

HYPOTHESIS TESTING IN THE PRESENCE OF MULTIPLE SAMPLES UNDER DENSITY RATIO MODELS

Song Cai¹, Jiahua Chen² and James V. Zidek²

¹*Carleton University and* ²*University of British Columbia*

Abstract: This paper presents a hypothesis testing method given independent samples from a number of connected populations. The method is motivated by a forestry project for monitoring change in the strength of lumber. Traditional practice has been built upon nonparametric methods which ignore the fact that these populations are connected. By pooling the information in multiple samples through a density ratio model, the proposed empirical likelihood method leads to more efficient inferences and therefore reduces the cost in applications. The new test has a classical chi-square null limiting distribution. Its power function is obtained under a class of local alternatives. The local power is found increased even when some underlying populations are unrelated to the hypothesis of interest. Simulation studies confirm that this test has better power properties than potential competitors, and is robust to model misspecification. An application example to lumber strength is included.

Key words and phrases: Dual empirical likelihood, empirical likelihood ratio test, information pooling, local power, long term monitoring, lumber quality, semiparametric inference.

1. Introduction

The paper presents a method for testing hypotheses about parameters of a given number of different population distributions with independent samples from each. The method was created as part of a research program aimed at developing statistical theory for monitoring change in the strength of lumber. Interest in such a program has been sparked by climate change, which will affect the way trees grow, as well by the changing resource mix, for example due to increasing reliance on plantation lumber. Added impetus comes from the increasing importance of wood as a construction material due to its sustainability as a building material. Moreover, the worldwide forest products industry is vast.

Desiderata for the statistical methods used in the long-term monitoring program of lumber includes two key goals. First the methods must be efficient to reduce the sizes of the required samples: testing lumber costs time and money. Toward the goal of efficiency, this paper proposes a method that borrows strength across the multiple samples by exploiting an obvious feature of the resource, that

distinct populations of lumber over years, species, regions and so on will share some latent strength characteristics. Second the methods should ideally be non-parametric in accordance with the well-ingrained practice in setting standards for forest products like those in American Society for Testing and Materials (ASTM) protocols (ASTM D1990 – 07).

These desiderata lead to the semiparametric density ratio model (DRM) adopted in this paper. More precisely, suppose we have $m+1$ lumber populations with cumulative distribution functions (CDFs) $F_k(x)$, $k = 0, \dots, m$. We link them through the DRM assumption:

$$dF_k(x) = \exp \{ \alpha_k + \boldsymbol{\beta}_k^\top \mathbf{q}(x) \} dF_0(x), \quad (1.1)$$

where x could be a single-valued or vector-valued variable, $\mathbf{q}(x)$, the *basis function*, is a prespecified d -dimensional function, and $\boldsymbol{\theta}_k^\top = (\alpha_k, \boldsymbol{\beta}_k^\top)$ are model parameters. The baseline distribution $F_0(x)$ is completely unspecified.

The DRM is flexible and covers many commonly used distribution families, including each member of the exponential family. For example, normal distributions $N(\mu_k, \sigma_k)$, $k = 0, \dots, m$, satisfy a DRM with basis function $\mathbf{q}(x) = (x, x^2)^\top$ and corresponding parameters $\alpha_k = \log(\sigma_0/\sigma_k) + \mu_0^2/(2\sigma_0^2) - \mu_k^2/(2\sigma_k^2)$, $\boldsymbol{\beta}_k = (\mu_k/\sigma_k^2 - \mu_0/\sigma_0^2, 1/(2\sigma_0^2) - 1/(2\sigma_k^2))^\top$. There is a close relationship between the logistic regression model in case-control studies and the two-sample DRM (Qin and Zhang (1997)).

The empirical likelihood (EL) is a natural platform for data analysis that in recent years has been widely studied in the context of DRM. See Chen and Liu (2013) and Zhang (2000) for quantile estimation, Fokianos (2004) for density estimation, and Keziou and Leoni-Aubin (2008) for a two-sample EL ratio test. Investigating the properties of tests constructed under the DRM assumption proves challenging since the parameters under the null hypothesis are often not interior points of the parameter space. Thus, the limiting distribution of the EL-based likelihood ratio test cannot be derived from the usual approach such as the ones given in Owen (2001) or Qin (1998). Instead, we study the properties of the *dual empirical likelihood ratio* (DELRL) test. We show that the proposed test statistic has a classical chi-square null limiting distribution under fairly general conditions. We further study its power function under a class of local alternatives and find that this local power often increases when additional samples are included in the data analysis even when their distributions are not related to the hypothesis. This result supports the use of the DRM for pooling information across multiple samples. Under a broad range of distributional settings, our simulations show that the proposed DELRL test is more powerful in detecting distributional changes over samples than many classical tests. The

new method is also found to be model robust: its size and power are resistant to mild violations to the DRM assumption.

An anonymous referee suggested the semi-parametric proportional hazards model (CoxPH) proposed by Cox (1972) as an alternative for analyzing multiple samples. The limitation of the CoxPH approach for multiple samples may be seen in the simulation results included in Sections 5.3 and 5.4. The power of the partial likelihood ratio test under the CoxPH is comparable to that of DRM approach when its proportional hazards assumption is true. Otherwise, the DRM-based test has a higher power.

The paper is organized as follows. We first review the EL methodology for multiple samples under the DRM. We then motivate the use of dual EL to overcome the associated boundary problem. In Section 3, we obtain the limiting distributions of the DELR statistic under various null hypotheses and local alternatives. Section 4 studies the effect of information pooling on power properties of the DELR test. The finite sample properties of the DELR test are assessed via simulation in Section 5. An application to lumber strength is given in Section 6. The simulation details are presented in the Appendix and the proofs are given in the supplementary material.

2. EL under the DRM

Denote the observations in the $m + 1$ samples as $\{x_{kj} : j = 1, \dots, n_k\}_{k=0}^m$ where $n_k > 0$ is the size of the k^{th} sample. Suppose these samples are independent of each other, and the observations within each sample are independent and identically distributed (iid). We denote the total sample size as $n = \sum_k n_k$. Let $dF_k(x) = F_k(x) - F_k(x^-)$, and put $p_{kj} = dF_0(x_{kj})$. For convenience, take $\alpha_0 = 0$, $\beta_0 = \mathbf{0}$, and $\theta_0 = (\alpha_0, \beta_0^T)^T$. Under the DRM assumption (1.1), the EL of the $\{F_k\}$ is defined to be

$$L_n(F_0, \dots, F_m) = \prod_{k,j} dF_k(x_{kj}) = \left\{ \prod_{k,j} p_{kj} \right\} \cdot \exp \left\{ \sum_{k,j} (\alpha_k + \beta_k^T \mathbf{q}(x_{kj})) \right\},$$

where the sum and product are over all possible (k, j) combinations. The DRM assumption and the fact that the $\{F_k\}$ are distribution functions imply that

$$1 = \int dF_k(x) = \int \exp\{\alpha_k + \beta_k^T \mathbf{q}(x)\} dF_0(x). \tag{2.1}$$

Let $\alpha = (\alpha_1, \dots, \alpha_m)^T$, $\beta = (\beta_1^T, \dots, \beta_m^T)^T$, and $\theta = (\alpha^T, \beta^T)^T$. We write the EL as $L_n(\theta, F_0)$. When there are tied observations, say $x_{kj} = x_{k't'j'}$ for some $(k, j) \neq (k', j')$, the L_n defined above may be interpreted as the limit of the EL based on $y_{kj} = x_{kj} + \delta\epsilon_{kj}$ as $\delta \rightarrow 0$, where ϵ_{kj} are some iid continuous random variables. With this convention, the L_n is well defined and motivated. We are

assured that inferences based on this L_n are the same as those based on the EL formulated with tied observations; unnecessary mathematical complexity is all that we lose (Owen (2001, Sec. 2.3)).

The maximum EL estimator (MELE) of $\boldsymbol{\theta}$ and F_0 is the maximum point of $L_n(\boldsymbol{\theta}, F_0)$ over the space of $\boldsymbol{\theta}$ and F_0 such that (2.1) is satisfied. For theoretical discussion and numerical computation, the maximization is carried out in two steps. First, we define the *profile log-EL*:

$$\tilde{l}_n(\boldsymbol{\theta}) = \sup \left\{ \log L_n(\boldsymbol{\theta}, F_0) : \sum_{k,j} \exp\{\alpha_r + \boldsymbol{\beta}_r^\top \mathbf{q}(x_{kj})\} p_{kj} = 1, r = 0, \dots, m, \right\}$$

where the supremum is over the space of F_0 with fixed $\boldsymbol{\theta}$. Based on the method of Lagrange multipliers, the supremum is found to be attained when

$$p_{kj}(\{\lambda_r\}, \boldsymbol{\theta}) = n^{-1} \left\{ 1 + \sum_{r=1}^m \lambda_r [\exp\{\alpha_r + \boldsymbol{\beta}_r^\top \mathbf{q}(x_{kj})\} - 1] \right\}^{-1}, \quad (2.2)$$

where the Lagrange multipliers $\{\lambda_r\}$ solve, for $t = 0, \dots, m$,

$$\sum_{k,j} \exp\{\alpha_t + \boldsymbol{\beta}_t^\top \mathbf{q}(x_{kj})\} p_{kj}(\{\lambda_r\}, \boldsymbol{\theta}) = 1. \quad (2.3)$$

The profile log-EL can hence be written as

$$\tilde{l}_n(\boldsymbol{\theta}) = - \sum_{k,j} \log \left\{ 1 + \sum_{r=1}^m \lambda_r [\exp\{\alpha_r + \boldsymbol{\beta}_r^\top \mathbf{q}(x_{kj})\} - 1] \right\} + \sum_{k,j} \{\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{q}(x_{kj})\}.$$

In the sequel, $p_{kj}(\{\lambda_r\}, \boldsymbol{\theta})$ will be simplified to p_{kj} if it does not lead to any confusion.

