

Comparison of empirical likelihood and **its dual likelihood** under density ratio model

Huapeng Li^{1,2}, Yang Liu¹, Yukun Liu^{1*}, and Riquan Zhang¹

¹School of Statistics, East China Normal University,
Shanghai, 200241, P. R. China

²School of Mathematics and Computer Sciences,
Shanxi Datong University, Datong, 037009, P. R. China

Abstract

Density ratio models (DRM) are commonly-used semiparametric models to link related populations. Empirical likelihood (EL) under DRM has been demonstrated to be a flexible and useful platform for semiparametric inferences. Since DRM-based EL has the same maximum point and maximum likelihood as its **dual form (dual EL)**, EL-based inferences under DRM are usually made through the latter. A natural question comes up: is there any efficiency loss of doing so? We make a careful comparison of the dual EL and DRM-based EL estimation methods from theory and numerical simulations. We find that their point estimators for any parameter are exactly the same, while they may have different performances in interval estimation. In terms of coverage accuracy, the two intervals are comparable for non or moderate skewed populations, and the DRM-based EL interval can be much superior for severely skewed populations. A real data example is analyzed for illustration purpose.

Key words: *Coverage accuracy; Density ratio model; Dual empirical likelihood; Empirical likelihood.*

AMS 2000 Subject Classifications: *Primary 62G15, Secondary 62G20.*

Short title: *Comparison of EL and Dual EL*

*Corresponding author. School of Statistics, East China Normal University, Shanghai, 200241, China. E-mail: ykliu@sfs.ecnu.edu.cn.

1 Introduction

Density ratio models (DRM) originated in logistic discriminant analysis (Anderson 1972,1974), and were formally proposed by Anderson (1979) to link two or more closely-related populations, which have similar characteristics. For example, the modulus of elasticity (MOE) of lumbers in Canada in 2007 may have similar distributions to those in 2010 (Chen and Liu 2013). The DRM postulates a completely parametric model on the density ratio of the two populations under study, and leaves the baseline distribution unspecified. To be specific, suppose we have samples $\{x_{ij} : j = 1, \dots, n_i\}$ ($i = 0, 1$) from two populations with cumulative distribution functions $F_0(x)$ and $F_1(x)$, respectively. The DRM assumes

$$dF_1(x)/dF_0(x) = \exp\{\theta^\top q(x)\}, \quad (1)$$

where $q(x)$ is a pre-specified d -dimensional basis function and θ a d -variate unknown parameter. We require the first element of $q(x)$ to be one so that the first element of θ is a normalization parameter.

The DRM is so flexible that it includes many commonly-used parametric distribution families as special cases. For example, the binomial and exponential distribution families, and the normal distribution family with a common variance parameter correspond to a DRM with $q(x) = (1, x)^\top$. The whole normal and Gamma distribution families correspond to DRMs with $q(x) = (1, x, x^2)^\top$ and $(1, x, \log(x))^\top$, respectively. See Kay and Little (1987) for more examples. The DRM can also be regarded as a semiparametric extension of these parametric probability models. With less model assumptions, DRM-based inferences can be less sensitive to model mis-specification than those based on parametric models.

As a semiparametric model, the DRM leaves the baseline distribution $F_0(x)$ completely unspecified. Owen (1988, 1990)'s empirical likelihood (EL) is a very suitable tool to handle the non-parametric baseline distribution. Since Owen's seminal paper, the EL has become remarkably popular because it has many nice properties paralleling to the parametric likelihood methods, e.g., it is range-preserving, transform-respect and Bartlett correctable (Hall and La Scala 1990; DiCiccio, Hall and Romano 1991; Owen 2001). It can be dated back to Vardi (1982, 1985) and Gill, Vardi and Wellner (1988) that the EL approach was applied to nonparametric parameter estimation problems under DRM, although it had not been refined into a universal method by then. Here and later on, the DRM was used to characterize length bias or more general selection bias (Qin 1993, 1999).

The DRM-based EL has attracted much attention in the past decades. Qin and Zhang (1997) showed that the logistic regression model commonly used in case-control studies can be described by the DRM. They studied the EL approach for parameter estimation and for goodness-of-fit tests of the regression model. Qin (1998) linked the prospective likelihood for case-control data to the DRM-based EL. Fokianos, Kedem, Qin and Short (2001) used the DRM-based EL approach for a classical one-way analysis-of-variance. Zhang (2000, 2002) investigated the DRM-based EL approach for quantile estimation and goodness-of-fit. Fokianos (2004) proposed to merge information from multiple samples by the DRM-based EL approach to construct more efficient density estimators for the multiple unknown populations. Although the common statistical tool in these developments is the DRM-based EL, the parameter estimation is often based on the so-called dual EL (Keziou and Leoin-Aubin 2008). Fortunately, Keziou and Leoin-Aubin (2008) formally established the equivalence of the maximum DRM-based EL estimators and the maximum dual EL estimators for both θ and F_0 in (1). Since the dual EL has a closed form and is easy to calculate, this makes the calculation and application of the DRM-based EL approach much convenient. Under this formulation, Chen and Liu (2013) disclosed that the DRM-based EL quantile estimator admits Bahadur representation, and Cai, Chen and Zidek (2017) studied hypothesis testing problems with the dual EL ratio tests.

Although it shares many properties with the DRM-based EL, the dual EL is not a real likelihood, let alone the real likelihood based on the data. Unlike the dual EL, the standard EL can not only use the constraint $\int \exp\{\theta^\top q(x)\}dF_0(x) = 1$ under DRM, but also easily incorporate auxiliary information defined through additional estimating equations (Qin and Lawless 1994). After modelling the case and control data with the logistic regression model or equivalently DRM, Qin et al. (2015) proposed using EL to increase the power of case-control studies by incorporating covariate-specific disease prevalence information. Under DRMs on the non-zero parts of multiple zero-inflated populations, Wang (2017) constructed EL ratio tests for the means after transforming the testing problems into general estimating equations.

Naturally we may wonder whether inferences based on dual EL are always the same as those based on the standard EL under DRM. If yes, we shall recommend the former instead of the latter since the former

is more convenient to use. Otherwise, when do they perform similarly and when does the former lose remarkable efficiency? These concerns motivate us to make a careful comparison of the dual EL and the standard EL estimations based on DRM. For simplicity, we consider the estimation problem of a general parameter ψ , which can be written as $\psi = \int u(x, \theta) dF_0(x)$ for a known function $u(\cdot, \cdot)$, and F_0 and θ the same as in (1). If only the mean of F_0 is of interest, we can choose $u(x) = x$. If the mean difference of the two populations is of interest, we can choose $u(x, \theta) = x - x \exp\{\theta^\top q(x)\}$. For the parameter ψ , our theoretical comparison indicates that their point estimators are still identical, however our numerical comparison shows that the corresponding interval estimators may have different performances, especially when the underlying populations are severely skewed.

The rest of the paper is organized as follows. In section 2, we briefly review the dual EL and present the asymptotical normality of the maximum dual EL estimators of ψ . Section 3 is devoted to the standard EL under DRM. We establish the limiting distributions of the maximum EL estimator and the likelihood ratio. A simulation study is provided in Section 4 and a real data example is presented in Section 5. All technical proofs are postponed to the Appendix.

2 Dual empirical likelihood

The dual EL is induced from the EL under DRM. Given the data $\{x_{ij} : j = 1, \dots, n_i; i = 0, 1\}$, let $p_{ij} = dF_0(x_{ij})$ with $dF_0(x) = F_0(x) - F_0(x-)$. Under the DRM (1), $dF_1(x_{ij}) = \exp\{\theta^\top q(x_{ij})\} dF_0(x_{ij}) = \exp\{\theta^\top q(x_{ij})\} p_{ij}$. The EL of (F_0, F_1) under the DRM (1) is

$$L(F_0, F_1) = \prod_{i=1}^{n_0} dF_0(x_{0i}) \prod_{j=1}^{n_1} dF_1(x_{1j}) = \left(\prod_{i=0}^1 \prod_{j=1}^{n_i} p_{ij} \right) \exp\left\{ \sum_{j=1}^{n_1} \theta^\top q(x_{1j}) \right\}. \quad (2)$$

The feasible p_{ij} 's satisfy

$$p_{ij} \geq 0, \quad \sum_{i,j} p_{ij} = 1, \quad \sum_{i,j} \exp\{\theta^\top q(x_{ij})\} p_{ij} = 1, \quad (3)$$

where $\sum_{i=0}^1 \sum_{j=1}^{n_i}$ is written as $\sum_{i,j}$ for short.

