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

Huapeng Li, Yang Liu, Yukun Liu & Riquan Zhang


To link to this article: https://doi.org/10.1080/10485252.2018.1457790
Comparison of empirical likelihood and its dual likelihood under density ratio model

Huapeng Li\textsuperscript{a,b}, Yang Li\textsuperscript{a}, Yukun Liu\textsuperscript{a} and Riquan Zhang\textsuperscript{a}

\textsuperscript{a}School of Statistics, East China Normal University, Shanghai, People’s Republic of China; \textsuperscript{b}School of Mathematics and Computer Sciences, Shanxi Datong University, Datong, People’s Republic of China

\section*{ABSTRACT}

Density ratio models (DRMs) 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 analysed for illustration purpose.

\section*{ARTICLE HISTORY}
Received 5 September 2017
Accepted 20 March 2018

\section*{KEYWORDS}
Coverage accuracy; density ratio model; dual empirical likelihood; empirical likelihood

\section*{AMS 2000 SUBJECT CLASSIFICATIONS}
Primary: 62G15; Secondary: 62G20

\section*{1. Introduction}

Density ratio models (DRMs) 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 \( \{x_{ij} : j = 1, \ldots, n_i\} (i = 0, 1) \) from two populations with cumulative distribution functions \( F_0(x) \) and \( F_1(x) \), respectively. The DRM assumes

\[
dF_1(x)/dF_0(x) = \exp\{\theta^\top q(x)\},
\]

where \( q(x) \) is a pre-specified \( d \)-dimensional basis function and \( \theta \) a \( d \)-variate unknown parameter. We require the first element of \( q(x) \) to be one so that the first element of \( \theta \) is a normalisation 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 nonparametric 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 characterise 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 Leoni-Aubin 2008). Fortunately, Keziou and Leoni-Aubin (2008) formally established the equivalence of the maximum DRM-based EL estimators and the maximum dual EL estimators for both \( \theta \) 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, Zhang, Li, Albanes and Yu (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 \( \psi \), which can be written as 
\[
\psi = \int u(x, \theta) \, dF_0(x)
\]
for a known function \( u(\cdot, \cdot) \), and \( F_0 \) and \( \theta \) 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 \( \psi \), 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 organised as follows. In Section 2, we briefly review the dual EL and present the asymptotical normality of the maximum dual EL estimators of \( \psi \). 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 \( \{x_{ij}: j = 1, \ldots, n_i; i = 0, 1\} \), 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
\[
L(F_0, F_1) = \prod_{i=1}^{n_0} dF_0(x_{0i}) \prod_{j=1}^{n_1} dF_1(x_{1j}) = \left( \prod_{i=0}^{1} \prod_{j=1}^{n_i} p_{ij} \right) \exp \left\{ \sum_{j=1}^{n_1} \theta^\top q(x_{1j}) \right\}.
\]
(2)

The feasible \( p_{ij} \)'s satisfy
\[
p_{ij} \geq 0, \quad \sum_{ij} p_{ij} = 1, \quad \sum_{ij} \exp\{\theta^\top q(x_{ij})\} p_{ij} = 1,
\]
(3)
where \( \sum_{i=0}^{1} \sum_{j=1}^{n_i} p_{ij} \) is written as \( \sum_{ij} \) for short.

Let \( n = n_0 + n_1 \) be the total sample size. By the Lagrange multiplier method, maximising \( \log(L(F_0, F_1)) \) with respect to \( p_{ij} \)'s under the constraint (3) gives
\[
p_{ij} = \frac{1}{n} \frac{1}{1 + \hat{\lambda} \left[ \exp\{\theta^\top q(x_{ij})\} - 1 \right]},
\]
where \( \hat{\lambda} = \hat{\lambda}(\theta) \) is the solution to
\[
\frac{1}{n} \sum_{ij} \frac{\exp\{\theta^\top q(x_{ij})\}}{1 + \hat{\lambda} \left[ \exp\{\theta^\top q(x_{ij})\} - 1 \right]} = 1.
\]
Accordingly, the profile empirical log-likelihood of $\theta$ (up to a constant) is

$$\ell^*_1(\theta) = - \sum_{i,j} \log(1 + \hat{\lambda} \{\exp[\theta^\top q(x_{ij})] - 1\}) + \sum_{j=1}^{n_1} \{\theta^\top q(x_{ij})\}.$$ 