The MELE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is then the point at which $\tilde{l}_n(\boldsymbol{\theta})$ is maximized. Given $\hat{\boldsymbol{\theta}}$, we solve for the Lagrange multipliers $\hat{\lambda}_r$ through (2.3). Interestingly, we always have $\hat{\lambda}_r = n_r/n$. Subsequently, we obtain \hat{p}_{kj} by plugging $\hat{\boldsymbol{\theta}}$ and $\hat{\lambda}_k$ into (2.2). Finally, the MELEs of the $\{F_k\}$ are given by

$$\hat{F}_k(x) = n^{-1} \sum_{r,j} \exp\{\hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^\top \mathbf{q}(x_{rj})\} \hat{p}_{rj} \mathbf{1}(x_{rj} \leq x),$$

where $\mathbf{1}(\cdot)$ is the indicator function.

In applications, giving a point estimation is a minor part of the data analysis. Assessing the uncertainty in the point estimator and testing hypotheses are judged of greater practical importance. Asymptotic properties of the point estimator and the likelihood function enable this, but classical asymptotic theories usually rely on differential properties of the likelihood function in the neighbourhood of the true parameter value. Consequently these results are applicable only if this neighbourhood lies in the parameter space.

According to (2.1), α_k is a normalizing constant satisfying

$$\alpha_k = -\log \int \exp\{\beta_k^\top \mathbf{q}(x)\} dF_0(x).$$

Thus, $\alpha_k = 0$ whenever $\beta_k = \mathbf{0}$. When the true value $\theta_1 = \mathbf{0}$, its neighborhood is not contained in the parameter space and DRM is not regular at this θ (Zou, Fine and Yandell (2002)). Regularity is also violated when $\beta_1 = \beta_2$ which implies $\alpha_1 = \alpha_2$. In our application, θ_k is the parameter of the lumber population at year k and $\theta_1 = \theta_2$ signifies the stability of the wood quality over these two years. Non-regularity denies a straightforward application of the EL ratio test to this hypothesis, creating a need for other effective inferential methods.

3. Dual EL and Its Properties

With $\theta = \hat{\theta}$, we have $\hat{\lambda}_r = n_r/n$. Hence,

$$l_n(\theta) = -\sum_{k,j} \log \left\{ \sum_{r=0}^m \frac{n_r}{n} \exp\{\alpha_r + \beta_r^\top \mathbf{q}(x_{kj})\} \right\} + \sum_{k,j} \{\alpha_k + \beta_k^\top \mathbf{q}(x_{kj})\}, \quad (3.1)$$

shares the same maximum point as well as maximum value as the profile log-EL $\tilde{l}(\theta)$. Therefore the MELE $\hat{\theta}$ is also given by $\hat{\theta} = \underset{\theta}{\operatorname{argmax}} l_n(\theta)$. Keziou and Leoni-Aubin (2008) derived $l_n(\theta)$ under a two-sample DRM by solving the dual problem of profiling the EL. Thus we call $l_n(\theta)$ the dual empirical likelihood (DEL) function. Compared to the EL under the DRM assumption, the DEL is well-defined for any θ in the corresponding Euclidean space, has a simple analytical form, and is concave. Under a two-sample DRM ($m = 1$), Keziou and Leoni-Aubin found that the corresponding likelihood ratio test statistic has the usual chi-square limiting distribution for $H_0 : \beta_1 = \mathbf{0}$, but this result does not apply when, for example, there are $m + 1 = 5$ samples and the hypothesis is

$$H_0 : \beta_1 = \mathbf{0} \quad \text{against} \quad H_1 : \beta_1 \neq \mathbf{0}, \quad (3.2)$$

where β_2, β_3 , and β_4 are nuisance parameters that do not appear in the hypothesis, or when

$$H_0 : \beta_1 = \mathbf{0} \text{ and } \beta_2 = \beta_3 \quad \text{against} \quad H_1 : \beta_1 \neq \mathbf{0} \text{ or } \beta_2 \neq \beta_3. \quad (3.3)$$

These are problems of interest in our lumber quality monitoring project.

Many of our inferential problems can be abstractly stated as testing

$$H_0 : \mathbf{g}(\beta) = \mathbf{0} \quad \text{against} \quad H_1 : \mathbf{g}(\beta) \neq \mathbf{0} \quad (3.4)$$

for some smooth function $\mathbf{g} : \mathbb{R}^{md} \rightarrow \mathbb{R}^q$, with $q \leq md$, the length of β , with m the number of non-baseline distributions, and d the dimension of the basis

function $\mathbf{q}(x)$. Here we assume that \mathbf{g} is thrice differentiable with a full rank Jacobian matrix $\partial\mathbf{g}/\partial\boldsymbol{\beta}$. The parameters $\{\alpha_k\}$ are usually not a part of the hypothesis, because their values are fully determined by the $\{\boldsymbol{\beta}_k\}$ and F_0 under the DRM assumption, although they are treated as independent parameters in the DEL.

Let $\tilde{\boldsymbol{\theta}}$ be the point at which the maximum of $l_n(\boldsymbol{\theta})$ is attained under the constraint $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}$. The DELR test statistic is defined to be

$$R_n = 2\{l_n(\hat{\boldsymbol{\theta}}) - l_n(\tilde{\boldsymbol{\theta}})\}.$$

Theorem 1. *Given $m+1$ random samples from populations with distributions of the DRM form (1.1) and a true parameter value $\boldsymbol{\theta}^*$ such that $\int \exp\{\boldsymbol{\beta}_k^\top \mathbf{q}(x)\} dF_0(x) < \infty$ for $\boldsymbol{\theta}$ in a neighbourhood of $\boldsymbol{\theta}^*$, $\int \mathbf{Q}(x)\mathbf{Q}^\top(x) dF_0(x)$ is positive definite with $\mathbf{Q}^\top(x) = (1, \mathbf{q}^\top(x))$, and $\hat{\lambda}_k = n_k/n = \rho_k + o(1)$ for some constant $\rho_k \in (0, 1)$.*

Under the null hypothesis $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}$, $R_n \rightarrow \chi_q^2$ in distribution as $n \rightarrow \infty$, where χ_q^2 is a chi-square random variable with q degrees of freedom.

The proof of Theorem 1 is given in the supplementary material. When $m = 1$ and $\mathbf{g}(\boldsymbol{\beta}) = \boldsymbol{\beta}_1$, Theorem 1 reduces to the result of Keziou and Leon-Aubin (2008). Theorem 1 covers additional ground, for instance, the hypothesis testing problems (3.2) and (3.3).

The null limiting distribution given by Theorem 1 is useful for approximating the p-value of a DELR test about $\boldsymbol{\beta}$ and for constructing approximate confidence regions for $\boldsymbol{\beta}$. Let $\mathbf{g}(\boldsymbol{\beta}) = \boldsymbol{\beta} - \boldsymbol{\beta}^*$ for a given value $\boldsymbol{\beta}^*$. Then the MELE of $\boldsymbol{\theta}$ under the constraint $\mathbf{g}(\boldsymbol{\beta}) = \boldsymbol{\beta} - \boldsymbol{\beta}^* = \mathbf{0}$ is a function of $\boldsymbol{\beta}^*$, and we denote it as $\tilde{\boldsymbol{\theta}}(\boldsymbol{\beta}^*)$. Consequently, the DELR statistic is a function of $\boldsymbol{\beta}^*$ given by $R_n(\boldsymbol{\beta}^*) = 2\{l_n(\hat{\boldsymbol{\theta}}) - l_n(\tilde{\boldsymbol{\theta}}(\boldsymbol{\beta}^*))\}$. When $\boldsymbol{\beta}^*$ is the true $\boldsymbol{\beta}$ parameter value, the limiting distribution of $R_n(\boldsymbol{\beta}^*)$ is χ_{md}^2 by Theorem 1. Hence an approximate $\alpha \times 100\%$, $0 < \alpha < 1$, confidence region for $\boldsymbol{\beta}^*$ is given by $\{\boldsymbol{\beta}^* : R_n(\boldsymbol{\beta}^*) \leq \chi_{md, \alpha}^2\}$, where $\chi_{md, \alpha}^2$ is the α th quantile of the χ_{md}^2 distribution.

For the power of the DELR test, we study the limiting distribution of R_n at a local alternative. Let $\{\boldsymbol{\beta}_k^*\}$ be a set of parameter values which form a null model satisfying $H_0 : \mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}$ under the DRM assumption. Let

$$\boldsymbol{\beta}_k = \boldsymbol{\beta}_k^* + n_k^{-1/2} \mathbf{c}_k \tag{3.5}$$

for some constants $\{\mathbf{c}_k\}$ be a set of parameter values which form a local alternative. We denote the distribution functions corresponding to $\boldsymbol{\beta}_k^*$ and $\boldsymbol{\beta}_k$ as F_k and G_k with $G_0 = F_0$, respectively. As $n \rightarrow \infty$, the limiting distribution of R_n under this local alternative is usually non-degenerate and provides useful information on the power of the test.