Let $n = n_0 + n_1$ be the total sample size. By the Lagrange multiplier method, maximizing $\log\{L(F_0, F_1)\}$ with respect to p_{ij} 's under the constraint (3) gives

$$p_{ij} = \frac{1}{n} \frac{1}{1 + \hat{\lambda}[\exp\{\theta^\top q(x_{ij})\} - 1]},$$

where $\hat{\lambda} = \hat{\lambda}(\theta)$ is the solution to

$$\frac{1}{n} \sum_{i,j} \frac{\exp\{\theta^\top q(x_{ij})\}}{1 + \hat{\lambda}[\exp\{\theta^\top q(x_{ij})\} - 1]} = 1.$$

Accordingly, the profile empirical log-likelihood of θ (up to a constant) is

$$\ell_1^*(\theta) = - \sum_{i,j} \log\{1 + \hat{\lambda}[\exp\{\theta^\top q(x_{ij})\} - 1]\} + \sum_{j=1}^{n_1} \{\theta^\top q(x_{1j})\}.$$

This profile log-likelihood has the same maximum value and maximum point as another function

$$\ell_1(\theta) = - \sum_{i,j} \log\{1 + \rho[\exp\{\theta^\top q(x_{ij})\} - 1]\} + \sum_{j=1}^{n_1} \{\theta^\top q(x_{1j})\},$$

where $\rho = n_1/n$ takes the place of $\hat{\lambda}$ in $\ell_1^*(\theta)$. See Keziou and Leoin-Aubin (2008) and Chen and Liu (2013). We call $\ell_1(\theta)$ dual EL as Keziou and Leoin-Aubin (2008) pointed out that it is a dual likelihood of the EL.

Compared with $\ell_1^*(\theta)$, the dual EL $\ell_1(\theta)$ has a closed form and is easier to maximize. Denote the maximum dual EL estimator of θ by $\hat{\theta} = \arg \max_{\theta} \ell_1(\theta)$. The maximum dual EL estimators of $F_0(x)$ and $F_1(x)$ are respectively

$$\hat{F}_0(x) = \sum_{i,j} \hat{p}_{ij} I(x_{ij} \leq x) \quad \text{and} \quad \hat{F}_1(x) = \sum_{i,j} \hat{p}_{ij} \exp\{\hat{\theta}^\top q(x_{ij})\} I(x_{ij} \leq x),$$

where $I(\cdot)$ is the indicator function and $\hat{p}_{ij} = n^{-1} \{1 + \rho[\exp\{\hat{\theta}^\top q(x_{ij})\} - 1]\}^{-1}$. Since the parameter of interest $\psi = \int u(x, \theta) dF_0(x)$ is a function of θ and F_0 , its maximum dual EL estimator is

$$\hat{\psi} = \sum_{i,j} \hat{p}_{ij} u(x_{ij}, \hat{\theta}) = \sum_{i,j} \frac{1}{n} \frac{u(x_{ij}, \hat{\theta})}{1 + \rho[\exp\{\hat{\theta}^\top q(x_{ij})\} - 1]}. \quad (4)$$

The asymptotical normalities of $\hat{\theta}$, $\hat{F}_0(x)$ and $\hat{F}_1(x)$ have been extensively studied in the literature. See Qin and Zhang (1997), Qin (1998), Zhang (2000, 2002), Keziou and Leoin-Aubin (2008) for the two sample case and Chen and Liu (2013) for the multiple sample case. We show that $\hat{\psi}$ also follows an asymptotically normal distribution.

For ease of presentation, we define some notations. Let θ_0 and ψ_0 be the true values of θ and ψ , respectively. Define $h(x, \theta) = 1 - \rho + \rho \exp\{\theta^\top q(x)\}$, $h_1(x, \theta) = \rho \exp\{\theta^\top q(x)\}/h(x, \theta)$ and $h_0(x, \theta) = 1 - h_1(x, \theta)$. We may drop θ_0 for notional simplicity. Let $u_\theta(x, \theta) = \partial u(x, \theta)/\partial \theta$ and e_1 be the d -variate vector $(1, 0, \dots, 0)^\top$. Define

$$\sigma^2 = \int \frac{u^2(x)}{h(x)} dF_0(x) + B^\top W^{-1} B - \psi_0^2 - \frac{\Delta^2}{\rho(1-\rho)}, \quad (5)$$

where $W = \int q(x)q^\top(x)h_0(x)h_1(x)h(x)dF_0(x)$, $B = \int \{h_1(x)u(x)q(x) - u_\theta(x)\}dF_0(x)$, and $\Delta = \rho\psi_0 - \int e_1^\top u_\theta(x)dF_0(x)$. We make the following assumptions on the populations and the DRM.

- (C1) The population distributions F_0 and F_1 satisfy the DRM (1) with true parameter θ_0 and $\int h(x, \theta)dF_0(x) < \infty$ in a neighborhood of θ_0 .
- (C2) The components of $q(x)$ are linearly independent and its first element is one.
- (C3) $n_1/n = \rho + o(1)$ for a constant $\rho \in (0, 1)$.

Theorem 1 *Suppose assumptions (C1)-(C3) are satisfied. As $n \rightarrow \infty$, $\sqrt{n}(\hat{\psi} - \psi_0)$ is asymptotically normal with mean 0 and variance σ^2 , which is defined in (5).*

The assumption that $\int h(x, \theta)dF_0(x) < \infty$ in a neighborhood of θ_0 implies the existence of the moment generating function of $q(x)$ and therefore all its finite moments. This together with the linearly independence of the elements of $q(x)$ guarantees that W is positive definite.

To construct interval estimators for ψ based on Theorem 1, we need a consistent estimator of σ^2 . The analytical form of σ^2 motivates us to estimate it by

$$\hat{\sigma}^2 = \int \frac{u^2(x, \hat{\theta})}{h(x, \hat{\theta})} d\hat{F}_0(x) + \hat{B}^\top \hat{W}^{-1} \hat{B} - \hat{\psi}^2 - \frac{\hat{\Delta}^2}{\rho(1-\rho)},$$

where $\hat{\Delta} = \rho\hat{\psi} - \int e_1^\top u_\theta(x, \hat{\theta})d\hat{F}_0(x)$ and

$$\begin{aligned} \hat{B} &= \int \{h_1(x, \hat{\theta})u(x, \hat{\theta})q(x) - u_\theta(x, \hat{\theta})\}d\hat{F}_0(x), \\ \hat{W} &= \int q(x)q^\top(x)h_0(x, \hat{\theta})h_1(x, \hat{\theta})h(x, \hat{\theta})d\hat{F}_0(x). \end{aligned}$$

The consistencies of $\hat{\theta}$ and $\hat{F}_0(x)$ implies the consistency of $\hat{\sigma}^2$. Hence a Wald interval estimator for ψ at level $1 - \alpha$ is

$$\hat{\psi} \pm n^{-1/2} z_{1-\alpha/2} \hat{\sigma}, \quad (6)$$

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

We remark that in (5), only the last two terms are directly related with the parameter ψ . In other words, for different ψ , the asymptotic variance σ^2 is different only in the last two terms or equivalently ψ and Δ . Let $\mu_i = \int x dF_i(x)$ ($i = 0, 1$) be the population means. When the mean of F_0 is of interest, $\psi = \mu_0$ and $u(x, \theta) = x$, which implies $\Delta = \rho\mu_0$. When the mean difference is of interest, $\psi = \mu_0 - \mu_1$ and $u(x, \theta) = x - x \exp\{\theta^\top q(x)\}$, leading to $\Delta = \rho\mu_0 + (1 - \rho)\mu_\theta$.

