Electronic Journal of Statistics
Vol. 16 (2022) 5321–5377
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2069

Semiparametric empirical likelihood inference with estimating equations under density ratio models*

Meng Yuan

Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
e-mail: meng.yuan@uwaterloo.ca

Pengfei Li

Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
e-mail: pengfei.li@uwaterloo.ca

and

Changbao Wu

Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
e-mail: cbwu@uwaterloo.ca

Abstract: The density ratio model (DRM) provides a flexible and useful platform for combining information from multiple sources. In this paper, we consider statistical inference under two-sample DRMs with additional parameters 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.

MSC2020 subject classifications: Primary 62G05, 62G10; Secondary 62G20.

Keywords and phrases: Auxiliary information, density ratio model, empirical likelihood, estimating equations.

Received November 2021.

*P. Li and C. Wu were supported in part by the Natural Sciences and Engineering Research Council of Canada.

Contents

1. Introduction

1.1. Problem setup

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. We assume that the CDFs F_0 and F_1 are linked through a semiparametric density ratio model (DRM) (Anderson, 1979, Qin, 2017),

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

where $dF_i(x)$ 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 a prespecified, nontrivial function of dimension d; and the baseline distribution F_0 is unspecified. We further assume that the main parameters of interest can be express and/or certain auxiliary information about F_0, F_1, and $\boldsymbol{\theta}$ is available in the form of functionally independent unbiased estimating equations (EEs):

$$E_0\{\boldsymbol{g}(X; \boldsymbol{\psi}, \boldsymbol{\theta})\} = \mathbf{0}, \tag{1.2}$$

where $E_0(\cdot)$ refers to the expectation operator with respect to F_0, $\boldsymbol{\psi}$ 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 paper, 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 Model (1.1) with unbiased EEs in (1.2);
(2) we propose a new testing procedure on the validity of (1.2) under Model (1.1), which leads to a practical validation method on the usefulness of the auxiliary information.

The semiparametric DRM in (1.1) provides a flexible and useful platform for combining information from multiple sources (Qin, 2017). It 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 (Chen and Liu, 2013, Cai et al., 2017, Zhuang et al., 2019). With the unspecified F_0, the DRM embraces many commonly used statistical models including distributions of exponential families (Kay and Little, 1987). For example, when $\boldsymbol{q}(x) = \log x$, the DRM 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; when $\boldsymbol{q}(x) = x$, it includes two normal distributions with different means but a common variance and two exponential distributions with different rates. Moreover, it has a natural connection to the well-studied logistic regression if one treats $D = 0$ and 1 as indicators for the observations from F_0 and F_1, respectively. Among others, Anderson (1979) and Qin and Zhang (1997) noticed that the DRM is equivalent to the logistic regression model via the fact that

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

where $\alpha^* = \alpha + \log\{P(D = 1)/P(D = 0)\}$.

The EEs in (1.2) 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 1 below and Section 7.1 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 and 3 below illustrate the use of auxiliary information through EEs in the form of (1.2).

Example 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* μ_0 *and* μ_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* δ*, 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 (1.1) 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* (1.1)*,* δ *can be defined through the following EE:*

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

with $\boldsymbol{\psi} = \delta$*. When additional information is available, we may add more EEs to improve the estimation efficiency; see Section 4.1 for further detail.*

Example 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 specified in* (1.3)*. Then, using the equivalence between the DRM and the logistic regression discussed above,* F_0 *and* F_1 *satisfy the DRM* (1.1)*.*

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*:* $\phi(a_{l-1}, a_l) = P(D = 1 | a_{l-1} < Y \le a_l)$ *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*

$$g_l(x; \boldsymbol{\psi}, \boldsymbol{\theta}) = I(a_{l-1} < x \le a_l) \left[\frac{\pi}{1-\pi} \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\} - \frac{\phi(a_{l-1}, a_l)}{1 - \phi(a_{l-1}, a_l)} \right]. \quad (1.4)$$

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 7.2, 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

$$\boldsymbol{g}(x; \boldsymbol{\psi}, \boldsymbol{\theta}) = [-(1-\pi)h(y) + \pi \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\}\{1 - h(y)\}](1, y^\top)^\top \tag{1.5}$$

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

$$\boldsymbol{g}(x; \boldsymbol{\theta}) = [-(1-\pi_E)h(y) + \pi_E \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(x)\}\{1 - h(y)\}](1, y^\top)^\top \tag{1.6}$$

with π_E being the proportion of diseased individuals in the external study.

Example 3 (A two-sample problem with common mean)**.** *Tsao and Wu (2006) considered two populations with a common mean. This type of problems 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 (1.1) with $\boldsymbol{q}(x) = (x, x^2)^\top$. Under this model, the common-mean assumption can be incorporated via the EE:*

$$E_0\{X \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X)\} - X\} = 0. \tag{1.7}$$

1.2. Literature review

The DRM has been investigated extensively because of its flexibility and efficiency. For example, it has been applied to multiple-sample hypothesis-testing problems (Fokianos et al., 2001, Cai et al., 2017, Wang et al., 2017, 2018) and quantile and quantile-function estimation (Zhang, 2000, Chen and Liu, 2013, Chen et al., 2021a). These inference problems can be viewed as special cases, without auxiliary information, of the first goal to be achieved in this paper.

Other applications of the DRM include receiver operating characteristic (ROC) analysis (Qin and Zhang, 2003, Chen et al., 2016, Yuan et al., 2021), inference under semiparametric mixture models (Qin, 1999, Zou et al., 2002, Li and Qin, 2011, Li et al., 2017, Yuan et al., 2022), the modeling of multivariate extremal distributions (de Carvalho and Davison, 2014), and dominance index estimation (Zhuang et al., 2019). Recently, Li et al. (2018) studied maximum empirical likelihood estimation (MELE) and empirical likelihood ratio (ELR) based confidence intervals (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. They did not consider auxiliary information, and because of the specific form of $\boldsymbol{\psi}$, their results do not apply to the mean ratio discussed in Example 1. Zhang et al. (2022) 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. Again, they 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 EEs in (1.2).

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 Model (1.1) with the unbiased EEs in (1.4). 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 7.2, we argue that their results are applicable to the MELE of $\boldsymbol{\theta}$ under Model (1.1) with the unbiased EEs in (1.5) but not to the MELE of $\boldsymbol{\theta}$ under Model (1.1) with the unbiased EEs in (1.6). 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.

1.3. Our contributions

With two-sample observations from the DRM (1.1), we use the empirical likelihood (EL) of Owen (1988, 2001) to incorporate the unbiased EEs in (1.2). We show that the MELE of $(\boldsymbol{\psi}, \boldsymbol{\theta})$ is asymptotically normal, and its asymptotic variance will not decrease when an EE in (1.2) is dropped. We also develop an ELR statistic for testing a general hypothesis about $(\boldsymbol{\psi}, \boldsymbol{\theta})$, and show that it has a χ^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 χ^2 null limiting distribution to test the validity of some or all of the EEs in (1.2).

