data Analysis*, 18:364–396.

Bantis, L. E., Yan, Q., Tsimikas, J. V., and Feng, Z. (2017). Estimation of smooth ROC curves for biomarkers with limits of detection. *Statistics in Medicine*, 36:3830–3843.

Bebu, I. and Mathew, T. (2008). Comparing the means and variances of a bivariate log-normal distribution. *Statistics in Medicine*, 27:2684–2696.

Bonetti, M., Gigliarano, C., and Muliere, P. (2009). The Gini concentration test for survival data. *Lifetime Data Analysis*, 15:493–518.

Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 26:211–243.

Cai, S. (2014). *On Dual Empirical Likelihood Inference under Semiparametric Density Ratio Models in the Presence of Multiple Samples with Applications to Long Term Monitoring of Lumber Quality*. PhD thesis, University of British Columbia.

Cai, S. and Chen, J. (2018). Empirical likelihood inference for multiple censored samples. *The Canadian Journal of Statistics*, 46:212–232.

Cai, S., Chen, J., and Zidek, J. V. (2017). Hypothesis testing in the presence of multiple samples under density ratio models. *Statistica Sinica*, 27:761–783.

Chatterjee, N., Chen, Y.-H., Maas, P., and Carroll, R. J. (2016). Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *Journal of the American Statistical Association*, 111:107–117.

Chen, B., Li, P., Qin, J., and Yu, T. (2016). Using a monotonic density ratio model to find the asymptotically optimal combination of multiple diagnostic tests. *Journal of the American Statistical Association*, 111:861–874.

Chen, J., Li, P., Liu, Y., and Zidek, J. V. (2021). Composite empirical likelihood for multisample clustered data. *Journal of Nonparametric Statistics*, 33:60–81.

Chen, J. and Liu, Y. (2013). Quantile and quantile-function estimations under density ratio model. *The Annals of Statistics*, 41:1669–1692.

Cowell, F. (2011). *Measuring Inequality*. Oxford University Press, Oxford.

David, H. (1968). Miscellanea: Gini's mean difference rediscovered. *Biometrika*, 55:573–575.

Davidson, R. (2009). Reliable inference for the Gini index. *Journal of Econometrics*, 150:30–40.

de Carvalho, M. and Davison, A. C. (2014). Spectral density ratio models for multivariate extremes. *Journal of the American Statistical Association*, 109:764–776.

DiCiccio, T., Hall, P., and Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *The Annals of Statistics*, 19:1053–1061.

Fluss, R., Faraggi, D., and Reiser, B. (2005). Estimation of the Youden index and its associated cutoff point. *Biometrical Journal*, 47:458–472.

Fokianos, K. (2004). Merging information for semiparametric density estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66:941–958.

Fokianos, K. (2007). Density ratio model selection. *Journal of Statistical Computation and Simulation*, 77:805–819.

Fokianos, K. and Kaimi, I. (2006). On the effect of misspecifying the density ratio model. *Annals of the Institute of Statistical Mathematics*, 58:475–497.

Fokianos, K., Kedem, B., Qin, J., and Short, D. A. (2001). A semiparametric approach to the one-way layout. *Technometrics*, 43:56–65.

Fokianos, K., Peng, A., and Qin, J. (1999). A generalized-moments specification test for the logistic link. *The Canadian Journal of Statistics*, 27:735–750.

Gastwirth, J. L. (1972). The estimation of the Lorenz curve and Gini index. *The Review of Economics and Statistics*, 54:306–316.

Giles, D. E. (2004). Calculating a standard error for the Gini coefficient: Some further results. *Oxford Bulletin of Economics and Statistics*, 66:425–433.

Gini, C. (1912). *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche*. Tipogr. di P. Cuppini, Bologna.

Gulick, R. M., Joan Hu, X., Fiscus, S. A., Fletcher, C. V., Haubrich, R., Cheng, H., Acosta, E., Lagakos, S. W., Swanstrom, R., Freimutn, W., Snyder, S., Mills, C., Fischl, M., Pettinelli, C., and Katzenstein, D. (2000). Randomized study of saquinavir with ritonavir or nelfinavir together with delavirdine, adefovir, or both in human immunodeficiency virus—infected adults with virologic failure on indinavir: AIDS Clinical Trials Group Study 359. *The Journal of Infectious Diseases*, 182:1375–1384.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review*, 58:109–127.