3 Empirical likelihood under DRM

The EL inducing the dual EL incorporates only one non-trivial constraint $\sum_{i,j} \exp\{\theta^\top q(x_{ij})\} p_{ij} = 1$, which comes from the DRM itself. The definition of ψ produces an estimating equation

$$\int \{\psi - u(x, \theta)\} dF_0(x) = 0. \quad (7)$$

which can be taken as another constraint on F_0 . When profiling F_0 out, we may maximize the logarithm of the likelihood (2) under both constraints (3) and (7). This leads to the standard profile empirical log-likelihood under DRM. Let $\xi = (\theta^\top, \psi)^\top$ and $m(x, \xi) = (u(x, \theta) - \psi, \exp\{\theta^\top q(x)\} - 1)^\top$. Similarly to Qin and Lawless (1994), the supremum is attained at

$$p_{ij} = \frac{1}{n} \frac{1}{1 + \tilde{\lambda}^\top m(x_{ij}, \xi)},$$

where $\tilde{\lambda} = \tilde{\lambda}(\xi)$ satisfies

$$\sum_{i,j} \frac{m(x_{ij}, \xi)}{1 + \tilde{\lambda}^\top m(x_{ij}, \xi)} = 0. \quad (8)$$

The resultant profile empirical log-likelihood of ξ (up to a constant) is

$$\ell_2(\xi) \equiv \ell_2(\theta, \psi) = - \sum_{i,j} \log\{1 + \tilde{\lambda}^\top m(x_{ij}, \xi)\} + \sum_{j=1}^{n_1} \{\theta^\top q(x_{1j})\}.$$

Denote the resultant maximum EL estimator of ξ by $\tilde{\xi} \equiv (\tilde{\theta}, \tilde{\psi}) = \arg \max_{\xi} \ell_2(\xi)$.

Theorem 2 *The maximum DRM-based EL estimators for all underlying parameters, such as θ , ψ , p_{ij} , F_0 and F_1 , are equal to the respective maximum dual EL estimators. In addition, $\sup_{\xi} \ell_2(\xi) = \sup_{\theta} \ell_1(\theta)$.*

Theorem 2 indicates the standard DRM-based EL and the dual EL share their maximum likelihood estimators of θ and ψ , and their maximum likelihoods. Hence Theorem 1 is also valid for $\tilde{\psi}$, and the Wald-type interval estimator of ψ based on the standard DRM-based EL is exactly (6), which is based on the dual EL. However unlike the latter, the former can be used to construct likelihood ratio confidence intervals for ψ , which is free of variance estimation. Define the EL ratio function of ψ

$$R(\psi) = 2\{\sup_{(\theta, \psi)} \ell_2(\theta, \psi) - \sup_{\theta} \ell_2(\theta, \psi)\}.$$

Theorem 3 *Suppose assumptions (C1)-(C3) are satisfied. As $n \rightarrow \infty$, $R(\psi_0)$ converges in distribution to χ_1^2 distribution.*

A likelihood ratio interval estimator for ψ at level $1 - \alpha$ can be constructed as

$$\{\psi : R(\psi) \leq \chi_{1,1-\alpha}^2\}, \quad (9)$$

where $\chi_{1,1-\alpha}^2$ is the $(1 - \alpha)$ quantile of the χ_1^2 distribution. By Theorems 1 and 3, both the Wald interval in (6) and the likelihood ratio interval (9) have the asymptotically correct coverage probabilities in theory. It seems that the dual EL has the same performance as the DRM-based EL in parameter estimation, since we have shown that they have the same point estimators for all underlying parameters. This is not always true since our simulation study indicates that their interval estimators may perform differently in some situations.

4 A simulation study

In this section, we carry out simulations to compare the finite-sample performances of the dual EL and the DRM-based EL. As they have the same point estimators for all underlying parameters, our comparison focuses on their interval estimators, namely, (6) and (9). We consider two parameters: the mean μ_0 of F_0 , and the mean difference $\mu_0 - \mu_1$ of F_0 and F_1 .

We generated data from three scenarios of population pair: (I) $F_0=N(2, 1)$ and $F_1=N(1, 1.5)$, (II) $F_0 = \text{Gamma}(4, 1)$ and $F_1 = \text{Gamma}(6, 1.1)$, (III) $F_0 = \text{LN}(1, 1)$ and $F_1 = \text{LN}(0, 1)$. Here $\text{Gamma}(a, b)$ is the Gamma distribution with shape parameter a and scale parameter b , and $\text{LN}(a, b)$ is the distribution of a random variable whose natural logarithm follows $N(a, b)$. The three scenarios satisfy the DRM in (1) with $q(x)$ being $q_1(x) = (1, x, x^2)^\top$, $q_2(x) = (1, x, \log(x))^\top$ and $q_3(x) = (1, \log(x))^\top$, respectively. We consider three sample size pairs: $(n_0, n_1) = (20, 20)$, $(40, 40)$ and $(60, 20)$. The nominal level is set to $1 - \alpha = 90\%$, 95% or 99% . The number of simulation repetitions is 2000.

For a generic two-sided interval $[x_l, x_u]$, there are two corresponding one-sided intervals $[x_l, \infty]$ (Lower Limit) and $[\infty, x_u]$ (Upper Limit). **Since one- and two-sided confidence intervals are designed for different purposes, we believe that it is more complete to compare two interval estimation methods from both one- and two-sided interval estimations than from only two-sided interval estimation. Hence we investigate the performances of the one- and two-sided intervals of the likelihood ratio interval (9) (EL) and the Wald interval (6) (DL). The simulated coverage probabilities and average interval lengths of the two-sided EL and DL intervals are reported in Tables (1) and (3), which correspond to μ_0 and $\mu_0 - \mu_1$, respectively. The simulated coverage probabilities of their one-sided intervals for both μ_0 and $\mu_0 - \mu_1$ are reported in Tables (2) and (4), respectively.**

We first examine the simulation results for μ_0 . From Tables 1 and 2, we find that the EL interval has uniformly better coverage accuracy than the DL interval in both one- and two-sided interval estimations. The coverage probabilities of the two-sided EL interval are always closer to the nominal levels than those of the two-sided DL interval in all the 9 combinations of scenario and sample size pair. This is still true in almost all cases for the one-sided EL and DL intervals. In scenarios I and II, the coverage probabilities of both the one- and two-sided EL intervals are almost equal to nominal levels. The one- and two-sided DL intervals have similar performances in most cases in scenarios I and II. However its two-sided coverage probabilities can be as low as 87.15% and 91.95% at nominal levels 90% and 95% respectively in scenario II with $(n_0, n_1) = (20, 20)$, and its one-sided coverage probabilities can be as low as **86.15%, 86.45% and 87.40% at the 90% level**, and as low as 91.35%, 92.45% and 92.5% at the 95% level in scenario II. When the populations become more skewed in scenario III, the relative priority of the EL intervals against the DL interval is more obvious, although both intervals are getting less accurate. The coverage gains of the EL interval against the DL interval can be larger than 2% in two-sided interval estimation, and larger than 4% in one-sided interval estimation.

In scenarios I and II, both interval estimators have acceptable coverage accuracies even when the sample size pair is as small as $(20, 20)$. However their coverage probabilities are much lower than the nominal levels in scenario III. This is probably because the log-normal distributions are far more skewed than the normal and Gamma distributions. When the total sample size increases from 40 to 80, both intervals have remarkably increasing coverage accuracies as expected.

When the parameter of interest is the mean difference $\mu_0 - \mu_1$, the advantage of the EL intervals against the DL intervals is not so obvious as the case for μ_0 . The EL interval still has better coverage probabilities in most situations, while the DL interval can have better two-sided coverage probabilities in scenario III at level 90%. However this priority of the DL interval is probably caused by its seriously inaccurate lower and upper limits. In scenario III, the lower limit of the DL interval has severe under-coverage while its upper limit has severe over-coverage. In all cases, the two-sided EL interval always has longer length than the two-sided DL interval for both μ_0 and $\mu_0 - \mu_1$, which is probably because of its larger coverage probabilities or its data-driven shape (Owen 1990).

To get more insights into the above simulations results, we display in Figures 1 and 2 the QQ-plots of $\text{sign}(\hat{\psi} - \psi_0)\sqrt{R(\psi_0)}$, the sign-root of $R(\psi_0)$, and $\sqrt{n}(\hat{\psi} - \psi_0)/\hat{\sigma}$ for both $\psi = \mu_0$ and $\mu_0 - \mu_1$ in scenarios I, II and III with sample sizes $(n_0, n_1) = (20, 20)$, $(40, 40)$ and $(60, 20)$. The closer their QQ-plots are to the solid line, the better their finite-sample distributions are approximated by their limiting distribution. It can be seen that compared with $\sqrt{n}(\hat{\psi} - \psi_0)/\hat{\sigma}$, the sign-root of $R(\psi_0)$ is always closer to the standard normal or equivalently $R(\psi_0)$ is always closer to the χ_1^2 distribution. This explains why the EL intervals have more accurate coverage probabilities than the DL intervals. In addition, both their QQ-plots are very close to the solid line in scenarios I and II, but become far away from it in scenario III. This coincides with the observation that both intervals have better performances in scenarios I and II but have poor performances in scenario III.