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 (1.2). Our results are also applicable to cases that are not covered by these two earlier studies, e.g., Example 2 with the EEs in (1.6) and Example 3.

(3) Our proposed ELR statistic, to our best 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 paper is organized as follows. In Section 2, we develop the EL 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 EEs in (1.2). In Section 3, we discuss inference procedures for (F_0, F_1) and their quantiles. Simulation results are reported in Section 4, and two real-data examples are presented in Section 5. We conclude the paper with a discussion in Section 6. For convenience of presentation, all the technical details are given in the Section 7.

2. Empirical likelihood and inference on $(\boldsymbol{\psi},\boldsymbol{\theta})$

In this section, we first develop the EL formulation under the DRM (1.1) with the unbiased EEs in (1.2). With two samples $\{X_{01},\dots,X_{0n_0}\}$ and $\{X_{11},\dots,X_{1n_1}\}$ 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 EL formulation of Owen (2001), the baseline distribution function $F_0(x)$ would have been modeled as $F_0(x)=\sum_{j=1}^{n_0} p_j I(X_{0j}\le x)$, where $p_j = dF_0(X_{0j})$ for $j=1,\dots,n_0$. Under the two-sample DRM (1.1), we use the combined sample to model the baseline function $F_0(x)$ as

$$F_0(x)=\sum_{i=0}^{1}\sum_{j=1}^{n_i} p_{ij} I(X_{ij}\le x), \tag{2.1}$$

where $p_{ij} = dF_0(X_{ij})$ for $i=0,1$ and $j=1,\dots,n_i$. Note that the size of the combined sample is $n=n_0+n_1$. With (2.1) and under the DRM (1.1), the EL function is given by

$$\mathcal{L}_n=\left\{\prod_{i=0}^{1}\prod_{j=1}^{n_i} p_{ij}\right\}\left[\prod_{j=1}^{n_1}\exp\left\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{1j})\right\}\right]. \tag{2.2}$$

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

$$\begin{aligned} \mathcal{C}_1 &= \Big\{(F_0, \boldsymbol{\theta}) : p_{ij} > 0,\ \sum_{i=0}^{1}\sum_{j=1}^{n_i} p_{ij} = 1, \\ &\qquad \sum_{i=0}^{1}\sum_{j=1}^{n_i} p_{ij}\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} = 1\Big\} \end{aligned} \tag{2.3}$$

and

$$\mathcal{C}_2 = \left\{(F_0, \boldsymbol{\psi}, \boldsymbol{\theta}) : \sum_{i=0}^{1}\sum_{j=1}^{n_i} p_{ij}\boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta}) = \mathbf{0}\right\}, \tag{2.4}$$

where the set of constraints $\mathcal{C}_1$ ensures that F_0 and F_1 are CDFs and the set of constraints $\mathcal{C}_2$ is induced by the EEs in (1.2).

Using the Lagrange multiplier method and for the given $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$, it can be shown that the maximizer of the EL 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 λ and $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_r)^\top$ are the solutions to the following set of $r + 1$ equations:

$$\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1}{1 + \lambda\left[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1\right] + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})} = 0, \tag{2.5}$$

$$\sum_{i=0}^{1}\sum_{j=1}^{n_i} \frac{\boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})}{1 + \lambda\left[\exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X_{ij})\} - 1\right] + \boldsymbol{\nu}^\top \boldsymbol{g}(X_{ij}; \boldsymbol{\psi}, \boldsymbol{\theta})} = \mathbf{0}. \tag{2.6}$$

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

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

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 $(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})$. Throughout the paper, we assume that the total sample size $n = n_0 + n_1 \to \infty$ and $\lambda^* \to n_1/n$ for some constant $\lambda^* \in (0, 1)$. This assumption 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.

Let $\boldsymbol{\eta} = (\boldsymbol{\psi}^\top, \boldsymbol{\theta}^\top)^\top$ be the vector of parameters and $\boldsymbol{\eta}^*$ be the true value of $\boldsymbol{\eta}$. We further define

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

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 1. *Assume that the regularity conditions in Section 7.3.1 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

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

In the absence of the constraints $\mathcal{C}_2$ in (2.4), we can maximize the EL function in (2.2) with respect only to the CDF constraints $\mathcal{C}_1$ in (2.3) to obtain the MELE $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. Qin and Zhang (1997) and Keziou and Leoni-Aubin (2008) noticed that $\tilde{\boldsymbol{\theta}}$ equivalently maximizes the following dual likelihood:

$$\begin{aligned}\ell_{nd}(\boldsymbol{\theta}) \quad = \quad & -\sum_{i=0}^{1}\sum_{j=1}^{n_i}\log\left\{1 + \lambda^*\left[\exp\left\{\boldsymbol{\theta}^\top\boldsymbol{Q}(X_{ij})\right\} - 1\right]\right\} \\ & + \sum_{j=1}^{n_1}\left\{\boldsymbol{\theta}^\top\boldsymbol{Q}(X_{1j})\right\}. \end{aligned} \tag{2.7}$$

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

Corollary 1. *Under the conditions of Theorem 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 EE in (1.2) is dropped.*

We provide some further comments on the results presented in Corollary 1. First, when the dimensions of the parameters $\boldsymbol{\psi}$ and the EEs 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 (2.4). Because of the result in Corollary 1(a), the estimators $\tilde{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\psi}}$ share the same asymptotic property. Second, Corollary 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. *Assume that the conditions of Theorem 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 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

$$\{\boldsymbol{\psi} : R_n^*(\boldsymbol{\psi}) \leq \chi^2_{q,1-a}\}, \tag{2.8}$$

where $\chi^2_{q,1-a}$ is the $100(1 - a)$th quantile of the χ^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 in this section is to construct an ELR statistic for testing the validity and usefulness of the auxiliary information. Let

$$\begin{aligned} W_n &= 2\left\{ \sup_{(\boldsymbol{\eta},F_0)\in\mathcal{C}_1} \log\mathcal{L}_n - \sup_{(\boldsymbol{\eta},F_0)\in\mathcal{C}_1\cap\mathcal{C}_2} \log\mathcal{L}_n \right\} \\ &= 2\left\{ \ell_{nd}(\tilde{\boldsymbol{\theta}}) - \ell_n(\hat{\boldsymbol{\psi}},\hat{\boldsymbol{\theta}}) \right\}. \end{aligned} \tag{2.9}$$

Theorem 3. *Under the conditions of Theorem 1 and as $n\to\infty$, we have $W_n \to \chi^2_{r-p}$ in distribution if (1.2) is correctly specified.*

We can also test the validity of some but not all of the EEs in (1.2). To do so, we partition the EEs in (1.2) 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 $r-m\geq p$. We are interested in testing $H_0: E_0\{\boldsymbol{g}_2(X;\boldsymbol{\psi},\boldsymbol{\theta})\} = \mathbf{0}$. 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 $E_0\{\boldsymbol{g}_1(x;\boldsymbol{\psi},\boldsymbol{\theta})\} = \mathbf{0}$. That is,