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

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

where $\rho = n_1/n$ takes the place of $\hat{\lambda}$ in $\ell^*_1(\theta)$. See Keziou and Leoni-Aubin (2008) and Chen and Liu (2013). We call $\ell_1(\theta)$ dual EL as Keziou and Leoni-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 maximise. Denote the maximum dual EL estimator of $\theta$ 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 $\theta$ 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 Leoni-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 $\theta_0$ and $\psi_0$ be the true values of $\theta$ and $\psi$, 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 $\theta_0$ for notional simplicity. Let $u_0(x, \theta) = \partial u(x, \theta)/\partial \theta$ and $e_1$ be the $d$-variate vector $(1, 0, \ldots, 0)^\top$. 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 $W = \int q(x)q^\top(x) h_0(x) h_1(x) h(x) \, dF_0(x) + B = \int [h_1(x)u(x)q(x) - u_0(x)] \, dF_0(x)$ and $\Delta = \rho \psi_0 - \int e_1 u_0(x) \, dF_0(x)$. 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 $\theta_0$ and $\int h(x, \theta) \, dF_0(x) < \infty$ in a neighourhood of $\theta_0$.
(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 2.1:** Suppose assumptions (C1)–(C3) are satisfied. As \( n \to \infty \), \( \sqrt{n}(\hat{\psi} - \psi_0) \) is asymptotically normal with mean 0 and variance \( \sigma^2 \), which is defined in (5).

The assumption that \( \int h(x, \theta) \, dF_0(x) < \infty \) in a neighbourhood of \( \theta_0 \) 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 \( \psi \) based on Theorem 2.1, we need a consistent estimator of \( \sigma^2 \). The analytical form of \( \sigma^2 \) motivates us to estimate it by

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

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

\[
\hat{B} = \int \{ h_1(x, \hat{\theta})u(x, \hat{\theta})q(x) - u_\theta(x, \hat{\theta}) \} \, d\hat{F}_0(x),
\]

\[
\hat{W} = \int q(x)q^T(x) h_0(x, \hat{\theta}) h_1(x, \hat{\theta}) h(x, \hat{\theta}) \, d\hat{F}_0(x).
\]

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

\[
\hat{\psi} \pm n^{-1/2} z_{1-\alpha/2} \hat{\sigma},
\]

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 \( \psi \). In other words, for different \( \psi \), the asymptotic variance \( \sigma^2 \) is different only in the last two terms or equivalently \( \psi \) and \( \Delta \). 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 \( u(x, \theta) = x - x \exp\{\theta^T q(x)\} \), leading to \( \Delta = \rho \mu_0 + (1 - \rho) \mu_\theta \).

### 3. EL under DRM

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

\[
\int \{ \psi - u(x, \theta) \} \, dF_0(x) = 0,
\]

which can be taken as another constraint on \( F_0 \). When profiling \( F_0 \) out, we may maximise the logarithm of the likelihood (2) under both constraints (3) and (7). This leads to the standard profile empirical log-likelihood under DRM. Let \( \xi = (\theta^T, \psi)^T \) and \( m(x, \xi) = (u(x, \theta) - \psi, \exp\{\theta^T q(x)\} - 1)^T \). 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
\[ \sum_{ij} \frac{m(x_{ij}, \xi)}{1 + \tilde{\lambda}^\top m(x_{ij}, \xi)} = 0. \]  

The resultant profile empirical log-likelihood of \( \xi \) (up to a constant) is
\[ \ell_2(\xi) \equiv \ell_2(\theta, \psi) = -\sum_{ij} \log\{1 + \tilde{\lambda}^\top m(x_{ij}, \xi)\} + \sum_{j=1}^{n_1}\{\theta^\top q(x_{1j})\}. \]

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

**Theorem 3.1:** 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. In addition, \( \sup_{\xi} \ell_2(\xi) = \sup_\theta \ell_1(\theta) \).