Overall, the finite-sample distribution of the likelihood ratio $R(\psi_0)$ is better approximated by its limiting χ_1^2 distribution than that of $\sqrt{n}(\hat{\psi} - \psi_0)/\hat{\sigma}$ is approximated by its normal distribution. Compared

with the DL intervals, the EL intervals have comparable or only slightly better coverage accuracy when the populations under study is non or moderate skewed, and become relatively much more accurate for severely skewed populations.

Table 1: Coverage probabilities (coverage, %) and average interval lengths (length) of the EL and DL intervals for μ_0

Scenario	(n_0, n_1)	Level	EL			DL		
			90%	95%	99%	90%	95%	99%
I	(20,20)	Coverage	88.30	94.50	98.50	87.35	92.80	97.85
		Length	0.73	0.88	1.19	0.70	0.84	1.10
	(40,40)	Coverage	89.80	94.45	99.10	89.20	94.05	98.65
		Length	0.52	0.62	0.83	0.51	0.61	0.80
	(60,20)	Coverage	90.75	95.10	98.95	90.35	94.75	98.80
		Length	0.42	0.51	0.68	0.42	0.50	0.66
II	(20,20)	Coverage	88.65	94.30	98.70	87.15	91.95	97.10
		Length	1.51	1.83	2.53	1.43	1.71	2.24
	(40,40)	Coverage	88.50	94.70	99.25	88.35	93.50	98.35
		Length	1.06	1.28	1.73	1.03	1.23	1.61
	(60,20)	Coverage	89.10	94.95	99.05	88.75	94.15	98.30
		Length	0.86	1.03	1.38	0.84	1.00	1.32
III	(20,20)	Coverage	80.70	86.70	94.00	79.30	84.55	91.25
		Length	3.60	4.34	5.81	3.50	4.17	5.48
	(40,40)	Coverage	84.35	90.10	96.50	83.85	88.50	93.95
		Length	2.79	3.37	4.56	2.69	3.20	4.21
	(60,20)	Coverage	84.50	91.20	97.35	84.00	89.20	94.70
		Length	2.36	2.85	3.86	2.28	2.70	3.55

5 A real data example

In this section, we further compare the DRM-based EL and dual EL interval estimation methods by analyzing the lumber data, which is obtained from tests conducted by the FPInnovations laboratory at the University of British Columbia. The data-set, [which is available upon request](#), consists of the MOE measurements for lumber produced in 2007 and in 2010 with sample sizes 98 and 282, respectively. See Chen and Liu (2013) for a more detailed description.

Let μ_0 and μ_1 be the means of MOE in 2010 and 2007. We wish to estimate the mean μ_0 and the differences $\mu_0 - \mu_1$. [For the choice of the basis function \$q\(x\)\$, we consider \$q_1\(x\)\$, \$q_2\(x\)\$, and \$q_3\(x\)\$, which are the basis functions used in our simulation study.](#) Table 5 tabulates the EL and DL interval estimates and the corresponding lengths for the two parameters at nominal level 95%. We observe that the two intervals are quite close to each other for both parameters in all the four cases, and the EL intervals are slightly longer than the DL intervals. This observation is consistent with those concluded from our simulation study. [The basis functions \$q_1\(x\)\$ and \$q_2\(x\)\$ produce almost the same results. When we change the base function from \$q_1\(x\)\$ or \$q_2\(x\)\$ to \$q_3\(x\)\$, the intervals have a relatively big change. Since \$q_3\(x\)\$ is a subvector of \$q_2\(x\)\$, DRM with \$q_2\(x\)\$ would be less risky to be misspecified than DRM with \$q_3\(x\)\$. Therefore in this situation, we believe that the results corresponding to \$q_1\(x\)\$ and \$q_2\(x\)\$ would be more reliable.](#)

Table 2: Coverage probabilities (%) of the one-sided EL and DL intervals for μ_0

Scenario	(n_0, n_1)	$1 - \alpha$	EL			DL		
			90%	95%	99%	90%	95%	99%
I	(20,20)	Lower Limit	88.80	93.75	98.40	88.55	93.55	98.10
		Upper Limit	89.15	94.55	99.05	88.45	93.80	98.20
	(40,40)	Lower Limit	90.00	94.80	98.80	89.90	94.70	98.70
		Upper Limit	90.05	95.00	99.05	89.90	94.50	98.70
	(60,20)	Lower Limit	89.20	95.00	99.05	89.30	94.90	99.00
		Upper Limit	90.75	95.75	98.85	90.55	95.45	98.65
II	(20,20)	Lower Limit	89.20	94.55	99.05	86.15	91.35	96.30
		Upper Limit	90.30	94.10	98.65	91.45	95.80	99.30
	(40,40)	Lower Limit	88.45	94.35	98.90	86.45	92.45	97.40
		Upper Limit	89.20	94.15	99.15	90.45	95.90	99.55
	(60,20)	Lower Limit	89.15	94.15	98.70	87.40	92.50	97.70
		Upper Limit	89.95	94.95	99.15	91.05	96.25	99.45
III	(20,20)	Lower Limit	78.55	85.05	93.30	75.65	80.65	88.95
		Upper Limit	91.10	95.65	98.95	95.25	98.65	99.85
	(40,40)	Lower Limit	83.15	88.75	95.65	79.00	85.45	92.30
		Upper Limit	90.55	95.60	99.15	94.90	98.40	99.95
	(60,20)	Lower Limit	82.45	89.05	95.95	79.85	85.20	93.30
		Upper Limit	91.05	95.45	99.40	95.20	98.80	1.00

6 Conclusion

This paper compares the dual EL and the standard EL in the context of two-sample DRM through both theoretical and numerical analyses. We found that the two methods produce the same point estimators for any underlying parameter, and both their corresponding intervals have asymptotically correct coverage probabilities. Even so, their intervals have different finite-sample performances in some situations.

Our primary goal is to highlight that there are differences between the two estimation methods although they are both based on EL and DRM, and produce the same point estimators. Hence we focus on parameters having the form of $\psi = \int u(x, \theta) dF_0(x)$ for simplicity. It would be interesting to extend our results to parameters that are defined through general estimating equations (Qin and Lawless 1994). Our theoretical results are still true in the just-identified case, namely the number of parameters is equal to the number of equations. For the over-identified case, where the number of parameters is greater than the number of equations, the maximum dual EL estimator of the parameters may not be well defined. In addition, the EL and DL methods would also have different powers in testing hypotheses. We would leave it as a further research topic.

Appendix: Proofs

We begin with a number of technical lemmas. Throughout, we assume the same conditions in Theorem 1. For the data $\{x_{ij} : j = 1, 2, \dots, n_i; i = 0, 1\}$, define $Y_n = \sum_{i,j} u(x_{ij}) / \{nh(x_{ij})\}$ and

$$Z_n = \frac{\partial \ell_1(\theta_0)}{\partial \theta} = \sum_{j=1}^{n_1} q(x_{1j}) - \sum_{i,j} h_1(x_{ij}) q(x_{ij}).$$

Also define $S = \{\rho(1 - \rho)\}^{-1} \text{diag}\{1, 0, \dots, 0\}$ and $c_{rs} = \int u(x) h_r(x) h_s(x) h(x) dF_0(x)$ for $0 \leq r, s \leq 1$.