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

where λ 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 (2.9) 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. *Under the conditions of Theorem 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.*

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 (2.5) and (2.6) 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

$$\hat{F}_0(x) = \sum_{i=0}^{1} \sum_{j=1}^{n_i} \hat{p}_{ij} I(X_{ij} \le x)$$

and

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

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

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

and

$$\boldsymbol{B}_0^*(x) = \begin{pmatrix} \boldsymbol{B}_{0\boldsymbol{\theta}}(x) \\ \boldsymbol{B}_{0\boldsymbol{u}}(x) \end{pmatrix}, \quad \boldsymbol{B}_1^*(x) = \begin{pmatrix} \boldsymbol{B}_{1\boldsymbol{\theta}}(x) \\ \boldsymbol{B}_{1\boldsymbol{u}}(x) \end{pmatrix},$$

where

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

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

Theorem 4. *Assume that the conditions of Theorem 1 are satisfied.*

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

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

where

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

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* $\Sigma_{ls}(x,y)$ *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* $\Sigma_{ls}(x,y)$ *cannot decrease if one EE in* (1.2) *is dropped.*

Theorem 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 τth-quantile of F_i as $\xi_{i,\tau} = \inf\{x : F_i(x) \ge \tau\}$ and its MELE as

$$\hat{\xi}_{i,\tau} = \inf\{x : \hat{F}_i(x) \ge \tau\}. \tag{3.1}$$

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

$$\tilde{\xi}_{i,\tau} = \inf\{x : \tilde{F}_i(x) \ge \tau\}. \tag{3.2}$$

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 5. *Assume that the conditions of Theorem 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 Barhadur 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 6. *Assume that the conditions in Theorem 5 hold for* $x = \xi_{l,\tau_l}$ *and* $x = \xi_{s,\tau_s}$.

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

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

where

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

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

(c) if $r > p$, the asymptotic variance matrix Ω_{ls} of the DRM-EE quantile estimators cannot decrease if one EE in (1.2) is dropped.

Using the results of Theorems 4 and 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 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 Model (1.1) with the unbiased EEs in (1.2) and $E_0\{g_1^*(X; \boldsymbol{\theta}, \zeta)\} = 0$. Then the ELR function of ζ 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 ξ at the quantile level τ 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 Model (1.1) with the unbiased EEs in (1.2) and $E_0\{g_2^*(X; \boldsymbol{\theta}, \xi)\} = 0$. Then the ELR function of ξ 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, we have the following results for $R_{n1}(\zeta^*)$ and $R_{n2}(\xi^*)$, where ζ^* and ξ^* are the true values of ζ and ξ.

Corollary 3. *Under the conditions of Theorem 1, as $n \to \infty$, both $R_{n1}(\zeta^*)$ and $R_{n2}(\xi^*)$ converge in distribution to χ_1^2.*

Corollary 3 enables us to construct the ELR-based CI for ζ and ξ. For example, the ELR-based CI for ξ with level $1-a$ can be constructed as $\{\xi : R_{n2}(\xi) \le \chi^2_{1,1-a}\}$.

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.

4.1. Simulation studies for inferences on ψ

4.1.1. 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 (1.1) 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 1.

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

$$Z = 1 + 0.5X + \epsilon \quad \text{and} \quad \epsilon \sim N(0, 1). \tag{4.1}$$

That is, given X_{ij}, Z_{ij} is generated from (4.1), 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 μ_{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 (1.2).

4.1.2. 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 1. The performance of each estimator is evaluated by the relative bias (RB) and the mean squared error (MSE). Simulation results on the three point estimators are presented in Table 1.

Table 1
RB (%) and MSE ($\times 100$) of three point estimators of the mean ratio

				DRM-EE				
(n_0, n_1)		EMP	DRM	$\kappa = 1$	$\kappa = 0.9$	$\kappa = 0.95$	$\kappa = 1.05$	$\kappa = 1.1$
$(50, 50)$	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

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 1 in Section 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.

The sensitivity of the DRM-EE estimator with respect to misspecified auxiliary information can be observed from the last four columns of Table 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 κ moves further away from 1.

4.1.3. Performance of confidence intervals

We compare four CIs for δ:

(i) EMP-NA: Wald-type CI for δ based on the asymptotic normality of $\log\bar{\delta}$;
(ii) EMP-EL: Owen (2001)'s ELR-based CI for δ;
(iii) DRM: the ELR-based CI for δ in (2.8) without auxiliary information;
(iv) DRM-EE: the ELR-based CI for δ in (2.8) 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.

TABLE 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

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

4.1.4. 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 3 and Corollary 2, the ELR statistic has a χ^2_1 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 3 gives the simulated power ($\kappa \neq 1$) and type I error rate ($\kappa = 1$) of the ELR test at the 5% significance level.

We observe from Table 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

TABLE 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

distribution for the ELR test works very well. As κ deviates from 1 and the sample size increases, the power of the test increases, as expected.

4.2. Simulation studies for inferences on quantiles

4.2.1. 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 $\tau = 0.10, 0.25, 0.5, 0.75, 0.90$.

4.2.2. 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 (3.2);
(iv) DRM-EE: our proposed quantile estimator in (3.1) with the common-mean assumption or the EE (1.7) in Example 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 4 presented here is only for the normal setting; the results under gamma distributions are included in Section 7.4.

Table 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

TABLE 4
RB (%) and MSE (×100) for quantile estimators (normal distributions)

(n_0, n_1)	τ		$N(18,4)$ EMP	EL	DRM	DRM-EE	$N(18,9)$ 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

the results agree with Theorem 6. We also notice that the EL and DRM quantile estimators are comparable.

4.2.3. 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., 2022);
(iii) DRM-EE: the proposed ELR-based CI.

The construction of CIs for the quantiles under the two-sample EL method with

Table 5
CP (%) and AL for 95% CIs of 100τ%-quantiles (normal distributions)

			$N(18,4)$			$N(18,9)$		
(n_0, n_1)	τ		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

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 5. The results for the gamma distributions display similar patterns and are included in Section 7.4.

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 3 works very well, and additional auxiliary information leads to shorter CIs.

5. Two 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 δ of the two populations and construct CIs for δ.

To use our proposed method to analyze the dataset, we need to choose an appropriate $\boldsymbol{q}(x)$ in the DRM (1.1). 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 analyze the data using DRM and the other methods discussed in Section 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 4.1.2, 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 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 4.1.3. The LBs of all three CIs are greater than 1, indicating that the seeded clouds slightly increase rainfall.

TABLE 6
Summary of 95% CIs for δ (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 radioimmunoassay (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 7. We set $\boldsymbol{q}(x)$ in the DRM (1.1) 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.