Theorem 3.1 indicates the standard DRM-based EL and the dual EL share their maximum likelihood estimators of \( \theta \) and \( \psi \), and their maximum likelihoods. Hence Theorem 2.1 is also valid for \( \tilde{\psi} \), and the Wald-type interval estimator of \( \psi \) 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 \( \psi \), which is free of variance estimation. Define the EL ratio function of \( \psi \)
\[ R(\psi) = 2 \left\{ \sup{(\theta, \psi)} \ell_2(\theta, \psi) - \sup_\theta \ell_2(\theta, \psi) \right\}. \]

**Theorem 3.2:** Suppose assumptions (C1)–(C3) are satisfied. As \( n \to \infty \), \( R(\psi_0) \) converges in distribution to \( \chi_1^2 \) distribution.

A likelihood ratio interval estimator for \( \psi \) at level \( 1 - \alpha \) can be constructed as
\[ \{\psi : R(\psi) \leq \chi_{1,1-\alpha}^2\}, \]
where \( \chi_{1,1-\alpha}^2 \) is the \( 1 - \alpha \) quantile of the \( \chi_1^2 \) distribution. By Theorems 2.1 and 3.2, 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 $\mu_0$ 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 $\mu_0$ and $\mu_0 - \mu_1$, respectively. The simulated coverage probabilities of their one-sided intervals for both $\mu_0$ and $\mu_0 - \mu_1$ are reported in Tables 2 and 4, respectively.

We first examine the simulation results for $\mu_0$. From Tables 1 and 2, we find that the EL interval has uniformly better coverage accuracy than the DL interval in both one- and