Let $U_n = -n^{-1}\partial^2 l_n(\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ be the empirical information matrix. Its almost sure limit under H_0 is a symmetric positive definite matrix that can be regarded as an information matrix U . We partition the entries of U in agreement with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and represent them as $U_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$, $U_{\boldsymbol{\alpha}\boldsymbol{\beta}}$, $U_{\boldsymbol{\beta}\boldsymbol{\alpha}}$, and $U_{\boldsymbol{\beta}\boldsymbol{\beta}}$. Let $\varphi_k(\boldsymbol{\theta}, x) = \exp\{\alpha_k + \boldsymbol{\beta}_k^\top \mathbf{q}(x)\}$, $k = 0, \dots, m$, and

$$\begin{aligned} \mathbf{h}(\boldsymbol{\theta}, x) &= (\rho_1\varphi_1(\boldsymbol{\theta}, x), \dots, \rho_m\varphi_m(\boldsymbol{\theta}, x))^\top, \\ s(\boldsymbol{\theta}, x) &= \rho_0 + \sum_{k=1}^m \rho_k\varphi_k(\boldsymbol{\theta}, x), \\ H(\boldsymbol{\theta}, x) &= \text{diag}\{\mathbf{h}(\boldsymbol{\theta}, x)\} - \frac{\mathbf{h}(\boldsymbol{\theta}, x)\mathbf{h}^\top(\boldsymbol{\theta}, x)}{s(\boldsymbol{\theta}, x)}. \end{aligned} \tag{3.6}$$

Let $E_0(\cdot)$ be the expectation operator with respect to F_0 . Then, the blockwise algebraic expressions of the information matrix U in terms of $H(\boldsymbol{\theta}^*, x)$ and $\mathbf{q}(x)$ can be written as

$$\begin{aligned} U_{\boldsymbol{\alpha}\boldsymbol{\alpha}} &= E_0\{H(\boldsymbol{\theta}^*, x)\}, \\ U_{\boldsymbol{\beta}\boldsymbol{\beta}} &= E_0\left\{H(\boldsymbol{\theta}^*, x) \otimes (\mathbf{q}(x)\mathbf{q}^\top(x))\right\}, \\ U_{\boldsymbol{\alpha}\boldsymbol{\beta}} &= U_{\boldsymbol{\beta}\boldsymbol{\alpha}}^\top = E_0\{H(\boldsymbol{\theta}^*, x) \otimes \mathbf{q}^\top(x)\}, \end{aligned} \tag{3.7}$$

where \otimes is the Kronecker product operator. We partition the Jacobian matrix of $\mathbf{g}(\boldsymbol{\beta})$ evaluated at $\boldsymbol{\beta}^*$, $\nabla = \partial\mathbf{g}(\boldsymbol{\beta}^*)/\partial\boldsymbol{\beta}$, into (∇_1, ∇_2) , with q and $md - q$ columns respectively. Without loss of generality, we assume that ∇_1 has a full rank. Let I_k be an identity matrix of size $k \times k$ and $J^\top = (-\nabla_1^{-1}\nabla_2)^\top, I_{md-q}$.

Theorem 2. *Under the conditions of Theorem 1 and the local alternative defined by (3.5), $R_n \rightarrow \chi_q^2(\delta^2)$ in distribution as $n \rightarrow \infty$, where $\chi_q^2(\delta^2)$ is a non-central chi-square random variable with q degrees of freedom and a nonnegative non-centrality parameter*

$$\delta^2 = \begin{cases} \boldsymbol{\eta}^\top \{\Lambda - \Lambda J(J^\top \Lambda J)^{-1} J^\top \Lambda\} \boldsymbol{\eta}, & \text{if } q < md, \\ \boldsymbol{\eta}^\top \Lambda \boldsymbol{\eta}, & \text{if } q = md, \end{cases}$$

where $\boldsymbol{\eta}^\top = (\rho_1^{-1/2} \mathbf{c}_1^\top, \rho_2^{-1/2} \mathbf{c}_2^\top, \dots, \rho_m^{-1/2} \mathbf{c}_m^\top)$ and $\Lambda = U_{\boldsymbol{\beta}\boldsymbol{\beta}} - U_{\boldsymbol{\beta}\boldsymbol{\alpha}} U_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{-1} U_{\boldsymbol{\alpha}\boldsymbol{\beta}}$. Moreover, $\delta^2 > 0$ except when $\boldsymbol{\eta}$ is in the column space of J .

The proof is given in the supplementary material. The following example demonstrates one usage of this result: computing local power of the DELR test under a given distributional setting.

Example 1 (Computing the local power of the DELR test for a composite hypothesis). Consider the situation where $m + 1 = 3$ samples are from a DRM with

basis function $\mathbf{q}(x) = (x, \log x)^\top$, and the sample proportions are $(0.4, 0.3, 0.3)$. Let F_k , $k = 1, 2$, be the distributions with parameters $\beta_1^* = (-1, 1)^\top$ and $\beta_2^* = (-2, 2)^\top$. Suppose H_0 is $\mathbf{g}(\beta) = 2\beta_1 - \beta_2 = 0$. Consider the local alternative

$$\beta_k = \beta_k^* + n_k^{-1/2} \mathbf{c}_k, \text{ for } k = 1, 2, \quad (3.8)$$

with $\mathbf{c}_1 = (2, 3)^\top$ and $\mathbf{c}_2 = (-1, 0)^\top$.

Under the above settings, we find $\nabla = (2I_2, -I_2)$ so $J = ((1/2)I_2, I_2)$, and $\boldsymbol{\eta} \approx (3.65, 5.48, -1.83, 0)^\top$. The information matrix U is F_0 dependent. When F_0 is $\Gamma(2, 1)$, where in general $\Gamma(\lambda, \kappa)$ denotes the gamma distribution with shape λ and rate κ , we obtain the information matrix (3.7) and hence Λ , based on numerical computation. We therefore get $\delta^2 \approx 10.29$.

Let $\chi_{d,p}^2$ denotes the p^{th} quantile of the χ_d^2 distribution. The null limiting distribution of R_n is χ_2^2 . Thus at the 5% significance level, the null hypothesis is rejected when $R_n \geq \chi_{2,0.95}^2 \approx 5.99$. Therefore at the current local alternative, the power of the DELR test is approximately $P(\chi_2^2(10.29) \geq 5.99) \approx 0.83$.

Theorem 2 is also useful for sample size calculation.

Example 2 (Sample size calculation for Example 1). Adopt the settings of Example 1. Suppose we require the power of the DELR test to be at least 0.8 at the alternative of $\beta_1 = \beta_1^* + (0.5, 1.5)^\top$ and $\beta_2 = \beta_2^* + (0.5, 0.5)^\top$ at the 5% significance level. This alternative corresponds to a local alternative of the form (3.8) with $\mathbf{c}_1 = (0.5\sqrt{n_1}, 1.5\sqrt{n_1})^\top = 0.5(\sqrt{0.3n}, 3\sqrt{0.3n})^\top$ and $\mathbf{c}_2 = (0.5\sqrt{n_2}, 0.5\sqrt{n_2})^\top = 0.5(\sqrt{0.3n}, \sqrt{0.3n})^\top$.

Using \mathbf{c}_1 , \mathbf{c}_2 , and sample proportions $(0.4, 0.3, 0.3)$, we obtain $\boldsymbol{\eta} = (0.3^{-1/2}\mathbf{c}_1^\top, 0.3^{-1/2}\mathbf{c}_2^\top)^\top = 0.5\sqrt{n}(1, 3, 1, 1)^\top$ as a function of the total sample size n . With the J , F_0 , and U obtained in Example 1, and applying the formula given in Theorem 2, we obtain the non-centrality parameter $\delta^2(n)$ as a function of n . We find that when $n \geq 50$, $P(\chi_2^2(\delta^2(n)) \geq \chi_{2,0.95}^2) \geq 0.8$.

Theorem 2 is also an effective tool for comparing the local powers of DELR tests formulated in different ways. The comparison helps us to determine the most efficient use of information contained in multiple samples. This point is discussed in the next section.

4. Power Properties of the DELR Test under the DRM

Our use of DRM is motivated by its ability to pool information across multiple samples. In general, we expect that the DELR test has higher power when more random samples included in the DRM. This section provides rigorous evidence of this.

Suppose we observe $m + 1$ random samples from distributions satisfying the DRM assumption (1.1), but a hypothesis testing problem of interest focuses on a characteristic of $r + 1$ of them, where $r < m$. Generally, we find that a DELR test based only on the $r + 1$ samples is less informative than the one based on all the $m + 1$ samples.

Without loss of generality, consider a null hypothesis regarding subpopulations F_0, \dots, F_r with $r < m$ and let $\zeta^\top = (\beta_1^\top, \dots, \beta_r^\top)$. The composite hypotheses are specified as

$$H_0 : \mathbf{g}(\zeta) = \mathbf{0} \quad \text{against} \quad H_1 : \mathbf{g}(\zeta) \neq \mathbf{0} \tag{4.1}$$

for some smooth function $\mathbf{g} : \mathbb{R}^{rd} \rightarrow \mathbb{R}^q$ with $q \leq rd$. A DELR test can be based either on samples from just F_0, \dots, F_r , or on the samples from all the populations F_0, \dots, F_m . We denote the corresponding test statistics as $R_n^{(1)}$ and $R_n^{(2)}$, respectively.