Lemma 1 For Y_n and Z_n defined above, we have

(a) $\mathbb{E}(Z_n) = 0$ and $\mathbb{V}\text{ar}(Z_n) = n(W - WSW)$;

Table 3: Coverage probabilities (coverage, %) and average interval lengths (length) of the EL and DL intervals for $\mu_0 - \mu_1$

Scenarios	(n_0, n_1)	Level	EL			DL		
			90%	95%	99%	90%	95%	99%
I	(20,20)	Coverage	88.55	94.65	98.45	87.50	94.00	98.35
		Length	1.15	1.37	1.83	1.13	1.34	1.77
	(40,40)	Coverage	89.20	94.25	99.00	88.85	94.00	98.80
		Length	0.81	0.97	1.29	0.81	0.96	1.27
	(60,20)	Coverage	88.05	94.00	98.75	87.65	93.70	98.20
		Length	0.98	1.18	1.57	0.96	1.15	1.51
II	(20,20)	Coverage	88.60	93.70	98.70	88.50	93.50	98.50
		Length	2.44	2.94	3.98	2.38	2.84	3.73
	(40,40)	Coverage	89.80	94.70	99.05	89.90	94.45	98.85
		Length	1.74	2.09	2.80	1.71	2.04	2.68
	(60,20)	Coverage	87.50	93.20	98.20	87.05	92.45	97.65
		Length	2.10	2.52	3.36	2.06	2.45	3.23
III	(20,20)	Coverage	81.40	88.50	95.30	81.95	87.65	93.75
		Length	3.81	4.61	6.29	3.65	4.35	5.72
	(40,40)	Coverage	85.05	90.50	96.75	85.30	90.65	95.55
		Length	2.89	3.49	4.76	2.76	3.29	4.33
	(60,20)	Coverage	86.05	91.85	98.05	87.00	91.90	97.85
		Length	2.68	3.26	4.48	2.55	3.04	3.99

$$(b) \mathbb{E}(Y_n) = \psi_0 \text{ and } n\text{Var}(Y_n) = \int \frac{u^2(x)}{h(x)} dF_0(x) - \frac{1}{(1-\rho)^3} c_{00}^2 - \frac{1}{\rho(1-\rho)^2} c_{01}^2;$$

$$(c) \text{Cov}(Y_n, Z_n) = \left\{ \frac{c_{00}}{(1-\rho)^2} - \frac{c_{01}}{\rho(1-\rho)} \right\} e_1^\top W.$$

Proof. Result (a) can be proved by direct calculation. See also the proof of Theorem 2.1 of Chen and Liu (2013).

For result (b), it is trivial to prove $\mathbb{E}(Y_n) = \psi_0$. We now calculate $\text{Var}(Y_n)$. It can be seen that

$$\begin{aligned} n\text{Var}(Y_n) &= (1-\rho)\text{Var}\left\{\frac{u(x_{01})}{h(x_{01})}\right\} + \rho\text{Var}\left\{\frac{u(x_{11})}{h(x_{11})}\right\} \\ &= (1-\rho)\mathbb{E}\left(\frac{u(x_{01})}{h(x_{01})}\right)^2 - (1-\rho)\left[\mathbb{E}\left\{\frac{u(x_{01})}{h(x_{01})}\right\}\right]^2 + \rho\mathbb{E}\left(\frac{u(x_{11})}{h(x_{11})}\right)^2 - \rho\left[\mathbb{E}\left\{\frac{u(x_{11})}{h(x_{11})}\right\}\right]^2 \\ &= \int \frac{u^2(x)}{h^2(x)} h_0(x)h(x) dF_0(x) - \frac{c_{00}^2}{(1-\rho)^3} + \int \frac{u^2(x)}{h^2(x)} h_1(x)h(x) dF_0(x) - \frac{c_{01}^2}{\rho(1-\rho)^2} \\ &= \int \frac{u^2(x)}{h(x)} dF_0(x) - \frac{c_{00}^2}{(1-\rho)^3} - \frac{c_{01}^2}{\rho(1-\rho)^2}. \end{aligned}$$

Table 4: Coverage probabilities (%) of the one-sided EL and DL intervals for $\mu_0 - \mu_1$

Scenario	(n_0, n_1)	1- α	EL			DL		
			90%	95%	99%	90%	95%	99%
I	(20,20)	Lower Limit	88.05	94.10	98.50	87.95	93.45	98.45
		Upper Limit	89.20	94.45	99.00	88.60	94.05	98.80
	(40,40)	Lower Limit	90.50	95.30	99.00	90.60	95.10	99.00
		Upper Limit	89.35	93.90	98.30	89.00	93.75	98.15
	(60,20)	Lower Limit	89.85	94.35	98.30	89.85	94.30	98.15
		Upper Limit	88.80	93.70	99.00	88.45	93.35	98.30
II	(20,20)	Lower Limit	89.35	94.70	99.00	89.30	94.80	98.75
		Upper Limit	88.10	93.90	98.40	87.80	93.70	98.35
	(40,40)	Lower Limit	90.40	95.15	98.95	90.50	95.35	98.80
		Upper Limit	89.05	94.65	99.00	89.05	94.55	98.75
	(60,20)	Lower Limit	89.50	95.35	98.90	90.70	95.80	99.05
		Upper Limit	87.25	92.15	98.05	86.35	91.25	97.15
III	(20,20)	Lower Limit	79.30	86.15	94.60	76.35	83.55	91.80
		Upper Limit	90.65	95.25	99.05	94.80	98.40	99.90
	(40,40)	Lower Limit	84.30	90.00	96.05	81.40	87.50	94.10
		Upper Limit	91.05	95.05	99.10	94.00	97.80	99.85
	(60,20)	Lower Limit	83.70	90.35	97.35	82.80	88.70	95.95
		Upper Limit	90.65	95.70	99.15	94.05	98.30	99.85

We now prove result (c). Similarly, we have

$$\begin{aligned}
& \text{Cov}(Y_n, Z_n) \\
&= (1 - \rho) \text{Cov} \left(\frac{u(x_{01})}{h(x_{01})}, -h_1(x_{01})q(x_{01}) \right) + \rho \text{Cov} \left(\frac{u(x_{11})}{h(x_{11})}, h_0(x_{11})q(x_{11}) \right) \\
&= (1 - \rho) \left\{ -\mathbb{E} \left(\frac{u(x_{01})}{h(x_{01})} h_1(x_{01}) q^\top(x_{01}) \right) + \mathbb{E} \left(\frac{u(x_{01})}{h(x_{01})} \right) \mathbb{E}(h_1(x_{01}) q^\top(x_{01})) \right\} \\
&\quad + \rho \mathbb{E} \left(\frac{u(x_{11})}{h(x_{11})} h_0(x_{11}) q^\top(x_{11}) \right) - \rho \mathbb{E} \left(\frac{u(x_{11})}{h(x_{11})} \right) \mathbb{E}(h_0(x_{11}) q^\top(x_{11})) \\
&= - \int \frac{u(x) h_1(x) h_0(x) q^\top(x)}{h(x)} h(x) dF_0(x) + \frac{c_{00}}{(1 - \rho)^2} \int h_1(x) q^\top(x) h_0(x) h(x) dF_0(x) \\
&\quad + \int \frac{u(x)}{h(x)} h_0(x) h_1(x) q^\top(x) h(x) dF_0(x) - \frac{c_{01}}{\rho(1 - \rho)} \int h_0(x) h_1(x) q^\top(x) h(x) dF_0(x) \\
&= \left\{ \frac{c_{00}}{(1 - \rho)^2} - \frac{c_{01}}{\rho(1 - \rho)} \right\} e_1^\top W.
\end{aligned}$$

□

Lemma 2 Assume the same conditions in Theorem 1. As n is large, $\hat{\theta} - \theta_0 = W^{-1}Z_n/n + o_p(n^{-1/2})$.

The result in Lemma 2 comes from the proof of Theorem 2.1 of Chen and Liu (2013). The proof is omitted here.

Proof of Theorem 1

Proof. Rewrite equation (4) as $\hat{\psi} = \sum_{i,j} u(x_{ij}, \hat{\theta}) / \{nh(x_{ij}, \hat{\theta})\}$. Using $\hat{\theta} = \theta_0 + O_p(n^{-1/2})$ and applying the delta method to (4), we have

$$\hat{\psi} = \sum_{i,j} \frac{u(x_{ij})}{nh(x_{ij})} - \sum_{i,j} \frac{h_\theta(x_{ij})u(x_{ij}) - u_\theta(x_{ij})h(x_{ij})}{nh^2(x_{ij})} (\hat{\theta} - \theta) + o(n^{-1/2}).$$

Table 5: DL and EL interval estimates at confidence level 95% with $q_1(x) = (1, x, x^2)^\top$, $q_2(x) = (1, x, \log(x))^\top$ and $q_3(x) = (1, \log(x))^\top$.