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$(n_0, n_1)$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EL</td>
<td>DL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>90%</td>
<td>95%</td>
<td>99%</td>
<td>90%</td>
<td>95%</td>
<td>99%</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>(20, 20)</td>
<td>Coverage</td>
<td>88.30</td>
<td>94.50</td>
<td>98.50</td>
<td>87.35</td>
<td>92.80</td>
<td>97.85</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>0.73</td>
<td>0.88</td>
<td>1.19</td>
<td>0.70</td>
<td>0.84</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(40, 40)</td>
<td>Coverage</td>
<td>89.80</td>
<td>94.45</td>
<td>99.10</td>
<td>89.20</td>
<td>94.05</td>
<td>98.65</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>0.52</td>
<td>0.62</td>
<td>0.83</td>
<td>0.51</td>
<td>0.61</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(60, 20)</td>
<td>Coverage</td>
<td>90.75</td>
<td>95.10</td>
<td>98.95</td>
<td>90.35</td>
<td>94.75</td>
<td>98.80</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>0.42</td>
<td>0.51</td>
<td>0.68</td>
<td>0.42</td>
<td>0.50</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>(20, 20)</td>
<td>Coverage</td>
<td>88.65</td>
<td>94.30</td>
<td>98.70</td>
<td>87.15</td>
<td>91.95</td>
<td>97.10</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>1.51</td>
<td>1.83</td>
<td>2.53</td>
<td>1.43</td>
<td>1.71</td>
<td>2.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(40, 40)</td>
<td>Coverage</td>
<td>88.50</td>
<td>94.70</td>
<td>99.25</td>
<td>88.35</td>
<td>93.50</td>
<td>98.35</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>1.06</td>
<td>1.28</td>
<td>1.73</td>
<td>1.03</td>
<td>1.23</td>
<td>1.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(60, 20)</td>
<td>Coverage</td>
<td>89.10</td>
<td>94.95</td>
<td>99.05</td>
<td>88.75</td>
<td>94.15</td>
<td>98.30</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>0.86</td>
<td>1.03</td>
<td>1.38</td>
<td>0.84</td>
<td>1.00</td>
<td>1.32</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>(20, 20)</td>
<td>Coverage</td>
<td>80.70</td>
<td>86.70</td>
<td>94.00</td>
<td>79.30</td>
<td>84.55</td>
<td>91.25</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>3.60</td>
<td>4.34</td>
<td>5.81</td>
<td>3.50</td>
<td>4.17</td>
<td>5.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(40, 40)</td>
<td>Coverage</td>
<td>84.35</td>
<td>90.10</td>
<td>96.50</td>
<td>83.85</td>
<td>88.50</td>
<td>93.95</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>2.79</td>
<td>3.37</td>
<td>4.56</td>
<td>2.69</td>
<td>3.20</td>
<td>4.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(60, 20)</td>
<td>Coverage</td>
<td>84.50</td>
<td>91.20</td>
<td>97.35</td>
<td>84.00</td>
<td>89.20</td>
<td>94.70</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>2.36</td>
<td>2.85</td>
<td>3.86</td>
<td>2.28</td>
<td>2.70</td>
<td>3.55</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Coverage probabilities (%) of the one-sided EL and DL intervals for $\mu_0$.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$(n_0, n_1)$</th>
<th>$1 - \alpha$</th>
<th>EL 90%</th>
<th>EL 95%</th>
<th>EL 99%</th>
<th>DL 90%</th>
<th>DL 95%</th>
<th>DL 99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (20,20)</td>
<td>Lower limit</td>
<td>88.80</td>
<td>93.75</td>
<td>98.40</td>
<td>88.55</td>
<td>93.55</td>
<td>98.10</td>
<td></td>
</tr>
<tr>
<td>Upper limit</td>
<td>89.15</td>
<td>94.55</td>
<td>99.05</td>
<td>88.45</td>
<td>93.80</td>
<td>98.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(40,40)</td>
<td>Lower limit</td>
<td>90.00</td>
<td>94.80</td>
<td>98.80</td>
<td>89.90</td>
<td>94.70</td>
<td>98.70</td>
<td></td>
</tr>
<tr>
<td>Upper limit</td>
<td>90.05</td>
<td>95.00</td>
<td>99.05</td>
<td>89.90</td>
<td>94.50</td>
<td>98.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(60,20)</td>
<td>Lower limit</td>
<td>89.20</td>
<td>95.00</td>
<td>99.05</td>
<td>89.30</td>
<td>94.90</td>
<td>99.00</td>
<td></td>
</tr>
<tr>
<td>Upper limit</td>
<td>90.75</td>
<td>95.75</td>
<td>98.85</td>
<td>90.55</td>
<td>95.45</td>
<td>98.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II (20,20)</td>
<td>Lower limit</td>
<td>89.20</td>
<td>94.55</td>
<td>99.05</td>
<td>86.15</td>
<td>91.35</td>
<td>96.30</td>
<td></td>
</tr>
<tr>
<td>Upper limit</td>
<td>90.30</td>
<td>94.10</td>
<td>98.65</td>
<td>91.45</td>
<td>95.80</td>
<td>99.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(40,40)</td>
<td>Lower limit</td>
<td>88.45</td>
<td>93.45</td>
<td>98.90</td>
<td>86.45</td>
<td>92.45</td>
<td>97.40</td>
<td></td>
</tr>
<tr>
<td>Upper limit</td>
<td>89.20</td>
<td>94.15</td>
<td>99.15</td>
<td>90.45</td>
<td>95.90</td>
<td>99.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(60,20)</td>
<td>Lower limit</td>
<td>89.15</td>
<td>94.15</td>
<td>98.70</td>
<td>87.40</td>
<td>92.50</td>
<td>97.70</td>
<td></td>
</tr>
<tr>
<td>Upper limit</td>
<td>89.95</td>
<td>94.95</td>
<td>99.15</td>
<td>91.05</td>
<td>96.25</td>
<td>99.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III (20,20)</td>
<td>Lower limit</td>
<td>78.55</td>
<td>85.05</td>
<td>93.30</td>
<td>75.65</td>
<td>80.65</td>
<td>88.95</td>
<td></td>
</tr>
<tr>
<td>Upper limit</td>
<td>91.10</td>
<td>95.65</td>
<td>98.95</td>
<td>95.25</td>
<td>98.65</td>
<td>99.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(40,40)</td>
<td>Lower limit</td>
<td>83.15</td>
<td>88.75</td>
<td>95.65</td>
<td>79.00</td>
<td>85.45</td>
<td>92.30</td>
<td></td>
</tr>
<tr>
<td>Upper limit</td>
<td>90.55</td>
<td>95.60</td>
<td>99.15</td>
<td>94.90</td>
<td>98.40</td>
<td>99.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(60,20)</td>
<td>Lower limit</td>
<td>82.45</td>
<td>89.05</td>
<td>95.95</td>
<td>79.85</td>
<td>85.20</td>
<td>93.30</td>
<td></td>
</tr>
<tr>
<td>Upper limit</td>
<td>91.05</td>
<td>95.45</td>
<td>99.30</td>
<td>95.20</td>
<td>98.80</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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