Theorem 1 implies that, under the null model of (4.1), $R_n^{(1)}$ and $R_n^{(2)}$ have the same χ_q^2 distribution in the limit. To compare their asymptotic powers, one can carry out simulations, or assess their asymptotic powers at local alternatives.

Theorem 2 provides a useful tool for the latter approach: $R_n^{(1)}$ and $R_n^{(2)}$ have non-central chi-square limiting distributions with the same q degrees of freedom, but with possibly different non-centrality parameter values at a local alternative. Our next result shows that $R_n^{(2)}$ always has a greater non-centrality parameter than $R_n^{(1)}$. Its proof is given in the supplementary material.

Theorem 3. *Under the conditions of Theorem 1, consider the composite hypothesis (4.1) and the local alternative*

$$\beta_k = \begin{cases} \beta_k^* + n_k^{-1/2} \mathbf{c}_k, & \text{for } k = 1, \dots, r \\ \beta_k^*, & \text{for } k = r + 1, \dots, m \end{cases} \tag{4.2}$$

with some given constants $\{\mathbf{c}_k\}$. If δ_1^2 and δ_2^2 are the non-centrality parameters of the limiting distributions of $R_n^{(1)}$ and $R_n^{(2)}$ under the local alternative model, then $\delta_2^2 \geq \delta_1^2$.

A standard result of Johnson, Kotz and Balakrishnan (1995, (29.25a)) has that if two non-central chi-square distributions have the same degrees of freedom, then the one with the greater non-central parameter stochastically dominates the other. Therefore, by Theorem 3, the local power of $R_n^{(2)}$ is greater than that of $R_n^{(1)}$ at all significance levels.

Example 3 (Effect of information pooling by DRM on the local power of the DELR test). Suppose $m + 1 = 3$, samples are from a DRM with basis function $\mathbf{q}(x) = (x, x^2)^\top$, and the sample proportions are (0.5, 0.25, 0.25). Let

F_k , $k = 1, 2$, be the distributions with parameters $\beta_1^* = (6, -1.5)^\top$ and $\beta_2^* = (-0.25, 0.375)^\top$. Suppose H_0 is given by $\mathbf{g}(\zeta) = \beta_1 - (6, -1.5)^\top = \mathbf{0}$, and the local alternative is $\beta_1 = \beta_1^* + n_1^{-1/2} \mathbf{c}_1$ and $\beta_2 = \beta_2^*$ with $\mathbf{c}_1 = (2, 2)^\top$.

Let $R_n^{(1)}$ and $R_n^{(2)}$ be the DELR test statistics based on F_0, F_1 , and on F_0, F_1, F_2 , respectively. When F_0 is the standard normal, we obtain information matrices (3.7), and hence $\Lambda = U_{\beta\beta} - U_{\beta\alpha} U_{\alpha\alpha}^{-1} U_{\alpha\beta}$, for $R_n^{(1)}$ and $R_n^{(2)}$ based on numerical computation. For $R_n^{(1)}$, $\boldsymbol{\eta} = (4, 4)^\top$ and $q = d = 2$, so by Theorem 2, $\delta_1^2 = \boldsymbol{\eta}^\top \Lambda \boldsymbol{\eta} \approx 5.90$. For $R_n^{(2)}$, $\boldsymbol{\eta} = (4, 4, 0, 0)^\top$, $\nabla = (I_2, 0_{2 \times 2})$, and $J = (0_{2 \times 2}, I_2)^\top$. And we get $\delta_2^2 \approx 6.67$. Since $\delta_1^2 < \delta_2^2$, $R_n^{(2)}$ is more powerful than $R_n^{(1)}$ at all significance levels. At the 5% significance level, for example, the powers of $R_n^{(1)}$ and $R_n^{(2)}$ are approximately 0.577 and 0.633, respectively.

5. Simulation Studies

We conducted simulations to study: (1) the approximation accuracy of the limiting distributions to the finite-sample distributions of the DELR statistic under both the null and the alternative models, (2) the power of the DELR test under correctly specified and misspecified DRMs, and (3) the effect of the number of samples used in the DRM to the local asymptotic power of the DELR test. The number of simulation runs was set to 10,000. Our simulations were more extensive, so we selected the most representative ones to include here; the other results are similar. All computations were carried out by our R package *drmdel* for EL inference under DRMs, which is available on the Comprehensive R Archive Network (CRAN).

5.1. Approximation to the distribution of the DELR under the null model

We studied how well the chi-square distribution approximates the finite-sample distribution of the DELR statistic under the null hypothesis of (3.4). With $m + 1 = 6$ and $\mathbf{g}(\boldsymbol{\beta}) = (\beta_1^\top, \beta_3^\top) - (\beta_2^\top, \beta_4^\top)$, the null hypothesis is equivalent to $F_1 = F_2$ and $F_3 = F_4$. We generated two sets of six samples of sizes (90, 60, 120, 80, 110, 30) from two distribution families. The first set were from normal distributions with means (0, 2, 2, 1, 1, 3.2) and standard deviations (1, 1.5, 1.5, 3, 3, 2). The second set of samples were from gamma distributions with shapes (3, 4, 4, 5, 5, 3.2) and rates (0.5, 0.8, 0.8, 1.1, 1.1, 1.5).

When the basis function $\mathbf{q}(x)$ is correctly specified, with $\mathbf{q}(x) = (x, x^2)^\top$ for the normal family and $\mathbf{q}(x) = (\log x, x)^\top$ for gamma family, the DELR statistic, R_n , has a χ_4^2 null limiting distribution. The quantile-quantile (Q-Q) plots of the distribution of R_n and χ_4^2 are shown in Figure 1. In both cases, the approximations are accurate. The type I error rates of R_n at 5% level are 0.056 and 0.058 for normal and gamma data respectively.

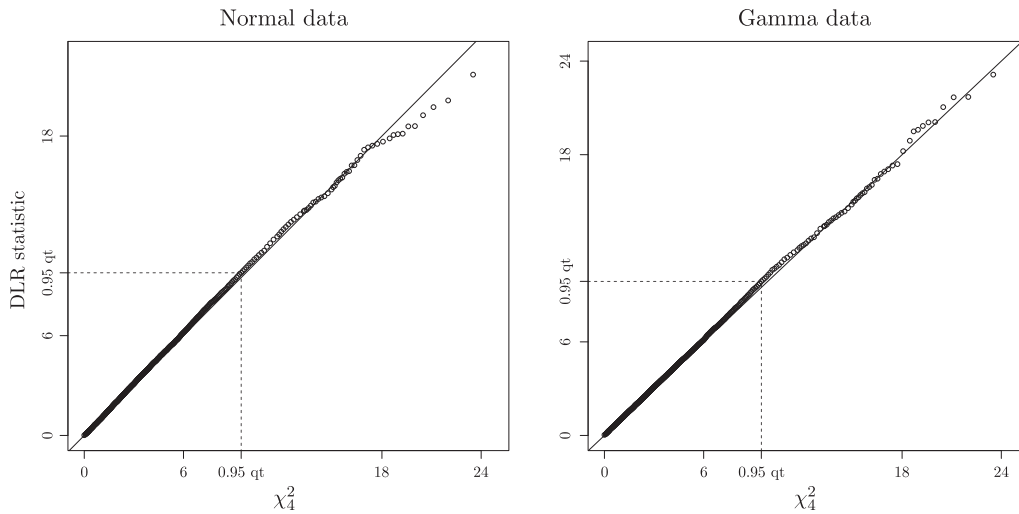


Figure 1. Q-Q plots of the simulated and the null limiting distributions of the DELR statistic.

In other simulations under various settings, we found more generally that the chi-square approximation has satisfactory precision when $n_k \geq 10qd$. When n_k is much smaller, a bootstrap or permutation test based on the DELR statistic can serve as an alternative.

5.2. Approximation to the distribution of the DELR under local alternatives

We examined the precision of the non-central chi-square distribution under the local alternative model (3.5), with $m + 1 = 4$ and sample sizes 120, 160, 80 and 60.

We first tested the hypothesis (3.4) with $\mathbf{g}(\boldsymbol{\beta}) = \boldsymbol{\beta}_1^\top - \boldsymbol{\beta}_2^\top$. The perceived null model is specified by $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_2^* = (0.25, 1.875)^\top$, $\boldsymbol{\beta}_3^* = (0.125, 1.97)^\top$ with basis function $\mathbf{q}(x) = (x, x^2)^\top$. The data were generated from $G_0 = N(0, 0.5^2)$, G_1 and G_3 with $\boldsymbol{\beta}_1^*$ and $\boldsymbol{\beta}_3^*$ respectively, and G_2 with $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^* + n_2^{-1/2}(1, 0)^\top$. According to Theorem 2, the limiting distribution of R_n is $\chi_2^2(2.67)$.

We also tested (3.4) with $\mathbf{g}(\boldsymbol{\beta}) = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_3^\top) - (\boldsymbol{\beta}_2^\top, (-6, 9)^\top)$. The perceived null model is specified by $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_2^* = (-4, 5)^\top$, $\boldsymbol{\beta}_3^* = (-6, 9)^\top$ with basis function $\mathbf{q}(x) = (\log x, x)^\top$. We generated data from $G_0 = \Gamma(3, 2)$ and $G_k, k = 1, 2, 3$, specified by (3.5) with $\mathbf{c}_1 = (0.5, 0.5)^\top$, $\mathbf{c}_2 = (1, 1)^\top$ and $\mathbf{c}_3 = (2, 2)^\top$. According to Theorem 2, the limiting distribution of R_n is $\chi_4^2(1.80)$.