Table 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 use the methods of Section 4.2 to analyze the independent samples. Table 8 summarizes the point estimates and 95% CIs. Note that the EL method does not specify how to construct CIs for quantiles with the common-mean assumption. We also provide the results of analyzing the original 56 pairs using the EMP method; these are recorded under "EMP–ALL" in Table 8 and serve as the benchmarks. Table 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 8
Summary of point estimates and 95% CIs for quantiles (cyclosporine data)

		HPLC ($D = 0$)				RIA ($D = 1$)			
τ		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

6. Discussion

We have proposed new and general semiparametric inference procedures to utilize the combined information from two samples as well as auxiliary information formulated through unbiased EEs. We have established the asymptotic normality of the MELEs of the unknown parameters in the DRMs and/or defined through EEs and the chi-square limiting distributions for the ELR statistics on the parameters. We have also derived efficiency results for estimating these parameters and obtained similar results for inference on the CDFs and population quantiles. We have developed an ELR test for checking the validity and usefulness of auxiliary information, and conducted simulation studies to evaluate the power of the test. Our theoretical results and simulation studies demonstrated that the use of DRMs and auxiliary information leads to improved efficiency of statistical inferences.

We have focused on two-sample data under the DRM (1.1) in the current paper. This leads to many interesting potential research topics. First, we may generalize our results to multiple-sample DRMs (Chen and Liu, 2013) with unbiased EEs. Second, we may study other types of parameters, such as the ROC curve and the area under the curve. Third, in Example 2 (a retrospective case-control study with auxiliary information), it is assumed that the ratio of the total sample size for the internal study to the total sample size for the external study goes to 0. This assumption ensures that the uncertainty of the regression coefficient from the external study is negligible. If the sample sizes of the internal and external studies are comparable, then the variation of the regression coefficient cannot be ignored. Simply discarding the uncertainty may not guarantee efficiency with the auxiliary information (Zhang et al., 2020). We may generalize the method of Zhang et al. (2020) from the one-sample case to case-control studies with uncertainty in the regression coefficient for the external study. We hope to address these problems in future research. Forth, in our asymptotic framework, we assume that $\lambda^* \in (0, 1)$, or both sample sizes n_0 and n_1 go to infinity at the same rate. It would be interesting to study the asymptotic results when $n_1/n \to 0$ or $n_0/n \to 0$. We leave this for future investigation. Fifth, as we observed in Section 4, misspecified auxiliary information may lead to biased results. This is the main motivation for our proposed test on the validity and usefulness of the auxiliary information. It would be desirable to have a robust method to incorporate auxiliary information, which results in improved inference when the auxiliary information is correctly specified and also leads to valid inference when the auxiliary information is misspecified. Li and Tseng (2008) and Li and Wu (2010) proposed two-stage estimators for two-stage design survival data by combining the first and the second stage estimators. The proposed methods improve the second stage estimator even when population distributions in the two stages are different. Chen et al. (2021b) and Zhai and Han (2022) suggested a penalized EL method under a one-sample setting to utilize the auxiliary information from heterogeneous populations. The ideas of two-stage estimators and the penalized EL method maybe potentially useful in our current setup to develop a robust method for combining auxiliary information. We plan to explore them in future research.

7. Appendix

7.1. Examples of summary quantities

In this section, we provide some examples to demonstrate that the EEs

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

can define many important summary quantities.

Example 4 (Means and variances)**.** *Let μ_i and σ_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 in the main paper, we can construct an 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 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 ith population, $i=0,1$. We assume that $GE_i^{(\xi)}$ exists. In our setup, $(GE_0^{(\xi)},GE_1^{(\xi)})^\top$ together with (μ_0,μ_1) can also be defined through the EEs. 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 (μ_0,μ_1) can be defined through $E_0\{\boldsymbol{g}(X;\boldsymbol{\psi},\boldsymbol{\theta})\}=\mathbf{0}$. For other values of ξ, we can define the corresponding EEs similarly.

Applying the results in Theorem 2 in the main paper, 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 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 in the main paper, we can also construct an ELR-based CI for ζ_0 or ζ_1 or an ELR-based confidence region for $(\zeta_0, \zeta_1)^\top$.

Example 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 or 3 in the main paper, we can also construct an ELR-based CI for ξ_{0,τ_0} or ξ_{1,τ_1} or an ELR-based confidence region for $(\xi_{0,\tau_0}, \xi_{1,\tau_1})^\top$.

7.2. 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:

$$h(Y; \alpha_Y, \boldsymbol{\beta}_Y) = P(D = 1|Y) = \frac{\exp(\alpha_Y + \boldsymbol{\beta}_Y^\top Y)}{1 + \exp(\alpha_Y + \boldsymbol{\beta}_Y^\top Y)}. \tag{7.1}$$

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 (7.1). 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 EEs with respect to F_0, which is the setup in the main paper. When the external study is a prospective case-control study, by defining the unknown overall disease prevalence $\pi = P(D = 1)$, we have

$$\begin{aligned} & E\{\boldsymbol{a}(\alpha_Y^*, \boldsymbol{\beta}_Y^*)\} \\ = \; & E\left[\{D - h(Y)\}(1, Y^\top)^\top\right] && (7.2) \\ = \; & E_0\Big([-(1-\pi)h(Y) + \pi \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X)\}\{1 - h(Y)\}](1, Y^\top)^\top\Big), && (7.3) \end{aligned}$$

where we have used the law of total expectation and the DRM (1.1) in the last step.

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

$$\begin{aligned} & E\{\boldsymbol{a}(\alpha_Y^*, \boldsymbol{\beta}_Y^*)\} \\ = \; & -(1-\pi_E)E_0\{h(Y)(1, Y^\top)^\top\} + \pi_E E_1[\{1 - h(Y)\}(1, Y^\top)^\top\}], && (7.4) \end{aligned}$$

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

Using the DRM (1.1), we further get

$$\begin{aligned} & E\{\boldsymbol{a}(\alpha_Y^*, \boldsymbol{\beta}_Y^*)\} \\ = \; & E_0\Big([-(1-\pi_E)h(Y) + \pi_E \exp\{\boldsymbol{\theta}^\top \boldsymbol{Q}(X)\}\{1 - h(Y)\}](1, Y^\top)^\top\Big). && (7.5) \end{aligned}$$

Summarizing (7.3) and (7.5), 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 EEs 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, (7.4) can be equivalently written as

$$\begin{aligned} & E\{\boldsymbol{a}(\alpha_Y^*, \boldsymbol{\beta}_Y^*)\} \\ = \; & E\Bigg[\frac{1-\pi_E}{1-\pi}(1 - D)\{D - h(Y)\}(1, Y^\top)^\top \end{aligned}$$

$$+ \frac{\pi_E}{\pi} D\{D - h(Y)\}(1, Y^\top)^\top \Bigg]. \tag{7.6}$$

Summarizing (7.2) and (7.6), 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.

7.3. Proofs

7.3.1. Regularity conditions

The asymptotic results in this paper 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 $\lambda^* \to n_1/n$ for some constant $\lambda^* \in (0, 1)$.
C2. The two CDFs F_0 and F_1 satisfy the DRM (1.1) 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. $E_0\{\boldsymbol{g}(X; \boldsymbol{\psi}^*, \boldsymbol{\theta}^*)\} = \mathbf{0}$, $E_0\{\partial \boldsymbol{g}(X; \boldsymbol{\psi}^*, \boldsymbol{\theta}^*)/\partial \boldsymbol{\eta}\}$ 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 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 $\ell_{nd}(\boldsymbol{\theta})$ in (2.7) is applicable. 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 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.