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

<table>
<thead>
<tr>
<th>Scenario (( n_0, n_1 ))</th>
<th>1 - ( \alpha )</th>
<th>EL</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90%</td>
<td>95%</td>
<td>99%</td>
</tr>
<tr>
<td>I (20,20)</td>
<td>Lower limit</td>
<td>88.05</td>
<td>94.10</td>
</tr>
<tr>
<td></td>
<td>Upper limit</td>
<td>89.20</td>
<td>94.45</td>
</tr>
<tr>
<td>I (40,40)</td>
<td>Lower limit</td>
<td>90.50</td>
<td>95.30</td>
</tr>
<tr>
<td></td>
<td>Upper limit</td>
<td>89.35</td>
<td>93.90</td>
</tr>
<tr>
<td>I (60,20)</td>
<td>Lower limit</td>
<td>89.85</td>
<td>94.35</td>
</tr>
<tr>
<td></td>
<td>Upper limit</td>
<td>88.80</td>
<td>93.70</td>
</tr>
<tr>
<td>II (20,20)</td>
<td>Lower limit</td>
<td>89.35</td>
<td>94.70</td>
</tr>
<tr>
<td></td>
<td>Upper limit</td>
<td>88.10</td>
<td>93.90</td>
</tr>
<tr>
<td>II (40,40)</td>
<td>Lower limit</td>
<td>90.40</td>
<td>95.15</td>
</tr>
<tr>
<td></td>
<td>Upper limit</td>
<td>89.05</td>
<td>94.65</td>
</tr>
<tr>
<td>II (60,20)</td>
<td>Lower limit</td>
<td>89.50</td>
<td>95.35</td>
</tr>
<tr>
<td></td>
<td>Upper limit</td>
<td>87.25</td>
<td>92.15</td>
</tr>
<tr>
<td>III (20,20)</td>
<td>Lower limit</td>
<td>79.30</td>
<td>86.15</td>
</tr>
<tr>
<td></td>
<td>Upper limit</td>
<td>90.65</td>
<td>95.25</td>
</tr>
<tr>
<td>III (40,40)</td>
<td>Lower limit</td>
<td>84.30</td>
<td>90.00</td>
</tr>
<tr>
<td></td>
<td>Upper limit</td>
<td>91.05</td>
<td>95.05</td>
</tr>
<tr>
<td>III (60,20)</td>
<td>Lower limit</td>
<td>83.70</td>
<td>90.35</td>
</tr>
<tr>
<td></td>
<td>Upper limit</td>
<td>90.65</td>
<td>95.70</td>
</tr>
</tbody>
</table>

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 \( \mu_0 \). 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 \( \mu_0 \) 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 sign(\( \hat{\psi} - \psi_0 \))/\( \sqrt{\hat{R}(\psi_0)} \), 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 $\chi^2_1$ 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 $\chi^2_1$ 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.

5. A real data example

In this section, we further compare the DRM-based EL and dual EL interval estimation methods by analysing the lumber data, which is obtained from tests conducted by the FPInnovations laboratory at the University of British Columbia. The dataset, 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 $\mu_0$ and $\mu_1$ be the means of MOE in 2010 and 2007. We wish to estimate the mean $\mu_0$ and the differences $\mu_0 - \mu_1$. For the choice of the basis function $q(x)$, we consider

\begin{table}[h]
\centering
\begin{tabular}{l l l l}
\hline
Parameter & $q(x)$ & Method & Confidence interval & Interval length \\
$\mu_0$ & $q_1(x)$ & DL & [1.4959, 1.5324] & 0.0365 \\
 & $q_2(x)$ & EL & [1.4960, 1.5327] & 0.0367 \\
 & $q_3(x)$ & DL & [1.4958, 1.5324] & 0.0365 \\
 & $q_3(x)$ & EL & [1.4960, 1.5327] & 0.0367 \\
$\mu_0 - \mu_1$ & $q_1(x)$ & DL & [-0.0244, 0.0420] & 0.0664 \\
 & $q_2(x)$ & EL & [-0.0248, 0.0420] & 0.0668 \\
 & $q_3(x)$ & DL & [-0.0244, 0.0420] & 0.0664 \\
 & $q_3(x)$ & EL & [-0.0248, 0.0420] & 0.0668 \\
\hline
\end{tabular}
\caption{DL and EL interval estimates at confidence level 95\% with $q_1(x) = (1, x, x^2)^T$, $q_2(x) = (1, x, \log(x))^T$ and $q_3(x) = (1, \log(x))^T$.}
\end{table}
$q_1(x)$, $q_2(x)$ 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.