The Q-Q plots under the two settings are shown in Figure 2. The non-central chi-square limiting distributions approximate those of R_n well. In other simulations, we generally found the approximation of the non-central chi-square to be satisfactory when $n_k \geq 15qd$.

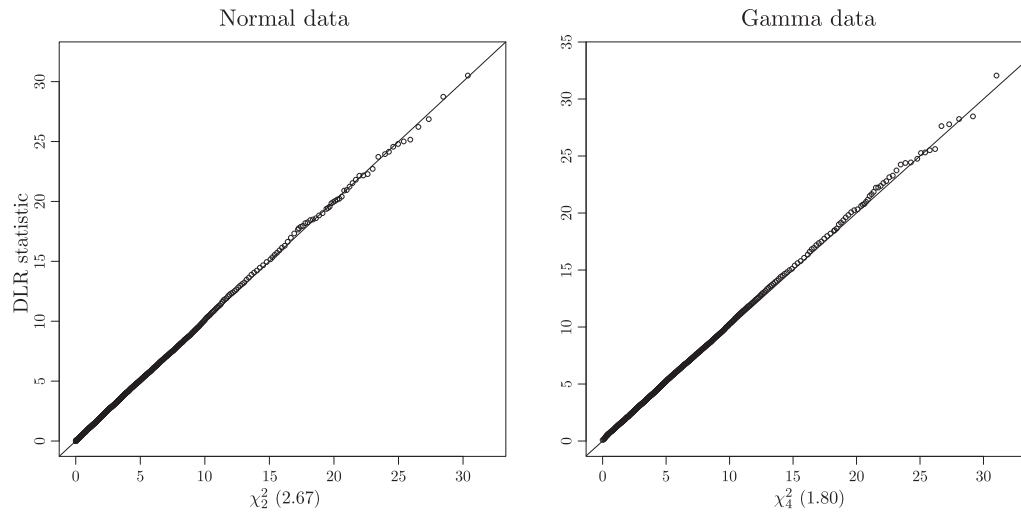


Figure 2. Q-Q plots of the distributions of the DELR statistics under the local alternative model against the corresponding asymptotic theoretical distributions.

5.3. Power comparison

We compared the power of the DELR test with a number of popular methods for detecting differences between distribution functions, testing $H_0 : F_0 = F_1 = \dots = F_m$. This is (3.4) with $\mathbf{g}(\boldsymbol{\beta}) = \boldsymbol{\beta}$. We used the nominal level of 5%.

Competitors were the Wald test based on the DRM (Wald) (Fokianos et al. (2001)), a one-way analysis of variance (ANOVA), the Kruskal–Wallis rank-sum test (KW) (Wilcox (1995)), the k-sample Anderson–Darling test (AD) (Scholz and Stephens (1987)), and the likelihood ratio test based on the partial likelihood under the CoxPH when observations are intrinsically positive. Under the CoxPH, we utilized m dummy covariates for data analysis. The corresponding likelihood ratio based on the partial likelihood has a χ_m^2 limiting distribution under the null hypothesis.

We compared powers based on normal data with $m + 1 = 2$ and sample sizes $n_0 = 30$ and $n_1 = 40$. We considered two scenarios with alternatives having $F_0 = N(0, 2^2)$. In the first, $F_1 = N(\mu, 2^2)$ with increasing μ . In the second, we considered seven parameter settings 0–6 for $F_1 = N(\mu, \sigma^2)$ with μ and σ taking values in $(0, 0.05, 0.1, 0.15, 0.25, 0.36, 0.55)$ and $(2, 1.9, 1.8, 1.7, 1.62, 1.56, 1.50)$, respectively.

The power curves are shown in Figure 3. In the two-sample case, all tests are found to have comparable powers. In the unequal variance case, the DELR test has much higher power than its competitors, and its type I error rate is close to the nominal 0.05.

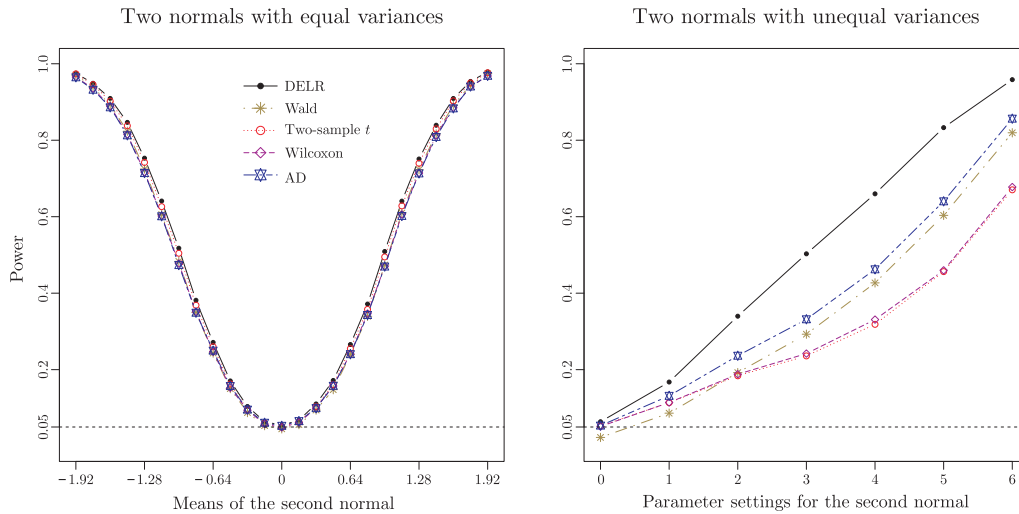


Figure 3. Power curve of R_n under normal data; the parameter setting 0 corresponds to the null model and settings 1–6 correspond to alternative models.

We also compared these tests on non-normal samples with $m + 1 = 5$ and sample sizes to be 30, 40, 25, 45 and 50. We generated data from the gamma, log-normal, Pareto with common support, and Weibull distributions with shape parameter equaling 0.8. The log-normal, Pareto and Weibull distributions are DRMs with basis functions $\mathbf{q}(x) = (\log x, \log^2 x)^\top$, $\mathbf{q}(x) = \log x$, and $\mathbf{q}(x) = x^{0.8}$, respectively.

For each distribution family, we obtain simulated power under six parameter settings 0–5 shown in Table 2 in the Appendix. Setting 0 satisfies the null hypothesis and settings 1–5 do not. The simulated rejection rates are shown in Figure 4. The DELR test has the highest power while its type I error rates are close to the nominal.

The gamma and log-normal families do not satisfy the conditions needed to justify use of the CoxPH approach. Consequently, the DELR test based on the DRM has a much higher power than the likelihood ratio test based on the partial likelihood under the CoxPH model. In contrast, the Pareto or Weibull families with known, common shapes do satisfy the CoxPH requirements; in these cases, the two tests have almost the same power. These results show that in general the DRM is a better choice for multiple samples.

5.4. DELR test under misspecified DRM

Examining the effect of misspecification is an important topic. Fokianos and Kaimi (2006) suggested that misspecifying the basis function $\mathbf{q}(x)$ has an

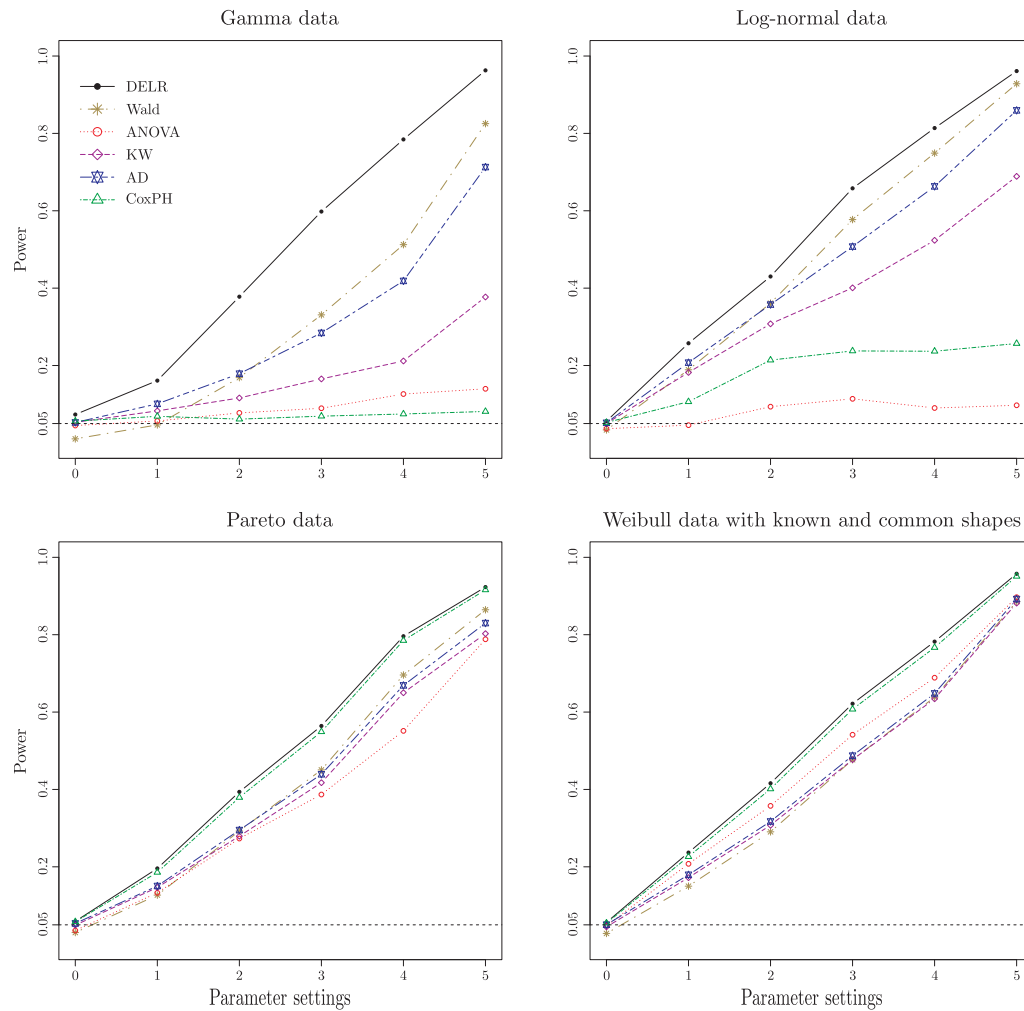


Figure 4. Power curves that obtain when the population distribution from which the data are sampled is non-normal; the parameter setting 0 corresponds to the null model and settings 1–5 correspond to alternative models.

adverse effect on estimating β . Chen and Liu (2013) found that estimation of population quantiles is robust against misspecification. In this section, we demonstrate that the effect of misspecification on DELR test is small for testing the equal population hypothesis.

