Accepted Manuscript

Semiparametric inference on the means of multiple nonnegative distributions with excess zero observations

Chunlin Wang, Paul Marriott, Pengfei Li

PII: S0047-259X(16)30233-0
DOI: https://doi.org/10.1016/j.jmva.2018.02.010
Reference: YJMVA 4332

To appear in: *Journal of Multivariate Analysis*

Received date : 12 December 2016

Please cite this article as: C. Wang, P. Marriott, P. Li, Semiparametric inference on the means of multiple nonnegative distributions with excess zero observations, *Journal of Multivariate Analysis* (2018), https://doi.org/10.1016/j.jmva.2018.02.010

Semiparametric inference on the means of multiple nonnegative distributions with excess zero observations

Chunlin Wang[a], Paul Marriott[b], Pengfei Li[b,*]

[a]Department of Statistics, School of Economics and Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, 361005, China
[b]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Abstract

A non-standard, but not uncommon, situation is to observe multiple samples of nonnegative data which have a high proportion of zeros. This is the so-called excess of zeros situation and this paper looks at the problem of making inferences about the means of the underlying distributions. Under the semiparametric setup, proposed by Wang et al. [31], we develop a unified inference framework, based on an empirical likelihood ratio (ELR) statistic, for making inferences on the means of multiple such distributions. A chi-square-type limiting distribution of this statistic is established under a general linear null hypothesis about the means. This result allows us to construct a new test for mean equality. Simulation results show favorable performance of the proposed ELR when compared with other existing methods for testing mean equality, especially when the correctly specified basis function in the density ratio model is the logarithm function. A real data set is analyzed to illustrate the advantages of the proposed method.

Keywords: Density ratio model, Empirical likelihood, Estimating equation, Multinomial logistic regression, Non-standard mixture model, Semi-continuous data
2010 MSC: 62H15, 62H10, 62E20

1. Introduction

Making reliable inferences on the means of multiple distributions is an important, and fundamental, topic in statistics. In this paper, we investigate this topic in the case of multiple skewed nonnegative distributions with an excess of zero values. Specifically, suppose we have $m+1$ independent samples modeled as

$$\forall_{i\in\{0,\ldots,m\}} \quad x_{i1},\ldots,x_{in_i} \sim F_i(x) = \nu_i \mathbf{1}(x=0) + (1-\nu_i)\mathbf{1}(x>0)G_i(x), \tag{1}$$

where n_i is the sample size of the ith group, $\mathbf{1}$ is an indicator function and the G_is are cumulative distribution functions with common support which may be continuous or discrete. Under the formulation (1), the mean of each mixture distribution F_i, with $i \in \{0,\ldots,m\}$, can be expressed as

$$\mu_i = \int_0^\infty x dF_i(x) = (1-\nu_i)\int_0^\infty x dG_i(x).$$

Our interest is to make inferences about $\mu_0,\ldots,\mu_m$. These include testing the null hypothesis $\mathcal{H}_0^*: \mu_0 = \cdots = \mu_m$, and constructing confidence intervals for $\mu_i - \mu_j$ and μ_i/μ_j, for $i \neq j$.

Multiple samples, with a non-standard mixture structure as shown in (1), frequently arise from many research areas; see [31] and also a recent special issue of the *Biometrical Journal* [2] and references therein. The mean of a population with excess zeros has been considered an important summary quantity. For example, in fishery and health economics studies, the population total often has a crucial scientific meaning. It can provide information for

*Corresponding author.
Email address: pengfei.li@uwaterloo.ca (Pengfei Li)

recovering the population total, e.g., the total egg production of Atlantic mackerel [20], and the total expenditure of patients [7].

A natural way to make inference on the means is by using fully parametric models. The papers [27] and [34] proposed modeling the G_is by a log-normal distribution, based on which they developed a Wald and a likelihood ratio test, for testing the overall mean equality. Several confidence intervals for the two-sample mean difference and mean ratio have been considered in [7] and [35], when the positive data in both samples follow log-normal distributions. Although the log-normal distributions are quite natural for modeling the positive observations, other parametric models, such as the Gamma distribution, have also been argued to be suitable in applications [16, 17]. However, as concluded in [17], "when sample sizes are not large, different parametric models that fit the data equally well can lead to substantially different inferences". This fact may pose an issue of model robustness in the fully parametric approach.