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.

Acknowledgements

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

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 the 111 Project (B14019) and Doctoral Fund of Ministry of Education of China under Grant number 20130076110004.

References


Appendix. Proofs

We begin with a number of technical lemmas. Throughout, we assume the same conditions in Theorem 2.1. For the data \( \{x_{ij} : j = 1, 2, \ldots, n_i ; i = 0, 1\} \), 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 \( S = \{\rho(1 - \rho)\}^{-1} \text{diag}\{1, 0, \ldots, 0\} \) 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 A.1:** For \( Y_n \) and \( Z_n \) defined above, we have

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

(b) \( \mathbb{E}(Y_n) = \psi_0 \) and \( n\mathbb{V}ar(Y_n) = \int (u^2(x)/h(x))\,dF_0(x) - (1/(1 - \rho)^3)c_{00}^2 - (1/\rho(1 - \rho)^2)c_{01}^2; \)

(c) \( \mathbb{C}ov(Y_n, Z_n) = [c_{00}/(1 - \rho)^2 - c_{01}/\rho(1 - \rho)]e_1^TW. \)

**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 \( \mathbb{V}ar(Y_n) \). It can be seen that

\[
n\mathbb{V}ar(Y_n) = (1 - \rho)\mathbb{V}ar\left(\frac{u(x_01)}{h(x_01)}\right) + \rho\mathbb{V}ar\left(\frac{u(x_{11})}{h(x_{11})}\right)
\]

\[
= (1 - \rho)\mathbb{E}\left(\frac{u(x_01)}{h(x_01)}\right)^2 - (1 - \rho)\left[\mathbb{E}\left(\frac{u(x_01)}{h(x_01)}\right)\right]^2
\]

\[
+ \rho\mathbb{E}\left(\frac{u(x_{11})}{h(x_{11})}\right)^2 - \rho\left[\mathbb{E}\left(\frac{u(x_{11})}{h(x_{11})}\right)\right]^2
\]

\[
= \int \frac{u^2(x)}{h^2(x)}h_0(x)h(x)\,dF_0(x) - \frac{c_{00}^2}{(1 - \rho)^3} + \int \frac{u^2(x)}{h^2(x)}h_1(x)h(x)\,dF_0(x) - \frac{c_{01}^2}{\rho(1 - \rho)^2}
\]

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

\[
\mathbb{C}ov(Y_n, Z_n)
\]

\[
= (1 - \rho)\mathbb{C}ov\left(\frac{u(x_01)}{h(x_01)}, -h_1(x_01)q(x_01)\right) + \rho\mathbb{C}ov\left(\frac{u(x_{11})}{h(x_{11})}, h_0(x_{11})q(x_{11})\right)
\]

\[
= (1 - \rho)\left\{-\mathbb{E}\left(\frac{u(x_01)}{h(x_01)}h_1(x_01)q^\top(x_01)\right) + \mathbb{E}\left(\frac{u(x_01)}{h(x_01)}\right)\mathbb{E}\left(h_1(x_01)q^\top(x_01)\right)\right\}
\]

\[
+ \rho\mathbb{E}\left(\frac{u(x_{11})}{h(x_{11})}h_0(x_{11})q^\top(x_{11})\right) - \rho\mathbb{E}\left(\frac{u(x_{11})}{h(x_{11})}\right)\mathbb{E}\left(h_0(x_{11})q^\top(x_{11})\right)
\]

\[
= -\int \frac{u(x)h_1(x)h_0(x)}{h(x)}h(x)\,dF_0(x) + \frac{c_{00}}{(1 - \rho)^2} \int h_1(x)q^\top(x)h_0(x)h(x)\,dF_0(x)
\]

\[
+ \int \frac{u(x)}{h(x)}h_0(x)h_1(x)q^\top(x)h(x)\,dF_0(x) - \frac{c_{01}}{\rho(1 - \rho)} \int h_0(x)h_1(x)q^\top(x)h(x)\,dF_0(x)
\]

\[
= \left\{\frac{c_{00}}{(1 - \rho)^2} - \frac{c_{01}}{\rho(1 - \rho)}\right\} e_1^TW.
\]
Lemma A.2: Assume the same conditions in Theorem 2.1. As $n$ is large, $\hat{\theta} - \theta_0 = W^{-1}Z_n/n + o_p(n^{-1/2})$.