We put $m + 1 = 5$ with sample sizes 90, 120, 75, 135 and 150. In the first simulation, we generated data from two-parameter Weibull distributions with density

$$f(x; a, b) = \left(\frac{a}{b}\right) \left(\frac{x}{b}\right)^{a-1} \exp\left\{-\left(\frac{x}{b}\right)^a\right\}, \quad x \geq 0.$$

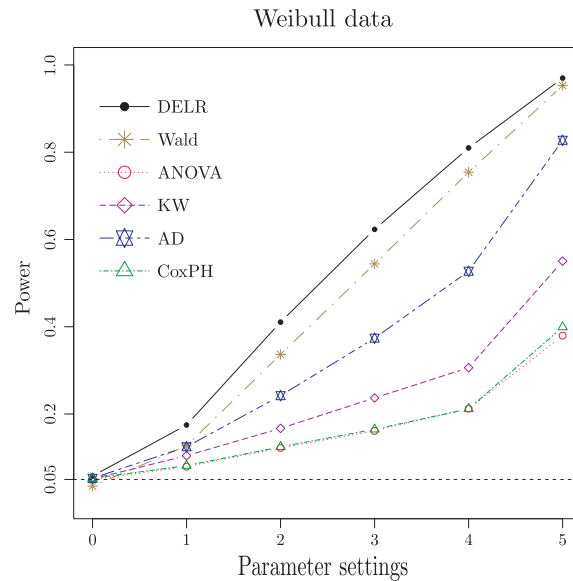


Figure 5. Power curves of five tests. Parameter setting 0 corresponds to the null model; settings 1–5 form alternative models.

The two-parameter Weibull family does not satisfy the DRM assumption (1.1). Nevertheless, we still fit a DRM with $\mathbf{q}(x) = (x, \log x)^T$ to the Weibull data. We used DELR test and Wald test under this DRM to test the equal distribution function hypothesis. We calculated the simulated power of these tests under six parameter settings (Table 3 in the Appendix) with setting 0 satisfying the null hypothesis.

We also applied ANOVA, KW, AD, and the CoxPH tests. The results are summarized as power curves in Figure 5. The DELR test has close to nominal type I error rates. It has superior power in detecting distributional differences. In particular, our DELR approach has a much higher power than the CoxPH.

We also experimented with other models. The results were similar.

5.5. Comparison of $R_n^{(1)}$ and $R_n^{(2)}$

It is helpful to have data from other populations in DELR analysis by Theorem 3. Here, we reaffirm this by means of simulations. We considered the same question about the Wald tests, $Wald^{(1)}$ and $Wald^{(2)}$, and included this in our simulations. The number of simulation repetitions was set to 10,000.

The first simulation used data from $m + 1 = 3$ normal populations with the null hypothesis $\beta_1 = (6, -1.5)^T$. The total sample size n was 240. We calculated the powers of $R_n^{(1)}$, $R_n^{(2)}$, $Wald^{(1)}$, and $Wald^{(2)}$ with the six DRM parameters as shown in the Appendix as the “Normal Case” in Table 4. The simulated power

curves are in Figure 6 (a). The power comparisons between $R_n^{(1)}$ and $R_n^{(2)}$ yield conclusions akin to that of Theorem 3. The Wald tests are not as powerful as the DELR tests, but $\text{Wald}^{(2)}$ seems to be more powerful than $\text{Wald}^{(1)}$.

Even if additional samples are from distributions not under comparison, they may well be helpful in estimating the baseline distribution F_0 , and, in turn, help to better identify the differences among the distributions under comparison. Let $m + 1 = 4$ and consider a hypothesis test for β_1 . The DELR test can be done using the first two samples ($R_n^{(1)}$) and then using all four samples ($R_n^{(2)}$).

We generated samples with sizes 60, 30, 40, and 90 from gamma distributions under two scenarios. In the first, the extra populations F_2 and F_3 are close to F_0 . Because of this, the samples from F_2 and F_3 are particularly helpful in accurately estimating F_0 . In the second scenario, F_2 and F_3 are rather distinct from F_0 and are less helpful at estimating F_0 . The density functions of F_0 , F_2 , and F_3 , along with their parameter values under both scenarios, are depicted in Figure 6 (b).

Under both scenarios, we considered the same null hypothesis $\beta_1 = (-2, 2)^\top$. We simulated the powers of the tests at six different values of β_1 (the ‘‘Gamma Case’’ in Table 4 of the Appendix) and the simulated power curves are shown in Figure 6 (c) and (d). The amounts of improvement of $R_n^{(2)}$ under two scenarios match our intuition, and are also evident for the Wald test.

Effects of the length of the basis function on DELR tests

If the additional populations are not exponential tilts of F_0 with a specific basis function, we may use a long basis function in the DRM to approximate the actual density ratios. This may have an adverse effect on the power. We investigate this issue here.

Let $m + 1 = 4$ and consider the hypothesis test for β_1 as in the last simulation. We compared the tests based on the first two samples and the ones based on all four samples.

We took the same distribution and parameter settings for F_0 and F_1 as in the last simulation but set F_2 to be log-normal with mean 0 and standard deviation 1 on log scale and F_3 to be Weibull with shape 2 and scale 3. We considered the basis functions (1) $\mathbf{q}(x) = (\log x, x)^\top$, (2) $\mathbf{q}(x) = (\log x, \sqrt{x}, x)^\top$, (3) $\mathbf{q}(x) = (\log x, \sqrt{x}, x, x^2)^\top$, and (4) $\mathbf{q}(x) = (\log x, \sqrt{x}, x, x^{1.5}, x^2)^\top$.

The simulation results are shown in Figure 7, where the parameter setting 0 corresponds to the null model. We see that $R_n^{(2)}$ is more powerful than $R_n^{(1)}$ in all cases. With the simplest basis function $\mathbf{q}(x) = (\log x, x)^\top$, $R_n^{(2)}$ has the type I error rate of 0.0625, which notably exceeds the nominal size of 5%; the type I error rate improves significantly when the dimension of the basis function increases. Moreover, the powers of all four tests can be seen to decrease as the dimension of the basis function increases.