Hawkins, D. M. (2002). Diagnostics for conformity of paired quantitative measurements. *Statistics in Medicine*, 21:1913–1935.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19:293–325.

Hsieh, F. and Turnbull, B. W. (1996). Nonparametric methods for evaluating diagnostic tests. *Statistica Sinica*, 6:47–62.

Imbens, G. W. and Lancaster, T. (1994). Combining micro and macro data in microeconometric models. *The Review of Economic Studies*, 61:655–680.

Jiang, S. and Tu, D. (2012). Inference on the probability $P(T_1 < T_2)$ as a measurement of treatment effect under a density ratio model and random censoring. *Computational Statistics & Data Analysis*, 56:1069–1078.

Karagiannis, E. and Kovacevic, M. (2000). A method to calculate the jackknife variance estimator for the Gini coefficient. *Oxford Bulletin of Economics and Statistics*, 62:119–122.

Kay, R. and Little, S. (1987). Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika*, 74:495–501.

Keziou, A. and Leoni-Aubin, S. (2008). On empirical likelihood for semiparametric two-sample density ratio models. *Journal of Statistical Planning and Inference*, 138:915–928.

Krishnamoorthy, K. and Mathew, T. (2003). Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals. *Journal of Statistical Planning and Inference*, 115:103–121.

Krzanowski, W. J. and Hand, D. J. (2009). *ROC Curves for Continuous Data*. Chapman and Hall/CRC, London.

Li, H., Liu, Y., Liu, Y., and Zhang, R. (2018). Comparison of empirical likelihood and its dual likelihood under density ratio model. *Journal of Nonparametric Statistics*, 30:581–597.

Li, P., Liu, Y., and Qin, J. (2017). Semiparametric inference in a genetic mixture model. *Journal of the American Statistical Association*, 112:1250–1260.

Li, P. and Qin, J. (2011). A new nuisance-parameter elimination method with application to the unordered homologous chromosome pairs problem. *Journal of the American Statistical Association*, 106:1476–1484.

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, 9:209–219.

Lv, X., Zhang, G., and Ren, G. (2017). Gini index estimation for lifetime data. *Lifetime Data Analysis*, 23:275–304.

Mises, R. v. (1947). On the asymptotic distribution of differentiable statistical functions. *The Annals of Mathematical Statistics*, 18:309–348.

Modarres, R. and Gastwirth, J. L. (2006). A cautionary note on estimating the standard error of the Gini index of inequality. *Oxford Bulletin of Economics and Statistics*, 68:385–390.

Murray, M. D., Harris, L. E., Overhage, J. M., Zhou, X.-H., Eckert, G. J., Smith, F. E., Buchanan, N. N., Wolinsky, F. D., McDonald, C. J., and Tierney, W. M. (2004). Failure of computerized treatment suggestions to improve health outcomes of outpatients with uncomplicated hypertension: Results of a randomized controlled trial. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 24:324–337.

Ogwang, T. (2000). A convenient method of computing the Gini index and its standard error. *Oxford Bulletin of Economics and Statistics*, 62:123–129.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75:237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18:90–120.

Owen, A. B. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, 19:1725–1747.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC, Boca Raton.

Peng, L. (2011). Empirical likelihood methods for the Gini index. *Australian & New Zealand Journal of Statistics*, 53:131–139.

Pepe, M. S. (2000). An interpretation for the ROC curve and inference using GLM procedures. *Biometrics*, 56:352–359.

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

Percy, M. E., Andrews, D. F., and Thompson, M. W. (1982). Duchenne muscular dystrophy carrier detection using logistic discrimination: serum creatine kinase, hemopexin, pyruvate kinase, and lactate dehydrogenase in combination. *American Journal of Medical Genetics*, 13:27–38.

Qin, J. (1993). Empirical likelihood in biased sample problems. *The Annals of Statistics*, 21:1182–1196.

Qin, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika*, 85:619–630.