7.3.2. Some preliminary results

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

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

where the Lagrange multipliers satisfy

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

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 λ 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 maximum empirical likelihood estimate (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

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

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

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

where

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

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

$$\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{u}^\top} \\ \boldsymbol{0} & \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} & \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\beta}\partial\boldsymbol{u}^\top} \\ \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{u}\partial\boldsymbol{\theta}^\top} & \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{u}\partial\boldsymbol{\beta}^\top} & \frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{u}\partial\boldsymbol{u}^\top} \end{pmatrix}, \tag{7.7}$$

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

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

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

- *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 (1.1) is satisfied as required in Condition C2.

Lemma 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

$$\begin{aligned}E\left\{\sum_{i=0}^{1}\sum_{j=1}^{n_i}\mathcal{S}(X_{ij})\right\} &= n_0E_0\{\mathcal{S}(X)\} + n_1E_0\{\omega(X)\mathcal{S}(X)\} \\ &= n\left[(1-\lambda^*)E_0\{\mathcal{S}(X)\} + \lambda^*E_0\{\omega(X)\mathcal{S}(X)\}\right] \\ &= nE_0\left[\{(1-\lambda^*) + \lambda^*\omega(X)\}\mathcal{S}(X)\right] \\ &= nE_0\{\omega(X)\mathcal{S}(X)\}.\end{aligned}$$

This completes the proof. □

Recall that

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

Applying Lemma 1, after some algebra, we have the following Lemma.

Lemma 2. *(a) With the form of* $\partial^2 l_n(\boldsymbol{\gamma}^*)/(\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top)$ *defined in (7.7), we have*

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

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

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

in distribution with

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

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

Proof. For (a): Note that Conditions C3 and C4 ensure that $\boldsymbol{A}$ is well defined. The results then follow by applying Lemma 1 to each term of $E\{\partial^2 l_n(\boldsymbol{\gamma}^*)/(\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top)\}$. 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 1 and the fact that $h_0(x)h(x) = 1-\lambda^*$, we have

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

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 1 to show that $E(\boldsymbol{S}_n^*) = \mathbf{0}$. For $E(\boldsymbol{S}_{n\theta})$,

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

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

The unbiasedness of the EEs leads to

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

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

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

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

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

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

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

With the help of Lemma 1, we have

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

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

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

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

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

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 $n^{-1/2}\boldsymbol{S}_{n\theta}$ and $n^{-1/2}\boldsymbol{S}_{n\boldsymbol{u}}$:

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

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

7.3.3. Proof of Theorem 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

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

With the law of large numbers and Lemma 2, we have

$$\frac{1}{n}\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top} = \frac{1}{n}E\left\{\frac{\partial^2 l_n(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right\} + o_p(1) = -\boldsymbol{A} + o_p(1). \tag{7.8}$$

Hence, we can write

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_{\theta\theta} \end{pmatrix} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + \begin{pmatrix} \boldsymbol{A}_{\psi\boldsymbol{u}} \\ \boldsymbol{A}_{\theta\boldsymbol{u}} \end{pmatrix} (\hat{\boldsymbol{u}} - \boldsymbol{u}_0) = \frac{1}{n} \begin{pmatrix} \mathbf{0} \\ \boldsymbol{S}_{n\theta} \end{pmatrix} + o_p(n^{-\frac{1}{2}}); \quad (7.9)$$

$$\begin{pmatrix} \boldsymbol{A}_{\boldsymbol{u}\psi} & \boldsymbol{A}_{\boldsymbol{u}\theta} \end{pmatrix} (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) - \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}}(\hat{\boldsymbol{u}} - \boldsymbol{u}_0) = \frac{1}{n}\boldsymbol{S}_{n\boldsymbol{u}} + o_p(n^{-\frac{1}{2}}). \quad (7.10)$$

Recall that

$$\boldsymbol{U} = \begin{pmatrix} \mathbf{0} & \boldsymbol{A}_{\psi\boldsymbol{u}} \\ \boldsymbol{A}_{\theta\theta} & \boldsymbol{A}_{\theta\boldsymbol{u}} \end{pmatrix}, \quad \boldsymbol{V} = \begin{pmatrix} \boldsymbol{A}_{\theta\theta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_{\boldsymbol{u}\boldsymbol{u}} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{J} = \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{U}^\top. \quad (7.11)$$

Conditions C3 and C4 ensure that $\boldsymbol{U}$, $\boldsymbol{V}$, and $\boldsymbol{J}$ have full rank. Then (7.9) and (7.10) 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 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

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

Since

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

we have

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

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

7.3.4. Proof of Corollary 1

Part (a). The results in Theorem 1 imply that

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

in distribution, where

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

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

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

and

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

We write

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

When $r = p$, we have

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

This leads to

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

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

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

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

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

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 EEs of $\boldsymbol{g}(\boldsymbol{x}; \boldsymbol{\eta})$. With Theorem 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

$$\boldsymbol{V}_m^{-1} \geq \begin{pmatrix} \boldsymbol{V}_{m-1}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & 0 \end{pmatrix}. \tag{7.12}$$

Therefore,

$$\begin{aligned} \boldsymbol{J}_m &= \boldsymbol{U}_m \boldsymbol{V}_m^{-1} \boldsymbol{U}_m^\top \\ &\geq (\boldsymbol{U}_{m-1}, u_m) \begin{pmatrix} \boldsymbol{V}_{m-1}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & 0 \end{pmatrix} (\boldsymbol{U}_{m-1}, u_m)^\top \\ &= \boldsymbol{J}_{m-1}, \end{aligned} \tag{7.13}$$

as required. This completes the proof.

7.3.5. Proof of Theorem 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

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

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 (7.8) and Condition C5, we have

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

$$-n(\boldsymbol{\eta}-\boldsymbol{\eta}^*)^\top \begin{pmatrix} \boldsymbol{A}_{\psi\boldsymbol{u}} \\ \boldsymbol{A}_{\theta\boldsymbol{u}} \end{pmatrix} (\boldsymbol{u}-\boldsymbol{u}^*) + o_p(1).$$

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

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

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

$$\begin{aligned} \ell_n(\boldsymbol{\psi},\boldsymbol{\theta}) &= l_n(\boldsymbol{\gamma}^*) + (\boldsymbol{\eta}-\boldsymbol{\eta}^*)^\top \boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{S}_n^* - \frac{n}{2}(\boldsymbol{\eta}-\boldsymbol{\eta}^*)^\top \boldsymbol{J}(\boldsymbol{\eta}-\boldsymbol{\eta}^*) \\ &\quad -\frac{1}{2n}\boldsymbol{S}_{n\boldsymbol{u}}^\top \boldsymbol{A}_{\boldsymbol{uu}}^{-1}\boldsymbol{S}_{n\boldsymbol{u}} + o_p(1). \end{aligned} \tag{7.14}$$