Parameter	$q(x)$	Method	Confidence Interval	Interval Length
μ_0	$q_1(x)$	DL	[1.4959, 1.5324]	0.0365
		EL	[1.4960, 1.5327]	0.0367
	$q_2(x)$	DL	[1.4958, 1.5324]	0.0365
		EL	[1.4960, 1.5327]	0.0367
	$q_3(x)$	DL	[1.4959, 1.5315]	0.0356
		EL	[1.4973, 1.5317]	0.0344
$\mu_0 - \mu_1$	$q_1(x)$	DL	[-0.0244, 0.0420]	0.0664
		EL	[-0.0248, 0.0420]	0.0668
	$q_2(x)$	DL	[-0.0244, 0.0420]	0.0664
		EL	[-0.0248, 0.0420]	0.0668
	$q_3(x)$	DL	[-0.0276, 0.0422]	0.0698
		EL	[-0.0279, 0.0421]	0.0699

where $h_\theta(x, \theta) = \partial h(x, \theta) / \partial \theta = h(x, \theta) h_1(x, \theta) q(x)$ and $h_\theta(x) = h_\theta(x, \theta_0)$. By the law of large numbers,

$$\sum_{i,j} \frac{h_\theta(x_{ij}) u(x_{ij}) - u_\theta(x_{ij}) h(x_{ij})}{nh^2(x_{ij})} = B + o_p(1),$$

where B is defined above Theorem 1. By Lemma 2, we have

$$\hat{\psi} = Y_n - B^\top (\hat{\theta} - \theta) + o_p(n^{-1/2}) = Y_n - B^\top W^{-1} Z_n / n + o_p(n^{-1/2}).$$

By Slutsky's theorem, $\sqrt{n}(\hat{\psi} - \psi_0)$ has the same limiting distribution as

$$\sqrt{n}\{Y_n - \psi_0\} - B^\top W^{-1} Z_n / n$$

which clearly has mean zero. It remains to calculate the variance of the latter.

Applying the results in Lemma 1, we have

$$\begin{aligned} \text{Cov}(Y_n, B^\top W^{-1} Z_n) &= \left\{ \frac{c_{00}}{(1-\rho)^2} - \frac{c_{01}}{\rho(1-\rho)} \right\} e_1^\top W (W^{-1} B) \\ &= \left\{ \frac{c_{00}}{(1-\rho)^2} - \frac{c_{01}}{\rho(1-\rho)} \right\} e_1^\top B. \end{aligned}$$

Since $\text{Var}(Z_n)/n = W - WSW$ and $S = \{\rho(1-\rho)\}^{-1} \text{diag}\{1, 0, \dots, 0\} = \{\rho(1-\rho)\}^{-1} e_1 e_1^\top$, it follows that

$$\begin{aligned} n\text{Var}(B^\top W^{-1} Z_n / n) &= B^\top W^{-1} (W - WSW) W^{-1} B \\ &= B^\top W^{-1} B - \{\rho(1-\rho)\}^{-1} B^\top e_1 e_1^\top B. \end{aligned}$$

Putting all these terms together, we have

$$\begin{aligned} \sigma^2 &= n\text{Var}(Y_n - \psi_0 - B^\top W^{-1} Z_n / n) \\ &= n\text{Var}(Y_n) - \text{Cov}(Y_n, B^\top W^{-1} Z_n) - \text{Cov}(B^\top W^{-1} Z_n, Y_n) \\ &\quad + B^\top W^{-1} \text{Var}(Z_n) W^{-1} B \\ &= \int \frac{u^2(x)}{h(x)} dF_0(x) - \frac{1}{(1-\rho)^3} c_{00}^2 - \frac{1}{\rho(1-\rho)^2} c_{01}^2 \\ &\quad - \left\{ \frac{c_{00}}{(1-\rho)^2} - \frac{c_{01}}{\rho(1-\rho)} \right\} e_1^\top B - B^\top e_1 \left\{ \frac{c_{00}}{(1-\rho)^2} - \frac{c_{01}}{\rho(1-\rho)} \right\}^\top \\ &\quad + B^\top W^{-1} B - \{\rho(1-\rho)\}^{-1} B^\top e_1 e_1^\top B \\ &= \int \frac{u^2(x)}{h(x)} dF_0(x) + B^\top W^{-1} B - \frac{\eta}{\rho(1-\rho)}, \end{aligned}$$

where $\eta = (B^\top e_1 + \frac{\rho}{1-\rho}c_{00} - c_{01})^2 + \frac{\rho}{1-\rho}(c_{00} + c_{01})^2$. Since $B^\top e_1 = \int \{h_1(x)u(x) - e_1^\top u_\theta(x)\}dF_0(x)$, we have

$$B^\top e_1 + \frac{\rho}{1-\rho}c_{00} - c_{01} = \rho\psi - \int e_1^\top u_\theta(x, \theta)dF_0(x) = \Delta.$$

Further using $c_{00} + c_{01} = (1 - \rho)\psi$ gives

$$\eta = \rho(1 - \rho)\psi^2 - \frac{\Delta^2}{\rho(1 - \rho)}.$$

This proves the expression of σ^2 and Theorem 1. \square

Proof of Theorem 2

Proof. Partition λ as $\tilde{\lambda} = (\lambda_1, \lambda_2)^\top$ and partition $\tilde{\lambda}$ in the same way. Setting $\partial \ell_2(\xi)/\partial \xi = 0$ gives

$$\frac{\partial \ell_2(\theta, \psi)}{\partial \theta} = - \sum_{i,j} \frac{\lambda_1 u_\theta(x_{ij}, \theta) + \lambda_2 \exp\{\theta^\top q(x_{ij})\}q(x_{ij})}{1 + \lambda_1 \{u(x_{ij}, \theta) - \psi\} + \lambda_2 [\exp\{\theta^\top q(x_{ij})\} - 1]} = 0, \quad (10)$$

$$\frac{\partial \ell_2(\theta, \psi)}{\partial \psi} = \sum_{i,j} \frac{\lambda_1}{1 + \lambda_1 \{u(x_{ij}, \theta) - \psi\} + \lambda_2 [\exp\{\theta^\top q(x_{ij})\} - 1]} = 0. \quad (11)$$

When $\psi = \tilde{\psi}$ and $\theta = \tilde{\theta}$, it follows from Equations (8) and (11) that $\tilde{\lambda}_1(\tilde{\theta}, \tilde{\psi}) = 0$. This implies that $(\tilde{\theta}, \tilde{\psi}, \tilde{\lambda}_2)$ is simply the solution to

$$\begin{cases} - \sum_{i,j} \frac{\lambda_2 \exp\{\theta^\top q(x_{ij})\}q(x_{ij})}{1 + \lambda_2 [\exp\{\theta^\top q(x_{ij})\} - 1]} = 0, \\ \sum_{i,j} \frac{u(x_{ij}, \theta) - \psi}{1 + \lambda_2 [\exp\{\theta^\top q(x_{ij})\} - 1]} = 0, \\ \sum_{i,j} \frac{\exp\{\theta^\top q(x_{ij})\} - 1}{1 + \lambda_2 [\exp\{\theta^\top q(x_{ij})\} - 1]} = 0, \end{cases}$$

which are Equations (8) and (10) with λ_1 replaced by 0. It is trivial to verify that the equations are exactly those which define $(\hat{\theta}, \hat{\psi}, \hat{\lambda})$. Consequently, the maximum DRM-based EL estimators for all underlying parameters, such as $\theta, \psi, p_{ij}, F_0$ and F_1 , are equal to the respective maximum dual EL estimators. This also indicates that $\sup_\xi \ell_2(\xi) = \sup_\theta \ell_1(\theta)$. \square

Proof of Theorem 3

Proof. It can be verified that $\ell_2(\xi) = \min_\lambda H(\lambda, \theta, \psi)$, where

$$H(\lambda, \theta, \psi) = - \sum_{i,j} \log\{1 + \lambda^\top m(x_{ij}, \xi)\} + \sum_{j=1}^{n_1} \{\theta^\top q(x_{1j})\}.$$

Since $\tilde{\xi}$ is root- n consistent, we need to study the behavior of $\ell_2(\xi)$ for $\xi = \xi_0 + O_p(n^{-1/2})$ with $\xi_0 = (\theta_0^\top, \psi_0)^\top$.

For $\xi = \xi_0 + O_p(n^{-1/2})$, it can be verified that the Lagrange multiplier $\tilde{\lambda} = \lambda_0 + O_p(n^{-1/2})$ where $\lambda_0 = (0, \rho)$. Denote $\zeta^\top = (\zeta_1, \zeta_2) = n^{1/2}((\lambda - \lambda_0)^\top, (\xi - \xi_0)^\top) = n^{1/2}((\lambda - \lambda_0)^\top, (\theta - \theta_0)^\top, (\psi - \psi_0)^\top)$, where $\zeta_1 = n^{1/2}((\lambda - \lambda_0)^\top, (\theta - \theta_0)^\top)$ and $\zeta_2 = n^{1/2}(\psi - \psi_0)^\top$. Define

$$Q(\zeta) = H(\zeta_{10} + n^{-1/2}\zeta_1, \psi_0 + n^{-1/2}\zeta_2)$$

with $\zeta_{10} = (\lambda_0, \theta_0)$.