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

**Proof of Theorem 2.1**

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

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

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,

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

where $B$ is defined above Theorem 2.1. By Lemma A.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 A.1, we have

$$\operatorname{Cov}(Y_n, B^\top W^{-1}Z_n) = \begin{bmatrix} c_{00} & c_{01} \\ \frac{c_{01}}{\rho(1-\rho)^2} & \frac{c_{01}}{\rho(1-\rho)} \end{bmatrix} e_1^\top W(W^{-1}B)$$

$$= \begin{bmatrix} c_{00} & c_{01} \\ \frac{c_{01}}{\rho(1-\rho)^2} & \frac{c_{01}}{\rho(1-\rho)} \end{bmatrix} e_1 e_1^\top B.$$

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

$$n\operatorname{Var}(B^\top W^{-1}Z_n/n) = B^\top W^{-1}(W - WSW)W^{-1}B$$

$$= B^\top W^{-1}B - [\rho(1-\rho)]^{-1}B e_1 e_1^\top B.$$

Putting all these terms together, we have

$$\sigma^2 = n\operatorname{Var}(Y_n - \psi_0 - B^\top W^{-1}Z_n/n)$$

$$= n\operatorname{Var}(Y_n) - \operatorname{Cov}(Y_n, B^\top W^{-1}Z_n) - \operatorname{Cov}(B^\top W^{-1}Z_n, Y_n)$$

$$+ B^\top W^{-1}\operatorname{Var}(Z_n) W^{-1}B$$

$$= \int \frac{u^2(x)}{h(x)} dF_0(x) - \frac{1}{(1-\rho)^2} c_{00} - \frac{1}{\rho(1-\rho)^2} c_{01}$$

$$- \left\{ \frac{c_{00}}{(1-\rho)^2} - \frac{c_{01}}{\rho(1-\rho)} \right\} e_1 B - B^\top e_1 \left\{ \frac{c_{00}}{(1-\rho)^2} - \frac{c_{01}}{\rho(1-\rho)} \right\}^\top$$

$$+ B^\top W^{-1}B - [\rho(1-\rho)]^{-1}B e_1 e_1^\top B$$

$$= \int \frac{u^2(x)}{h(x)} dF_0(x) + B^\top W^{-1}B - \frac{\eta}{\rho(1-\rho)},$$
where \( \eta = (B^\top e_1 + (\rho/(1-\rho))c_{00} - c_{01})^2 + (\rho/(1-\rho))(c_{00} + c_{01})^2. \) Since \( B^\top e_1 = \int [h_1(x)u(x) - e_1^\top u_0(x)] \, dF_0(x), \) we have
\[
B^\top e_1 + \frac{\rho}{1 - \rho} c_{00} - c_{01} = \rho \psi - \int e_1^\top u_0(x, \theta) \, dF_0(x) = \Delta.
\]
Further using \( c_{00} + c_{01} = (1 - \rho) \psi \) gives
\[
\eta = \rho (1 - \rho) \psi^2 - \Delta^2 / \rho(1 - \rho).
\]
This proves the expression of \( \sigma^2 \) and Theorem 2.1. ■