With the approximation of $\hat{\boldsymbol{\eta}}$ in (7.14), we then have

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

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

$$\frac{\partial \ell_n^*(\check{\boldsymbol{\psi}},\check{\boldsymbol{\theta}},\check{\boldsymbol{v}})}{\partial \boldsymbol{\psi}} = \boldsymbol{0}, \quad \frac{\partial \ell_n^*(\check{\boldsymbol{\psi}},\check{\boldsymbol{\theta}},\check{\boldsymbol{v}})}{\partial \boldsymbol{\theta}} = \boldsymbol{0}, \quad \frac{\partial \ell_n^*(\check{\boldsymbol{\psi}},\check{\boldsymbol{\theta}},\check{\boldsymbol{v}})}{\partial \boldsymbol{v}} = \boldsymbol{0}. \tag{7.15}$$

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 (7.14), we approximate $\ell_n^*(\boldsymbol{\psi},\boldsymbol{\theta},\boldsymbol{v})$ at an $n^{-1/2}$ neighbor of $(\boldsymbol{\psi}_0^\top, \boldsymbol{\theta}_0^\top, \boldsymbol{0}_{1\times q})^\top$ with

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

Applying the first-order Taylor expansion to (7.15), we have

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

Hence,

$$n^{1/2}(\check{\boldsymbol{\eta}}-\boldsymbol{\eta}^*)$$

$$
\begin{aligned}
&= (\boldsymbol{I},\boldsymbol{0})\begin{pmatrix}\boldsymbol{J} & -\boldsymbol{h}^{*\top}\\ -\boldsymbol{h}^{*\top} & \boldsymbol{0}\end{pmatrix}^{-1}\begin{pmatrix} n^{-1/2}\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{S}_n\\ \boldsymbol{0}\end{pmatrix} + o_p(1)\\
&= \{\boldsymbol{J}^{-1} - \boldsymbol{J}^{-1}\boldsymbol{h}^{*\top}(\boldsymbol{h}^*\boldsymbol{J}^{-1}\boldsymbol{h}^{*\top})^{-1}\boldsymbol{h}^*\boldsymbol{J}^{-1}\}\boldsymbol{U}\boldsymbol{V}^{-1}(n^{-1/2}\boldsymbol{S}_n^*) + o_p(1), \quad (7.16)
\end{aligned}
$$

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

Substituting the expression of $\check{\boldsymbol{\eta}}$ in (7.16) into (7.14) gives

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

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 χ_q^2 under H_0.

7.3.6. Proofs of Theorem 3 and Corollary 2

We start with the proof of Theorem 3. Recall that the ELR statistic for testing the validity of the EEs 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 $\boldsymbol{\nu}^* = \boldsymbol{0}$ implies $\ell_{nd}(\boldsymbol{\theta}^*) = l_n(\boldsymbol{\gamma}^*)$. According to Qin and Zhang (1997), it is easy to verify that

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

and

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

Then

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

Hence, the ELR statistic can be written as

$$\begin{aligned} W_n &= 2\left\{\ell_{nd}(\tilde{\boldsymbol{\theta}}) - \ell_n(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}})\right\} \\ &= \frac{1}{n}\boldsymbol{S}_{n\boldsymbol{\theta}}^\top \boldsymbol{A}_{\theta\theta}^{-1}\boldsymbol{S}_{n\boldsymbol{\theta}} + \frac{1}{n}\boldsymbol{S}_{n\boldsymbol{u}}^\top \boldsymbol{A}_{\boldsymbol{uu}}^{-1}\boldsymbol{S}_{n\boldsymbol{u}} - \frac{1}{n}\boldsymbol{S}_n^{*\top}\boldsymbol{V}^{-1}\boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U}\boldsymbol{V}^{-1}\boldsymbol{S}_n^* \\ &= \frac{1}{n}\boldsymbol{S}_n^{*\top}\boldsymbol{V}^{-1}(\boldsymbol{V} - \boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U})\boldsymbol{V}^{-1}\boldsymbol{S}_n^* + o_p(1). \end{aligned} \tag{7.17}$$

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

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

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

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

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, 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 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 $r-p$ and $\boldsymbol{C}^{*\top}\boldsymbol{V}^{-1}\boldsymbol{C} = \boldsymbol{0}$, we have

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

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

We now prove Corollary 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

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

Following the techniques used to obtain (7.17), 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:

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

With the technique used to prove Corollary 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 $n \to \infty$.

7.3.7. *Proof of Theorem 4*

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

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

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

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 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 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*

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

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

Note that $\boldsymbol{B}_{0\boldsymbol{\theta}}(x)$, $\boldsymbol{B}_{1\boldsymbol{\theta}}(x)$, $\boldsymbol{B}_{0\boldsymbol{u}}(x)$, and $\boldsymbol{B}_{1\boldsymbol{u}}(x)$ have been defined before Theorem 4. 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 3, we have

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

Using the relationship in (7.10) and the definitions of the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in (7.11), we have

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

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 1, we have

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

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

$$\boldsymbol{B}_{0\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}} = E_0\left\{\frac{\omega(X)I(X \le x)}{h(X)}\right\} - E_0\left\{\frac{I(X \le x)}{h(X)}\right\},$$

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

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

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

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

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

and

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

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

$$nCov\{F_l(x, \boldsymbol{\gamma}^*),\ F_s(y, \boldsymbol{\gamma}^*)\} \quad = \quad E_0\left\{\frac{\omega^{l+s}(X) I(X \le x \wedge y)}{h(X)}\right\} - F_l(x)F_s(y)$$

$$-\lambda^*(1-\lambda^*)\boldsymbol{B}_{l\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{s\boldsymbol{u}}(y). \quad (7.18)$$

Next, we consider the cross-terms with $\boldsymbol{S}_n^*$. We present the calculation of $Cov\{F_0(x,\boldsymbol{\gamma}^*),\boldsymbol{S}_n^*\}$ as an illustration. Using Lemma 1, we get

$$\begin{aligned}
& Cov\{F_0(x,\boldsymbol{\gamma}^*),\boldsymbol{S}_{n\theta}\} \\
= \ & \frac{1}{n}Cov\left\{\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{I(X_{ij}\le x)}{h(X_{ij})},\ \sum_{j=1}^{n_1}h_0(X_{1j})\boldsymbol{Q}(X_{1j})^\top - \sum_{j=1}^{n_0}h_1(X_{0j})\boldsymbol{Q}(X_{0j})^\top\right\} \\
= \ & \lambda^* Cov\left\{\frac{I(X_{11}\le x)}{h(X_{11})},\ h_0(X_{11})\boldsymbol{Q}(X_{11})^\top\right\} \\
& -(1-\lambda^*)Cov\left\{\frac{I(X_{01}\le x)}{h(X_{01})},\ h_1(X_{01})\boldsymbol{Q}(X_{01})^\top\right\} \\
= \ & \left[E_0\{h_0(X)I(X\le x)\} - \frac{1-\lambda^*}{\lambda^*}E_0\{h_1(X)I(X\le x)\}\right]E_0\{h_1(X)\boldsymbol{Q}(X)^\top\}.
\end{aligned}$$