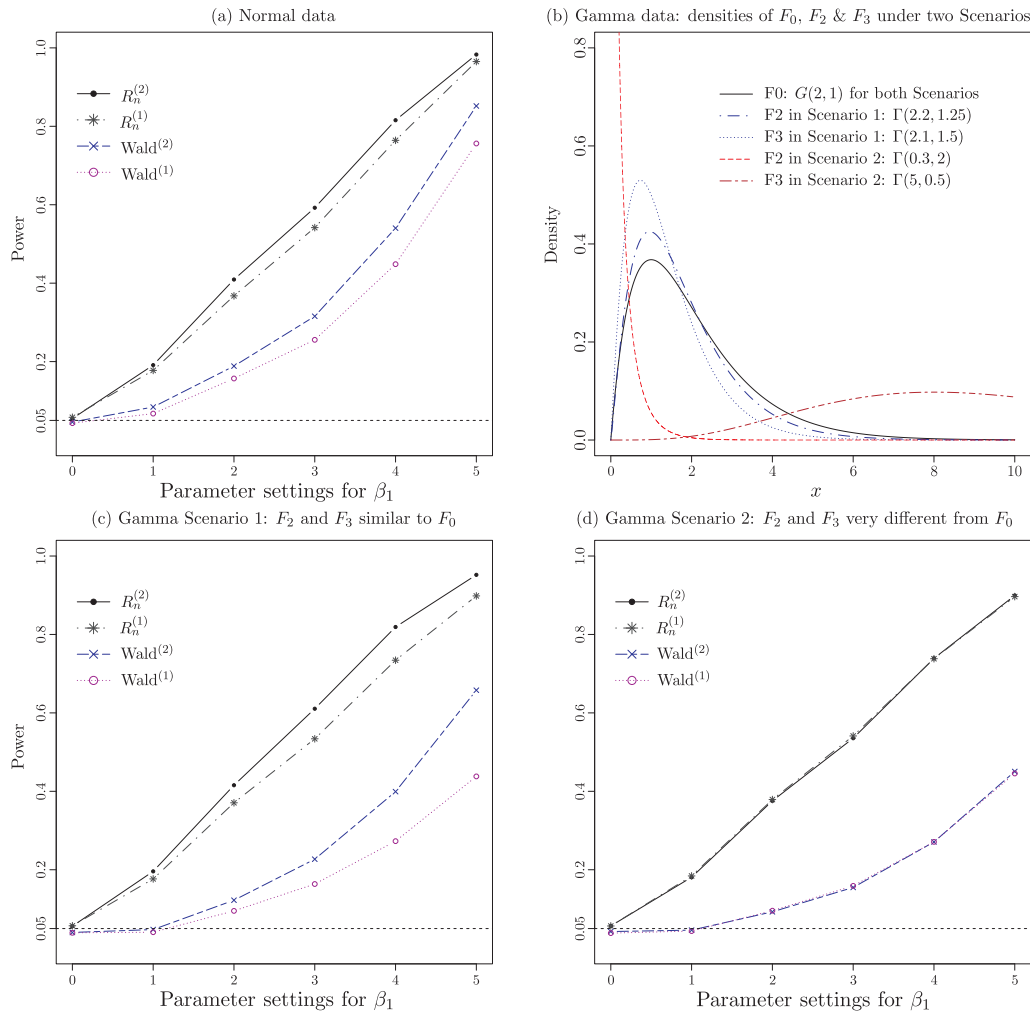


Figure 6. Power curves of $R_n^{(1)}$, $R_n^{(2)}$, Wald⁽¹⁾ and Wald⁽²⁾; Parameter setting 0 corresponds to the null model; settings 1–5 correspond to alternative models.

In this particular case, choosing a three dimensional basis function gives the best overall result: a reasonably accurate type I error rate and also a good power. The issue on how to choose basis function to achieve such a balance in general is rather delicate, and we will study it in the near future.

6. Analysis of Lumber Properties

The authors are members of the Forest Products Stochastic Modeling Group centered at the University of British Columbia and, in that capacity, are helping develop methods for assessing the engineering strength properties of lumber.

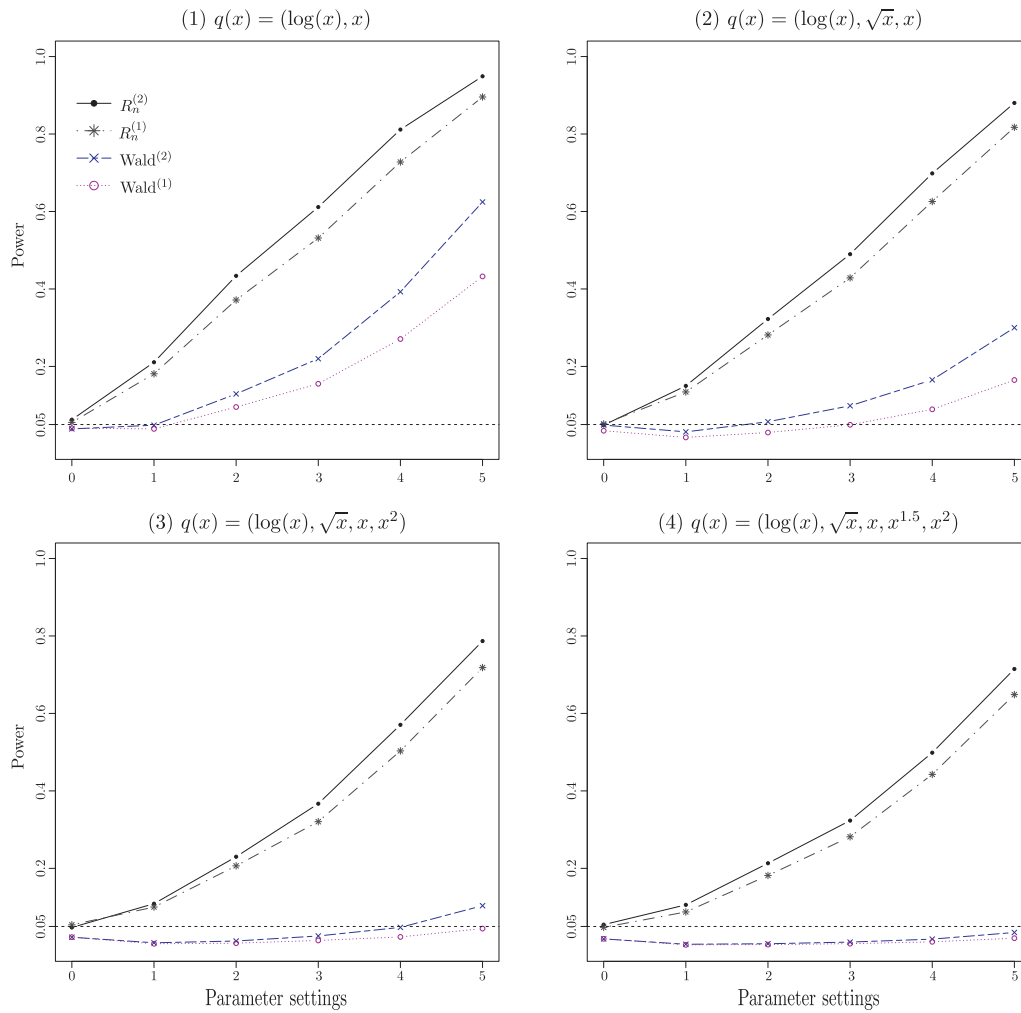


Figure 7. Power curves of $R_n^{(1)}$, $R_n^{(2)}$, Wald⁽¹⁾ and Wald⁽²⁾ under DRMs with basis functions of different dimensions; Parameter setting 0 corresponds to the null model; settings 1–5 correspond to alternative models.

A primary goal is an effective but relatively inexpensive long-term monitoring program for the strength of lumber. Of primary importance is the so-called modulus of rupture (MOR) or “bending strength”, which is measured in units of 10^3 pound-force per square inch (psi). The Forest Products Stochastic Modeling Group collected three MOR samples in year 2007, 2010, and 2011 with sample sizes 98, 282, and 445, respectively. Our interest in change over time led us to test the hypothesis that the three samples come from the same lumber population.

We used basis function $\mathbf{q}(x) = (\log x, x, x^2)^\top$ for the DRM, chosen according to the characteristics of the kernel density estimators of the MOR samples, shown

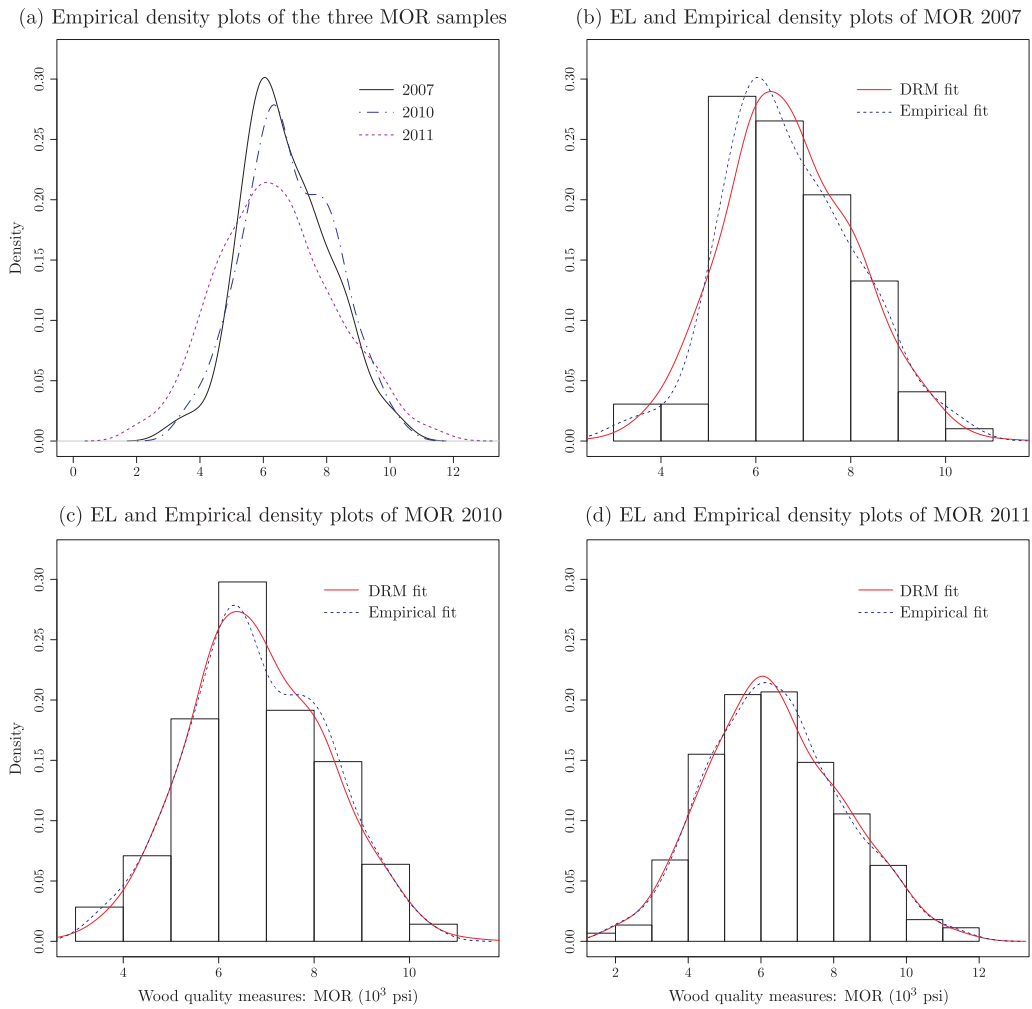


Figure 8. EL and empirical kernel density plots of the MOR samples.

in Figure 8 (a). They seem to be well approximated by either a Gamma or a normal distribution. Hence, we chose a basis function that includes both $(\log x, x)$ and (x, x^2) . To examine the adequacy of this basis function for fitting the MOR samples, we obtained EL kernel density estimates based on $\{x_{kj}\}$ with weights $\{\hat{p}_{kj}\}$ in addition to the usual kernel density estimates. These density estimates along with histograms of the MOR samples are shown in Figure 8 (b)–(d). We see that the EL kernel density estimates based on the DRM (the DRM fits) agree reasonably with the usual kernel density estimators (the Empirical fits) and the histograms.