Qin, J. (1999). Empirical likelihood ratio based confidence intervals for mixture proportions. *The Annals of Statistics*, 27:1368–1384.

Qin, J. (2017). *Biased Sampling, Over-identified Parameter Problems and Beyond*. Springer, Singapore.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22:300–325.

Qin, J. and Lawless, J. (1995). Estimating equations, empirical likelihood and constraints on parameters. *The Canadian Journal of Statistics*, 23:145–159.

Qin, J. and Zhang, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, 84:609–618.

Qin, J. and Zhang, B. (2003). Using logistic regression procedures for estimating receiver operating characteristic curves. *Biometrika*, 90:585–596.

Qin, J. and Zhang, B. (2005). Density estimation under a two-sample semiparametric model. *Nonparametric Statistics*, 17:665–683.

Qin, J., Zhang, H., Li, P., Albanes, D., and Yu, K. (2015). Using covariate-specific disease prevalence information to increase the power of case-control studies. *Biometrika*, 102:169–180.

Qin, Y., Rao, J., and Wu, C. (2010). Empirical likelihood confidence intervals for the Gini measure of income inequality. *Economic Modelling*, 27:1429–1435.

Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment. *Sankhyā: The Indian Journal of Statistics, Series A*, 27:311–324.

Ren, J.-J. (2008). Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. *The Annals of Statistics*, 36:147–166.

Ruopp, M. D., Perkins, N. J., Whitcomb, B. W., and Schisterman, E. F. (2008). Youden index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biometrical Journal*, 50:419–430.

Schisterman, E. F. and Perkins, N. (2007). Confidence intervals for the Youden index and corresponding optimal cut-point. *Communications in Statistics – Simulation and Computation*, 36:549–563.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

Shan, G. (2015). Improved confidence intervals for the Youden index. *PloS One*, 10:e0127272.

Shen, Y., Ning, J., and Qin, J. (2012). Likelihood approaches for the invariant density ratio model with biased-sampling data. *Biometrika*, 99:363–378.

Shen, Y., Qin, J., and Costantino, J. P. (2007). Inference of tamoxifen's effects on prevention of breast cancer from a randomized controlled trial. *Journal of the American Statistical Association*, 102:1235–1244.

Simpson, J., Olsen, A., and Eden, J. C. (1975). A Bayesian analysis of a multiplicative treatment effect in weather modification. *Technometrics*, 17:161–166.

Tsao, M. and Wu, C. (2006). Empirical likelihood inference for a common mean in the presence of heteroscedasticity. *The Canadian Journal of Statistics*, 34:45–59.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, New York.

Wan, S. and Zhang, B. (2007). Smooth semiparametric receiver operating characteristic curves for continuous diagnostic tests. *Statistics in Medicine*, 26:2565–2586.

Wan, S. and Zhang, B. (2008). Comparing correlated ROC curves for continuous diagnostic tests under density ratio models. *Computational Statistics & Data Analysis*, 53:233–245.

Wang, C. (2017). *Empirical Likelihood and Bootstrap Inference with Constraints*. PhD thesis, University of Waterloo.

Wang, C., Marriott, P., and Li, P. (2017a). Testing homogeneity for multiple nonnegative distributions with excess zero observations. *Computational Statistics & Data Analysis*, 114:146–157.

Wang, C., Marriott, P., and Li, P. (2018). Semiparametric inference on the means of multiple nonnegative distributions with excess zero observations. *Journal of Multivariate Analysis*, 166:182–197.

Wang, C., Tan, Z., and Louis, T. A. (2011). Exponential tilt models for two-group comparison with censored data. *Journal of Statistical Planning and Inference*, 141:1102–1117.

Wang, D., Tian, L., and Zhao, Y. (2017b). Smoothed empirical likelihood for the Youden index. *Computational Statistics & Data Analysis*, 115:1–10.

Wang, D. and Zhao, Y. (2016). Jackknife empirical likelihood for comparing two Gini indices. *The Canadian Journal of Statistics*, 44:102–119.

Wang, D., Zhao, Y., and Gilmore, D. W. (2016). Jackknife empirical likelihood confidence interval for the Gini index. *Statistics & Probability Letters*, 110:289–295.