It can be checked that

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

Then we have

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

Similarly,

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

where in the last step we used the facts that

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

Recall that

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

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

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

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

$$\begin{aligned} & n^{-1}Var(\boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{S}_n^*) - 2Cov\left\{F_l(x,\boldsymbol{\gamma}^*), \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{S}_n^*\right\} \\ = & \left[\boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{\Gamma} - 2Cov\left\{F_l(x,\boldsymbol{\gamma}^*), \boldsymbol{S}_n^*\right\}\right]\boldsymbol{W}\boldsymbol{B}_l^*(x) \\ = & \left[\boldsymbol{B}_l^*(x)^\top \boldsymbol{V}^{-1}\boldsymbol{U}^\top \boldsymbol{J}^{-1}\boldsymbol{U} + \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{B}_{l\boldsymbol{u}}(x) \end{pmatrix}^\top + \boldsymbol{B}_{l\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{C}^\top\right]\boldsymbol{W}\boldsymbol{B}_l^*(x) \\ = & \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{B}_l^*(x) + \lambda^*(1-\lambda^*)\boldsymbol{B}_{l\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{l\boldsymbol{u}}(x). \end{aligned} \quad (7.19)$$

Combining (7.18) and (7.19) leads to

$$\sigma_{ll}(x,x) = E_0\left\{\frac{\omega^{2l}(X)I(X \le x)}{h(X)}\right\} - F_l(x)^2 + \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{B}_l^*(x). \quad (7.20)$$

Using similar steps to derive (7.19), we find that the summation of the last three terms in $\sigma_{ls}(x,y)$ is

$$\boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{B}_s^*(y) + \lambda^*(1-\lambda^*)\boldsymbol{B}_{l\boldsymbol{u}}(x)^\top \boldsymbol{e}_{\boldsymbol{u}}\boldsymbol{e}_{\boldsymbol{u}}^\top \boldsymbol{B}_{s\boldsymbol{u}}(y). \quad (7.21)$$

Combining (7.18) and (7.21) gives

$$\sigma_{ls}(x,y) = E_0\left\{\frac{\omega^{l+s}(X)I(X \le x \wedge y)}{h(X)}\right\}$$

$$-F_l(x)F_s(y) + \boldsymbol{B}_l^*(x)^\top \boldsymbol{W}\boldsymbol{B}_s^*(y). \tag{7.22}$$

Summarizing (7.20) and (7.22), we conclude that for any $i, j \in \{l, s\}$

$$\begin{aligned}\sigma_{ij}(x,y) \quad &= \quad E_0\left\{\frac{\omega^{i+j}(X)I(X \le x \wedge y)}{h(X)}\right\} \\ &\quad -F_i(x)F_j(y) + \boldsymbol{B}_i^*(x)^\top \boldsymbol{W}\boldsymbol{B}_j^*(y),\end{aligned} \tag{7.23}$$

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

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

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

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

$$= \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\theta\theta}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} - \{\lambda^*(1-\lambda^*)\}^{-1} \begin{pmatrix} \boldsymbol{e}_{\boldsymbol{\theta}} \\ -\lambda^*(1-\lambda^*)\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix} \begin{pmatrix} \boldsymbol{e}_{\boldsymbol{\theta}} \\ -\lambda^*(1-\lambda^*)\boldsymbol{e}_{\boldsymbol{u}} \end{pmatrix}^\top .$$

Substituting $\boldsymbol{W}$ into (7.23) and using the fact that

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

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

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

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 EEs 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 EEs.

From the definitions of these matrices and vectors, we notice the relationships: $\boldsymbol{U}_m = (\boldsymbol{U}_{m-1}, u_m)$ and

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

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

With the fact that

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

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

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

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 (7.12) and (7.13), we have

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

This implies that $\boldsymbol{\Sigma}_{ls}^{(m-1)}(x,y) - \boldsymbol{\Sigma}_{ls}^{(m)}(x,y) \geq \boldsymbol{0}$. This completes the proof of (c).

7.3.8. Proof of Theorem 5

We first introduce two lemmas that will be helpful in the proof of Theorem 5. The following lemma establishes the convergence rate of $\hat{\xi}_{i,\tau}$.

Lemma 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., 2021a)

$$\Delta_n = O_p(n^{-1/2}). \tag{7.24}$$

Define

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

Then

$$\Delta_n = \sup_x |\hat{F}_0(x) - F_0(x)|$$

$$\begin{aligned} &\leq \sup_x |\hat{F}_0(x) - \bar{F}_0(x)| + \sup_x |\bar{F}_0(x) - F_i(x)| \\ &= \Delta_{n1} + \Delta_{n2}, \end{aligned}$$

where

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

and

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

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

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

With this result, the claim (7.24) is proved if $\Delta_{n1} = O_p(n^{-1/2})$.

As preparation, we argue that

$$(n\hat{p}_{ij})^{-1} = 1 + \hat{\lambda}[\exp\{\hat{\boldsymbol{\theta}}^\top \boldsymbol{Q}(X_{ij})\} - 1] + \hat{\boldsymbol{\nu}}^\top \boldsymbol{g}(X_{ij}; \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\theta}}) \geq 1 - \lambda^* + o_p(1) \quad (7.25)$$

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

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

By Condition C5,

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

which, together with $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^* = O_p(n^{-1/2})$, implies that (7.25) is valid.

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

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

Using (7.25), we have

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

$$+O_p(1/n)\sum_{i=0}^{1}\sum_{j=1}^{n_i}\frac{|\hat{\boldsymbol{\nu}}^\top \boldsymbol{g}(X_{ij};\hat{\boldsymbol{\psi}},\hat{\boldsymbol{\theta}})|}{1-\lambda^*}I(X_{ij}\le x). \qquad (7.26)$$

By Condition C5,

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

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 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 $x\ge\xi_{0,\tau}$. Note that

$$\begin{aligned}&|\{\hat{F}_0(x)-\hat{F}_0(\xi_{0,\tau})\}-\{F_0(x)-F_0(\xi_{0,\tau})\}|\\ \le\ &|\{\hat{F}_0(x)-\hat{F}_0(\xi_{0,\tau})\}-\{\bar{F}_0(x)-\bar{F}_0(\xi_{0,\tau})\}|\\ &+|\{\bar{F}_0(x)-\bar{F}_0(\xi_{0,\tau})\}-\{F_0(x)-F_0(\xi_{0,\tau})\}|. \qquad (7.27)\end{aligned}$$

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

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

Consequently, we need to show only that the first term in (7.27) has a higher order than $n^{-3/4}(\log(n))^{1/2}$ uniformly in $0\le x-\xi_{0,\tau}<cn^{-1/2}$.