Define

$$A = n^{-1/2} \begin{pmatrix} \frac{\partial H}{\partial \lambda} \\ \frac{\partial H}{\partial \theta} \\ \frac{\partial H}{\partial \psi} \end{pmatrix}_{(\lambda_0, \xi_0)}, \quad C_n = n^{-1} \begin{pmatrix} \frac{\partial^2 H}{\partial \lambda \partial \lambda^\top} & \frac{\partial^2 H}{\partial \lambda \partial \theta^\top} & \frac{\partial^2 H}{\partial \lambda \partial \psi^\top} \\ \frac{\partial^2 H}{\partial \theta \partial \lambda^\top} & \frac{\partial^2 H}{\partial \theta \partial \theta^\top} & \frac{\partial^2 H}{\partial \theta \partial \psi^\top} \\ \frac{\partial^2 H}{\partial \psi \partial \lambda^\top} & \frac{\partial^2 H}{\partial \psi \partial \theta^\top} & \frac{\partial^2 H}{\partial \psi \partial \psi^\top} \end{pmatrix}_{(\lambda_0, \xi_0)}.$$

It can be verified that $A_n = O_p(n^{-1/2})$ and $C_n = C + o_p(1)$, where $C = \mathbb{E}(C_n)$. Applying the second-order Taylor expansion to $Q(\zeta)$, we have

$$Q(\zeta) = H(\zeta_{10}, \psi_0) + A^\top \zeta + \frac{1}{2} \zeta^\top C \zeta + o_p(1).$$

Setting $\partial Q / \partial \zeta = 0$ gives $\hat{\zeta} = -C^{-1}A + o_p(1)$. Hence

$$\sup_{\xi} \ell_2(\xi) = Q(\hat{\zeta}) = H(\zeta_{10}, \psi_0) - \frac{1}{2} A^\top C^{-1} A + o_p(1).$$

Partition $A = (A_1^\top, A_2^\top)^\top$ and C in the same way as ζ . Setting $\partial Q / \partial \zeta_1$ to zero gives

$$\hat{\zeta}_1 = -C_{11}^{-1} A_1 + o_p(1).$$

Consequently,

$$\sup_{\theta} \ell_2(\theta, \psi_0) = Q(\hat{\zeta}_1, 0) = H(\zeta_{10}, \psi_0) - \frac{1}{2} A_1^\top C_{11}^{-1} A_1 + o_p(1).$$

Thus we have

$$R_2(\psi_0) = 2 \sup_{(\theta, \psi)} \ell_2(\theta, \psi) - 2 \sup_{\theta} \ell_2(\theta, \psi_0) = A_1^\top C_{11}^{-1} A_1 - A^\top C^{-1} A + o_p(1).$$

Since

$$C^{-1} = \begin{pmatrix} C_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} C_{11}^{-1} C_{12} \\ -I \end{pmatrix} C_{22 \cdot 1}^{-1} \begin{pmatrix} C_{21} C_{11}^{-1} & -I \end{pmatrix}$$

with $C_{22 \cdot 1} = C_{22} - C_{21} C_{11}^{-1} C_{12}$, it can be verified that

$$A^\top C^{-1} A = A_1 C_{11}^{-1} A_1^\top + (A_1, A_2) \begin{pmatrix} C_{11}^{-1} C_{12} \\ -I \end{pmatrix} C_{22 \cdot 1}^{-1} \begin{pmatrix} C_{21} C_{11}^{-1} & -I \end{pmatrix} \begin{pmatrix} A_1^\top \\ A_2^\top \end{pmatrix}.$$

Consequently,

$$R_2(\psi_0) = (A_1 C_{11}^{-1} C_{12} - A_2) C_{22 \cdot 1}^{-1} (A_1 C_{11}^{-1} C_{12} - A_2)^\top + o_p(1).$$

After length algebra, we can show that

$$\sqrt{n}(A_1 C_{11}^{-1} C_{12} - A_2) \xrightarrow{d} N(0, C_{22 \cdot 1})$$

which directly implies $R_2(\psi_0) \xrightarrow{d} \chi_1^2$. This completes the proof. \square

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grant numbers 11771144, 11371142, 11501354, 11571112, and [11501208](#); and Chinese Ministry of Education the 111 Project under Grant number B14019; and Doctoral Fund of Ministry of Education of China under Grant number 20130076110004.

Acknowledgement

We thank the editor, the Editorial Board member, and an anonymous referee for their constructive suggestions that significantly improved the paper.

References

- Anderson, J. A. (1972), ‘Separate Sample Logistic Discrimination’, *Biometrika*, 59, 19–35.
- Anderson, J. A. (1974), ‘Diagnosis by Logistic Discriminant Function: Further Practical Problems and Results’, *Applied Statistics*, 23, 397–404.
- Anderson, J. A. (1979), ‘Multivariate Logistic Compounds’, *Biometrika*, 66, 17–26.
- Chen, J. and Liu, Y. (2013), ‘Quantile and Quantile-function Estimations Under Density Ratio Model’, *The Annals of Statistics*, 41, 1669–1692.
- 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.
- DiCiccio, T., Hall, P., and Romano, J. (1991), ‘Empirical Likelihood is Bartlett-correctable’, *The Annals of Statistics*, 19, 1053–1061.
- Fokianos, K., B. Kedem, J. Qin, and D. A. Short (2001), ‘A Semiparametric Approach to the One-way Layout’, *Technometrics*, 43, 56–65.
- Fokianos, K. (2004), ‘Merging Information for Semiparametric Density Estimation’, *Journal of the Royal Statistical Society: Series B*, 66, 941–958.
- Gill, R., Vardi, Y., and Wellner, J. A. (1988), ‘Large Sample Theory of Empirical Distributions in Biased Sampling Models’, *The Annals of Statistics*, 16, 1069–1112.
- Hall, P. and La Scala, B. (1990), ‘Methodology and Algorithms of Empirical Likelihood’, *International Statistical Review*, 58, 109–127.
- Kay, R. and Little, S. (1987), ‘Transformations of the Explanatory Variables in the Logistic Regression Model for Binary Data’, *Biometrika*, 74, 495–501.
- Keziou, A. and Leoni-Aubin, S. (2008), ‘On Empirical Likelihood for Semiparametric Two-sample Density Ratio Models’, *Journal of Statistical Planning & Inference*, 138, 915–928.
- Owen, A. B. (1988), ‘Empirical Likelihood Ratio Confidence Intervals for a Single Functional’, *Biometrika*, 75, 237–249.
- Owen, A. B. (1990), ‘Empirical Likelihood Ratio Confidence Regions’, *The Annals of Statistics*, 18, 90–120.
- Owen, A. B. (2001), ‘*Empirical Likelihood*’, Chapman and Hall/CRC, New York.
- Qin, J. (1993), ‘Empirical Likelihood in Biased Sample Problems’, *The Annals of Statistics*, 21, 1182–1196.
- Qin, J. (1998), ‘Inferences for Case-control and Semiparametric Two-sample Density Models’, *Biometrika*, 85, 619–630.
- Qin, J. (1999), ‘Empirical Likelihood Ratio Based Confidence Intervals for Mixture Proportions’, *The Annals of Statistics*, 27, 1368–1384.
- Qin, J. and Lawless, J. (1994), ‘Empirical Likelihood and General Estimating Equations’, *The Annals of Statistics*, 22, 300–325.
- Qin, J. and Zhang, B. (1997), ‘A Goodness-of-fit Test for Logistic Regression Models Based on Case-control Data’, *Biometrika*, 84, 609–618.
- Qin, J., 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.

- Vardi, Y. (1982), ‘Nonparametric Estimation in the Presence of Length Bias’, *The Annals of Statistics*, 10, 616–620.
- Vardi, Y. (1985), ‘Empirical Distributions in Selection Bias Models’, *The Annals of Statistics*, 13, 178–203.
- Wang, C. (2017), ‘Empirical Likelihood and Bootstrap Inference With Constraints’, *PhD Thesis*, University of Waterloo.
- Zhang, B. (2000), ‘Quantile Estimation Under a Two-sample Semiparametric Model’, *Bernoulli*, 6, 491–511.
- Zhang, B. (2002), ‘Assessing Goodness-of-fit of Generalized Logit Models Based on Case-control Data’, *Journal of Multivariate Analysis*, 82, 17–38.