With F_{07} , F_{10} and F_{11} distributions for 2007, 2010, and 2011, the p-values obtained using the DELR test, Wald test, ANOVA and Kruskal–Wallis tests for

Table 1. The p-values of pairwise comparisons among three MOR populations.

	DELR	Wald	t-test	KW
$H_0: F_{07}=F_{10}$	0.871	0.875	0.516	0.431
$H_0: F_{07}=F_{11}$	5.40e-4	7.01e-3	0.0579	0.0604
$H_0: F_{10}=F_{11}$	4.54e-8	1.82e-6	6.09e-4	3.95e-4

$H_0 : F_{07} = F_{10} = F_{11}$ were respectively 3.05e-8, 2.04e-6, 2.90e-3 and 1.08e-3. The DRM-based tests, especially the DELR test, had much smaller p-values.

Given rejection of that hypothesis it is natural to look at pairwise comparisons. The p-values for them are given in Table 1. The two DRM-based tests strongly suggest F_{11} is markedly different from F_{07} and F_{10} , while F_{07} and F_{10} are not significantly different. The other two tests arrive at the same conclusion, but without statistical significance at 5% level. The conclusion does not change at the 5% level when a Bonferroni correction is applied to account for the multiple comparison.

In addition, if the 5% size is strictly observed, the t-test and the KW test would imply $F_{07} = F_{10}$ and $F_{07} = F_{11}$, but $F_{10} \neq F_{11}$. This is harder to interpret.

7. Concluding Remarks

Our work was motivated in developing a new long-term monitoring program for the North American lumber industry. The need for efficiency and hence small sample sizes led to our DRM approach where common information across samples is pooled to gain efficiency. The demonstration of the use of the method on three lumber samples, shows our method to give a more incisive assessment than competitors through paired comparisons of the populations.

Our R package *drmdel* for EL inference under DRMs, available on CRAN, can carry out all computations in this paper, and those in Chen and Liu (2013).

Supplementary Materials

The online supplementary material accompanying this paper presents detailed proofs of the theorems.

Acknowledgements

This work was sponsored in part by the FPInnovations and the National Sciences and Engineering Research Council of Canada (NSERC).

We would like to thank the Associate Editor and the two reviewers for their comments that helped us improve our presentation.

Appendix: Parameter Values in Simulation Studies

Table 2. Parameter values for power comparison under non-normal distributions (Section 5.3). F_0 remains unchanged across parameter settings 0–5. $\Gamma(\lambda, \kappa)$: gamma distribution with shape λ and rate κ ; $LN(\mu, \sigma)$: log-normal distribution with mean μ and standard deviation σ on log scale; $Pa(\gamma)$: Pareto distribution with shape γ and common support of $x > 1$; $W(b)$: Weibull distribution with scale b and common shape of 0.8.

F_0	Parameter settings										
	1		2		3		4		5		
	λ	κ	λ	κ	λ	κ	λ	κ	λ	κ	
$\Gamma(0.2, 0.8)$	F_1 :	0.18	0.7	0.17	0.6	0.16	0.5	0.155	0.45	0.14	0.4
	F_2 :	0.22	0.85	0.24	0.95	0.255	1.05	0.18	0.7	0.17	0.6
	F_3 :	0.23	0.95	0.255	1.2	0.275	1.25	0.29	1.4	0.33	1.6
	F_4 :	0.24	1.05	0.27	1.3	0.29	1.4	0.31	1.55	0.35	1.85
$LN(0, 1.5)$		μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
	F_1 :	0.44	1.3	0.7	1.2	0.9	1.15	1	1	1.2	0.85
	F_2 :	0.22	1.32	0.57	1.30	0.62	1.25	0.67	1.20	0.87	1
	F_3 :	0.18	1.35	0.63	1.33	0.73	1.30	0.83	1.28	0.85	1.28
$Pa(2)$		γ		γ		γ		γ		α	
	F_1 :	1.9		1.85		1.8		1.75		1.7	
	F_2 :	2.1		2.2		2.3		1.85		1.75	
	F_3 :	2.35		2.55		2.70		2.85		3.25	
$W(1)$		b		b		b		b		b	
	F_1 :	0.76		0.65		0.59		0.53		0.42	
	F_2 :	1.2		1.26		1.31		1.35		1.42	
	F_3 :	1.08		1.05		1.10		1.12		1.14	
F_4 :	0.90		0.89		0.85		0.82		0.78		

Table 3. Parameter values for power comparison under misspecified DRMs (Section 5.4). F_0 remains unchanged across parameter settings 0–5. $W(a, b)$: Weibull distribution with shape a and scale b .

F_0	Parameter settings									
	1		2		3		4		5	
	a	b	a	b	a	b	a	b	a	b
$W(1, 1)$	F_1 : 0.9	0.95	0.85	0.94	0.82	0.92	0.79	0.91	0.75	0.88
	F_2 : 0.98	0.98	0.96	0.96	0.95	0.95	0.94	0.94	0.91	0.92
	F_3 : 1.03	1.04	1.05	1.06	1.07	1.07	1.09	1.08	1.12	1.12
	F_4 : 1.01	0.95	1.02	0.92	1.03	0.90	1.05	0.89	1.07	0.85

Table 4. Parameter settings for power comparison of $R_n^{(1)}$ and $R_n^{(2)}$ (Section 5.5).

Common parameter settings: $F_0 : N(0, 1), F_2 : N(-1, 2)$						
Normal Case	Parameter settings for F_1					
	0	1	2	3	4	5
	$N(1.5, 0.5)$	$N(1.57, 0.45)$	$N(1.58, 0.41)$	$N(1.6, 0.39)$	$N(1.62, 0.36)$	$N(1.64, 0.31)$
Common parameter settings: $F_0 : \Gamma(2, 1)$						
Gamma Case	Parameter settings for F_1					
	0	1	2	3	4	5
	$\Gamma(4, 3)$	$\Gamma(5.3, 4.3)$	$\Gamma(6.3, 5.3)$	$\Gamma(7.1, 6.1)$	$\Gamma(8.3, 7.3)$	$\Gamma(10, 9)$

References

- ASTM (2007). *ASTM D1990 - 07: Standard Practice for Establishing Allowable Properties for Visually-Graded Dimension Lumber from In-Grade Tests of Full-Size Specimens*. West Conshohocken, USA: ASTM International.
- Chen, J. and Liu, Y. (2013). Quantile and quantile-function estimations under density ratio model. *Ann. Statist.* **41**, 1669-1692.
- Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.
- Fokianos, K. (2004). Merging information for semiparametric density estimation. *J. Roy. Statist. Soc. Ser. B* **66**, 941-958.
- Fokianos, K. and Kaimi, I. (2006). On the effect of misspecifying the density ratio model. *Ann. Inst. Statist. Math.* **58**, 475-497.
- Fokianos, K., Kedem, B., Qin, J. and Short, D. A. (2001). A semiparametric approach to the one-way layout. *Technometrics* **43**, 56-65.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*. Volume 2, 2nd edition. Wiley-Interscience, Hoboken.
- Keziou, A. and S. Leoni-Aubin (2008). On empirical likelihood for semiparametric two-sample density ratio models. *J. Statist. Plann. Inference* **138**, 915-928.
- Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall, New York.

- Qin, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika* **85**, 619-630.
- Qin, J. and B. Zhang (1997). A goodness of fit test for the logistic regression model based on case-control data. *Biometrika* **84**, 609-618.
- Scholz, F. W. and M. A. Stephens (1987). K-sample Anderson-Darling tests. *J. Amer. Statist. Assoc.* **82**, 918-924.
- Wilcox, R. R. (1995). Anova: A paradigm for low power and misleading measures of effect size. *Rev. Educational Res.* **65**, 51-77.
- Zhang, B. (2000). Quantile estimation under a two-sample semi-parametric model. *Bernoulli* **6**, 491-511.
- Zou, F., Fine, J. P. and Yandell, B. S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika* **89**, 61-75.

School of Mathematics and Statistics, Carleton University, 4302 Herzberg Laboratories, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6, Canada.

E-mail: scai@math.carleton.ca

Department of Statistics, University of British Columbia, 3182 Earth Sciences Building, 2207 Main Mall, Vancouver, BC, V6T 1Z4, Canada.

E-mail: jhchen@stat.ubc.ca

Department of Statistics, University of British Columbia, 3182 Earth Sciences Building, 2207 Main Mall, Vancouver, BC, V6T 1Z4, Canada.

E-mail: jim@stat.ubc.ca

(Received May 2014; accepted April 2016)