With the technique used to obtain (7.26), we have

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

By Condition C5, $E_0\{1+R^{1/3}(X)\}<\infty$ and $E_1\{1+R^{1/3}(X)\}<\infty$, then uniformly in x

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

and

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

Therefore,

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

This completes the proof. □

We are now ready to prove Theorem 5. By Lemma 4, for $i = 0, 1$,

$$\begin{aligned} F_i(\hat{\xi}_{i,\tau}) - F_i(\xi_{i,\tau}) &= f_i(\xi_{i,\tau})(\hat{\xi}_{i,\tau} - \xi_{i,\tau}) + O_p((\hat{\xi}_{i,\tau} - \xi_{i,\tau})^2) \\ &= f_i(\xi_{i,\tau})(\hat{\xi}_{i,\tau} - \xi_{i,\tau}) + O_p(n^{-1}). \end{aligned} \tag{7.28}$$

Note that $\hat{F}_i(\hat{\xi}_{i,\tau}) = \tau + O_p(n^{-1})$. Replacing x by $\hat{\xi}_{i,\tau}$ in Lemma 5 and using (7.28) 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.

7.3.9. Proof of Theorem 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 4(c) implies the results in (c). This completes the proof.

7.4. Additional simulation under the gamma distributional setting

Table 9 presents the four quantile estimates under gamma distributions. Table 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 9
RB (%) and MSE (×100) for quantile estimators (gamma distributions)

(n_0, n_1)	τ		$\Gamma(8, 1.125)$ EMP	EL	DRM	DRM-EE	$\Gamma(6, 1, 5)$ 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

References

[1] Anderson, J. A. (1979). Multivariate logistic compounds. *Biometrika*, 66:17–26. MR0529143

[2] Bebu, I. and Mathew, T. (2008). Comparing the means and variances of a bivariate log-normal distribution. *Statistics in Medicine*, 27:2684–2696. MR2440059

[3] 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. MR3674695

[4] 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*

Table 10
CP (%) and AL for three 95% CIs of 100τ%-quantile (gamma distributions)

			$\Gamma(8, 1.125)$			$\Gamma(6, 1.5)$		
(n_0, n_1)	τ		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

American Statistical Association, 111:107–117. MR3494641

[5] 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. MR3538711

[6] Chen, J., Li, P., Liu, Y., and Zidek, J. V. (2021a). Composite empirical likelihood for multisample clustered data. *Journal of Nonparametric Statistics*, 33:60–81. MR4261898

[7] Chen, J. and Liu, Y. (2013). Quantile and quantile-function estimations under density ratio model. *The Annals of Statistics*, 41:1669–1692. MR3113825

[8] Chen, Z., Ning, J., Shen, Y., and Qin, J. (2021b). Combining primary cohort

data with external aggregate information without assuming comparability. *Biometrics*, 77:1024–1036. MR4320675

[9] de Carvalho, M. and Davison, A. C. (2014). Spectral density ratio models for multivariate extremes. *Journal of the American Statistical Association*, 109:764–776. MR3223748

[10] Fokianos, K., Kedem, B., Qin, J., and Short, D. A. (2001). A semiparametric approach to the one-way layout. *Technometrics*, 43:56–65. MR1819908

[11] Hawkins, D. M. (2002). Diagnostics for conformity of paired quantitative measurements. *Statistics in Medicine*, 21:1913–1935. MR2900861

[12] Imbens, G. W. and Lancaster, T. (1994). Combining micro and macro data in microeconometric models. *The Review of Economic Studies*, 61:655–680. MR1299309

[13] Kay, R. and Little, S. (1987). Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika*, 74:495–501. MR0909354

[14] 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. MR2384498

[15] 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. MR1972942

[16] Li, G. and Tseng, C.-H. (2008). Non-parametric estimation of a survival function with two-stage design studies. *Scandinavian Journal of Statistics*, 35:193–211. MR2418736

[17] Li, G. and Wu, T. T. (2010). Semiparametric additive risks regression for two-stage design survival studies. *Statistica Sinica*, 20:1581–1607. MR2777337

[18] 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. MR3843042

[19] Li, P., Liu, Y., and Qin, J. (2017). Semiparametric inference in a genetic mixture model. *Journal of the American Statistical Association*, 112:1250–1260. MR3735374

[20] 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. MR2896850

[21] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75:237–249. MR0946049

[22] Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC, Boca Raton.

[23] Qin, J. (1999). Empirical likelihood ratio based confidence intervals for mixture proportions. *The Annals of Statistics*, 27:1368–1384. MR1740107

[24] Qin, J. (2017). *Biased Sampling, Over-identified Parameter Problems and Beyond*. Springer, Singapore. MR3675467

[25] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22:300–325. MR1272085
[26] Qin, J. and Lawless, J. (1995). Estimating equations, empirical likelihood and constraints on parameters. *The Canadian Journal of Statistics*, 23:145–159. MR1345462
[27] Qin, J. and Zhang, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, 84:609–618. MR1603924
[28] Qin, J. and Zhang, B. (2003). Using logistic regression procedures for estimating receiver operating characteristic curves. *Biometrika*, 90:585–596. MR2006837
[29] 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. MR3335103
[30] Simpson, J., Olsen, A., and Eden, J. C. (1975). A Bayesian analysis of a multiplicative treatment effect in weather modification. *Technometrics*, 17:161–166.
[31] 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. MR2267709
[32] Wang, C., Marriott, P., and Li, P. (2017). Testing homogeneity for multiple nonnegative distributions with excess zero observations. *Computational Statistics & Data Analysis*, 114:146–157. MR3660845
[33] 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. MR3799642
[34] Wu, C. and Thompson, M. E. (2020). *Sampling Theory and Practice*. Springer, Cham. MR4180686
[35] 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. MR1908510
[36] Yuan, M., Li, P., and Wu, C. (2021). Semiparametric inference of the Youden index and the optimal cut-off point under density ratio models. *The Canadian Journal of Statistics*, 49:965–986. MR4303199
[37] Yuan, M., Wang, C., Lin, B., and Li, P. (2022). Semiparametric inference on general functionals of two semicontinuous populations. *Annals of the Institute of Statistical Mathematics*, 74:451–472. MR4417367
[38] Zhai, Y. and Han, P. (2022). Data integration with oracle use of external information from heterogeneous populations. *Journal of Computational and Graphical Statistics*, In press.
[39] Zhang, A. G., Zhu, G., and Chen, J. (2022). Empirical likelihood ratio test on quantiles under a density ratio model. *Electronic Journal of Statistics*, 15:6191–6227. MR4355706
[40] Zhang, B. (2000). Quantile estimation under a two-sample semi-parametric model. *Bernoulli*, 6:491–511. MR1762557
[41] Zhang, H., Deng, L., Schiffman, M., Qin, J., and Yu, K. (2020). Generalized integration model for improved statistical inference by leveraging external

summary data. *Biometrika*, 107:689–703. MR4138984

[42] 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.

[43] Zhuang, W., Hu, B., and Chen, J. (2019). Semiparametric inference for the dominance index under the density ratio model. *Biometrika*, 106:229–241. MR3912393

[44] Zou, F., Fine, J. P., and Yandell, B. S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika*, 89:61–75. MR1888346