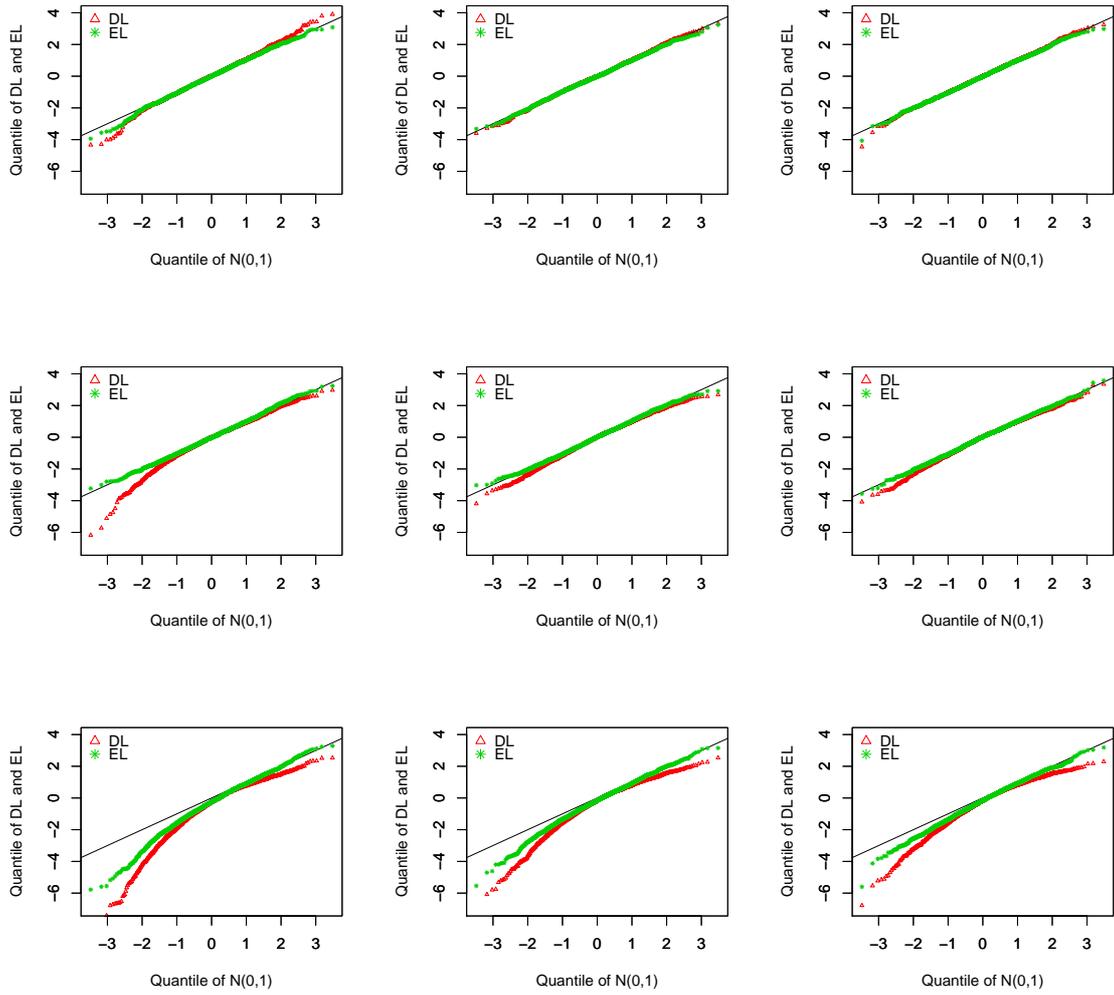


Figure 1: QQ-plots of $\text{sign}(\tilde{\psi} - \psi_0)\sqrt{R(\psi_0)}$ (EL) and $\sqrt{n}(\hat{\psi} - \psi_0)/\hat{\sigma}$ (DL) for $\psi = \mu_0$ in scenarios I, II and III (from top to bottom) with sample sizes $(n_0, n_1) = (20, 20), (40, 40)$ and $(60, 20)$ (from left to right).

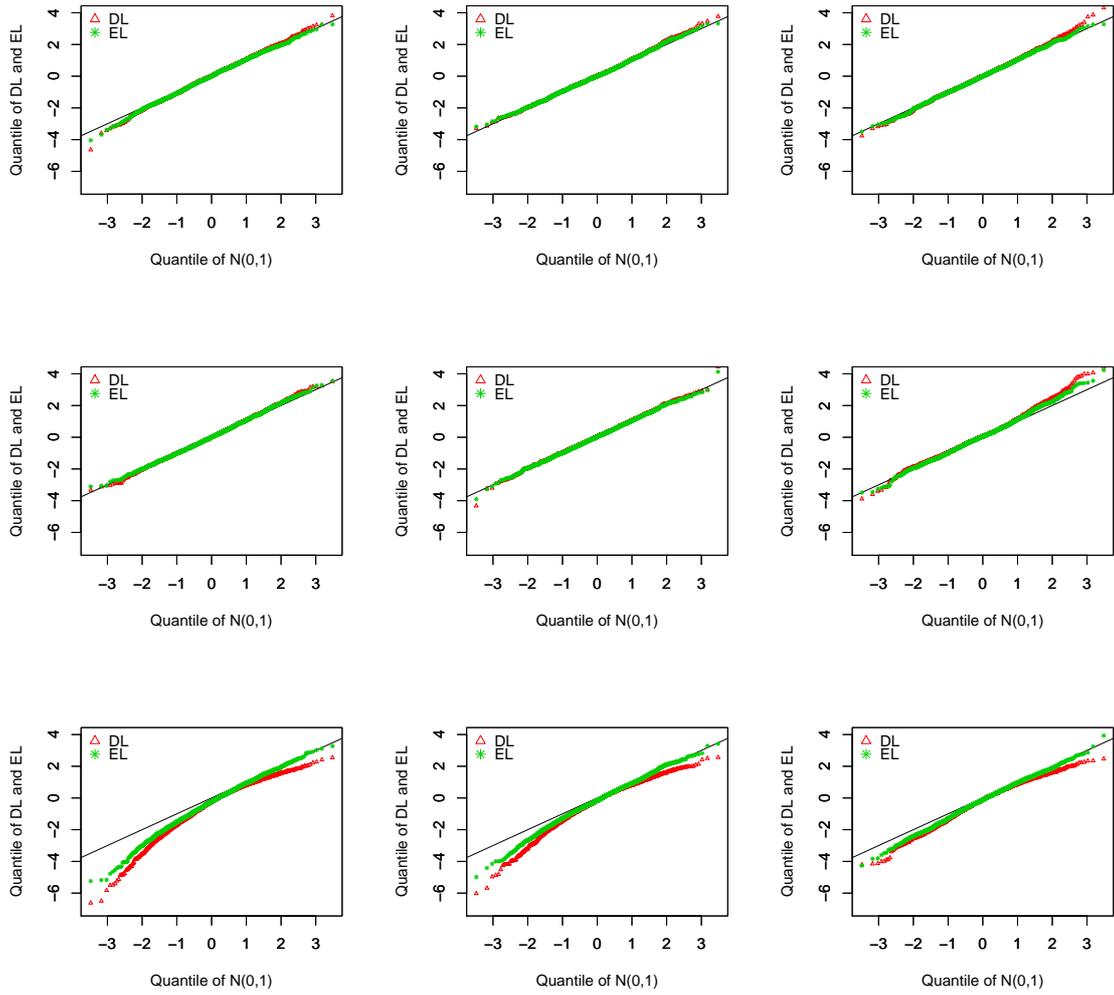


Figure 2: QQ-plots of $\text{sign}(\tilde{\psi} - \psi_0)\sqrt{R(\psi_0)}$ (EL) and $\sqrt{n}(\hat{\psi} - \psi_0)/\hat{\sigma}$ (DL) for $\psi = \mu_0 - \mu_1$ in scenarios I, II and III (from top to bottom) with sample sizes $(n_0, n_1) = (20, 20)$, $(40, 40)$ and $(60, 20)$ (from left to right).