Wang, S. and Zhang, B. (2014). Semiparametric empirical likelihood confidence intervals for AUC under a density ratio model. *Computational Statistics & Data Analysis*, 70:101–115.

Wei, W. and Zhou, Y. (2016). Semiparametric maximum likelihood estimation for a two-sample density ratio model with right-censored data. *The Canadian Journal of Statistics*, 44:58–81.

Wu, C. and Thompson, M. E. (2020). *Sampling Theory and Practice*. Springer, Cham.

Wu, J., Jiang, G., Wong, A., and Sun, X. (2002). Likelihood analysis for the ratio of means of two independent log-normal distributions. *Biometrics*, 58:463–469.

Yin, J. and Tian, L. (2014). Joint inference about sensitivity and specificity at the optimal cut-off point associated with Youden index. *Computational Statistics & Data Analysis*, 77:1–13.

Yitzhaki, S. (1991). Calculating jackknife variance estimators for parameters of the Gini method. *Journal of Business & Economic Statistics*, 9:235–239.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3:32–35.

Yuan, M., Li, P., and Wu, C. (2021a). Semiparametric empirical likelihood inference with estimating equations under density ratio models. arXiv:2102.13232.

Yuan, M., Li, P., and Wu, C. (2021b). Semiparametric inference of the Youden index and the optimal cut-off point under density ratio models. *The Canadian Journal of Statistics*, 49:965–986.

Yuan, M., Li, P., and Wu, C. (2021c). Semiparametric inference on Gini indices of two semicontinuous populations under density ratio models. arXiv:2106.02741.

Yuan, M., Wang, C., Lin, B., and Li, P. (2021d). Semiparametric inference on general functionals of two semicontinuous populations. *Annals of the Institute of Statistical Mathematics*. https://doi.org/10.1007/s10463-021-00804-4.

Zhang, A. G. and Chen, J. (2021). Density ratio model with data-adaptive basis function. arXiv:2103.03445.

Zhang, A. G., Zhu, G., and Chen, J. (2020). Empirical likelihood ratio test on quantiles under a density ratio model. arXiv:2007.10586.

Zhang, B. (2000). Quantile estimation under a two-sample semi-parametric model. *Bernoulli*, 6:491–511.

Zhang, B. (2002). Assessing goodness-of-fit of generalized logit models based on case-control data. *Journal of Multivariate Analysis*, 82:17–38.

Zhang, B. (2006). A semiparametric hypothesis testing procedure for the ROC curve area under a density ratio model. *Computational Statistics & Data Analysis*, 50:1855–1876.

Zhang, D. and Zhang, B. (2014). Semiparametric empirical likelihood confidence intervals for the difference of areas under two correlated ROC curves under density ratio model. *Biometrical Journal*, 56:678–696.

Zhou, H. and Qin, G. (2012). New nonparametric confidence intervals for the Youden index. *Journal of Biopharmaceutical Statistics*, 22:1244–1257.

Zhou, X.-H. and Cheng, H. (2008). A computer program for estimating the re-transformed mean in heteroscedastic two-part models. *Computer Methods and Programs in Biomedicine*, 90:210–216.

Zhou, X.-H., Gao, S., and Hui, S. L. (1997). Methods for comparing the means of two independent log-normal samples. *Biometrics*, 53:1129–1135.

Zhou, X.-H. and McClish, D. K. (2002). *Statistical Methods in Diagnostic Medicine*. John Wiley & Sons, New York.

Zhou, X.-H. and Tu, W. (1999). Comparison of several independent population means when their samples contain log-normal and possibly zero observations. *Biometrics*, 55:645–651.

Zhou, X.-H. and Tu, W. (2000). Interval estimation for the ratio in means of log-normally distributed medical costs with zero values. *Computational Statistics & Data Analysis*, 35:201–210.

Zhuang, W., Hu, B., and Chen, J. (2019). Semiparametric inference for the dominance index under the density ratio model. *Biometrika*, 106:229–241.

Zou, F., Fine, J. P., and Yandell, B. S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika*, 89:61–75.

Zou, K. H., Liu, A., Bandos, A. I., Ohno-Machado, L., and Rockette, H. E. (2011). *Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis*. CRC Press, Boca Raton.