**Proof of Theorem 3.1**

**Proof:** Partition \( \lambda \) 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
\[
\frac{\partial \ell_2(\theta, \psi)}{\partial \theta} = -\sum_{i,j} \frac{\lambda_1 u_{ij}(x_{ij}, \theta) + \lambda_2 \exp[\theta^\top q(x_{ij})] q(x_{ij})}{1 + \lambda_1 [u(x_{ij}, \theta) - \psi] + \lambda_2[\exp[\theta^\top q(x_{ij})] - 1]} = 0, \tag{A1}
\]
\[
\frac{\partial \ell_2(\theta, \psi)}{\partial \psi} = \sum_{i,j} \frac{\lambda_1}{1 + \lambda_1 [u(x_{ij}, \theta) - \psi] + \lambda_2[\exp[\theta^\top q(x_{ij})] - 1]} = 0. \tag{A2}
\]
When \( \psi = \tilde{\psi} \) and \( \theta = \tilde{\theta} \), it follows from Equations (8) and (A2) that \( \tilde{\lambda}_1(\tilde{\theta}, \tilde{\psi}) = 0 \). This implies that \( (\tilde{\theta}, \tilde{\psi}, \tilde{\lambda}_2) \) is simply the solution to
\[
-\sum_{i,j} \frac{\lambda_2 \exp[\theta^\top q(x_{ij})] q(x_{ij})}{1 + \lambda_2[\exp[\theta^\top q(x_{ij})] - 1]} = 0,
\]
\[
\sum_{i,j} \frac{u(x_{ij}, \theta) - \psi}{1 + \lambda_2[\exp[\theta^\top q(x_{ij})] - 1]} = 0,
\]
\[
\sum_{i,j} \frac{\exp[\theta^\top q(x_{ij})] - 1}{1 + \lambda_2[\exp[\theta^\top q(x_{ij})] - 1]} = 0,
\]
which are Equations (8) and (A1) with \( \lambda_1 \) replaced by 0. It is trivial to verify that the equations are exactly those which define \( (\tilde{\theta}, \tilde{\psi}, \tilde{\lambda}). \) 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 \( \sup_\xi \ell_2(\xi) = \sup_\theta \ell_1(\theta). \) ■

**Proof of Theorem 3.2**

**Proof:** It can be verified that \( \ell_2(\xi) = \min_\lambda H(\lambda, \theta, \psi), \) where
\[
H(\lambda, \theta, \psi) = -\sum_{i,j} \log[1 + \lambda^\top m(x_{ij}, \xi)] + \sum_{j=1}^{n_j} \{\theta^\top q(x_{ij})\}.
\]
Since \( \xi \) is root-\( n \) consistent, we need to study the behaviour 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 \( \xi^\top = (\xi_1, \xi_2) = n^{1/2}(\lambda - \lambda_0)^\top, (\xi - \xi_0)^\top = n^{1/2}(\lambda - \lambda_0)^\top, (\theta - \theta_0)^\top, (\psi - \psi_0)^\top \), where \( \xi_1 = n^{1/2}(\lambda - \lambda_0)^\top, (\theta - \theta_0)^\top \) and \( \xi_2 = n^{1/2}(\psi - \psi_0)^\top \). Define
\[
Q(\xi) = H(\xi_0 + n^{-1/2}\xi_1, \psi_0 + n^{-1/2}\xi_2)
\]
with \( \xi_{10} = (\lambda_0, \theta_0). \)
Consequently, after length algebra, we can show that

\[ R_n = O_p(n^{-1/2}) \text{ and } C_n = C + o_p(1), \]

where \( C = \mathbb{E}(C_n) \). Applying the second-order Taylor expansion to \( Q(\xi) \), we have

\[ Q(\xi) = H(\xi_{10}, \psi_0) + A^\top \xi + \frac{1}{2} \xi^\top C_\xi + o_p(1). \]

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

\[ \sup \ell_2(\xi) = Q(\hat{\xi}) = H(\xi_{10}, \psi_0) - \frac{1}{2} A^\top C^{-1}A + o_p(1). \]

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

\[ \hat{\xi}_1 = -C_{11}^{-1}A_1 + o_p(1). \]

Consequently,

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

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

\[ C^{-1} = \begin{pmatrix} C_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} C_{11}^{-1}C_{12} \\ C_{21}C_{11}^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ -I \end{pmatrix}, \]

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

\[ A^\top C^{-1}A = A_1^\top C_{11}^{-1}A_1^\top + \left( A_1, A_2 \right) \begin{pmatrix} C_{11}^{-1}C_{12} \\ -I \end{pmatrix} C_{22,1}^{-1} \begin{pmatrix} C_{21}C_{11}^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ -I \end{pmatrix} \begin{pmatrix} A_1^\top \\ A_2^\top \end{pmatrix}. \]

Consequently,

\[ R_2(\psi_0) = \left( A_1C_{11}^{-1}C_{12} - A_2 \right) \left( A_1C_{11}^{-1}C_{12} - A_2 \right)^\top + o_p(1). \]

After length algebra, we can show that

\[ \sqrt{n}(A_1C_{11}^{-1}C_{12} - A_2) \xrightarrow{d} N(0, C_{22,1}) \]

which directly implies \( R_2(\psi_0) \xrightarrow{d} \chi^2_1 \). This completes the proof.