Another approach is to use nonparametric methods. The nonparametric ANOVA-type statistic (ATS) and the Wald-type permutation statistic (WTPS) are two representative methods for comparing the means of multiple non-normal and heteroscedastic samples. The ATS, proposed in [3], is an extension of the classical ANOVA F-test for heteroscedastic factorial designs. Brunner et al. [3] suggested using an F-distribution with random degrees of freedom to approximate the finite-sample distribution of the ATS. It can be shown that the ATS is equivalent to the Welch two-sample t-test [32] when $m = 1$. More recently, the WTPS was proposed in [19] for testing a linear hypothesis about the means which does not make any distributional assumptions, and is appropriate under very general heteroscedastic factorial designs. In practice, the homoscedastic variance assumption is usually difficult to justify for multiple groups of observations with excess zeros; see, e.g., [34] and Section 4 of this paper. Hence, it is appropriate to directly apply the ATS and WTPS methods. The empirical likelihood method has also received considerable interest in dealing with such problems. Chen et al. [5] and Chen and Qin [6] used it to construct the confidence interval for the mean of a population with excess zeros. For the two-sample case, Taylor and Pollard [25] considered a test for the means, and Kang et al. [13] and Wu and Yan [33] studied the construction of confidence intervals for the mean difference, by using the empirical likelihood.

To borrow information across similar populations to improve the inferential results, Wang et al. [31] proposed a semiparametric setup for modeling the distributions F_is in (1), in which G_is are linked by the semiparametric *density ratio model* (DRM) of Anderson [1]. Under this framework, Wang et al. [31] proposed a procedure to test for homogeneity in distributions, i.e., $F_0 = \cdots = F_m$. They demonstrate that the proposed procedure is robust to changes of underlying distributions, is competitive to, and sometimes more powerful than, the existing methods. In this paper, we study the inference procedures for the means under the semiparametric framework of [31]. For the convenience of presentation, we concentrate on continuous distributions G_is whose support consists of all nonnegative real numbers. The proposed inference method can be similarly applied to discrete distributions G_is for count data, as discussed in [31].

Let $dG_i(x)$ denote the density of $G_i(x)$, for $i \in \{0, \ldots, m\}$. The DRM postulates that, for each $i \in \{0, \ldots, m\}$,

$$dG_i(x) = \exp\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x)\} dG_0(x), \tag{2}$$

where $\mathbf{q}(x)$ is a non-trivial, pre-specified, basis function of dimension d, and α_is and $\boldsymbol{\beta}_i$s are unknown parameters. Obviously, $\alpha_0 = 0$ and $\boldsymbol{\beta}_0 = \mathbf{0}_{d\times 1}$ for an arbitrarily selected baseline group indexed, without loss, by 0. Here and after, we use bold notation for a vector or matrix. Under (1) and (2), we develop a unified framework, based on an empirical likelihood ratio (ELR) statistic, for making inferences on the means of multiple distributions, without having to fully specify their distributions. We show that the ELR statistic has a simple χ^2-type limiting distribution under a general linear null hypothesis about the means. This result allows us to construct a new test for mean equality. Software implementing the proposed ELR for testing overall mean equality, with basis function $\mathbf{q}(x) = \ln(x)$ in the DRM, has been developed in R language [24], and is available in [30].

We note that testing the general linear null hypothesis about the means, considered in (3) of Section 2, is essentially different from testing homogeneity in distributions, i.e., $F_0 = \cdots = F_m$, considered in [31]. Note that the full models are the same for these two hypothesis testing problems, but the null models are distinct. Hence, different testing procedures are required in these two testing problems, although both testing procedures are developed by using the empirical likelihood ratio principle. The empirical likelihood under the null hypothesis that $F_0 = \cdots = F_m$ has a simple form as shown in [31]. However, as demonstrated in Section 2, the empirical likelihood under the null model considered here actually involves sample selection bias problem induced by the DRM in (2) together with estimating equations induced by the linear null hypothesis. Unlike existing works, e.g., [4, 22, 29, 31], we no longer have a simple analytical form for the profile empirical likelihood or dual empirical likelihood under the general linear null hypothesis. This makes the theoretical derivation more subtle and complicated; see [23] and our proofs in the Online Supplement. Additionally, due to the non-standard mixture structure (1), the summations in the definition of the ELR (see Section 2) are over random numbers, i.e., the number of positive observations in each group. Hence, standard large-sample techniques may not be directly applicable. Despite of these above challenges, we show that the ELR enjoys a simple χ^2-type limiting distribution.

The structure of this paper is as follows. In Section 2, we formulate the research problem, construct the empirical likelihood ratio statistic, and study its asymptotic properties. Simulation results are reported in Section 3, and a real data set is analyzed in Section 4. Some concluding remarks are provided in Section 5. For the convenience of presentation, numerical implementation of the proposed ELR is discussed in Appendix A, and the form of $\mathbf{U}$ matrix, needed in regularity conditions, is defined in Appendix B. The proofs are given in the Online Supplement.

2. Empirical likelihood inference under the DRM

2.1. Notation and problem setup

Let us first introduce some notation. Let n_{i0} and n_{i1} denote the (random) numbers of zero and positive observations for the ith group, for $i \in \{0, \dots, m\}$. Define $n_{\cdot 0} = n_{00} + \cdots + n_{m0}$ and $n_{\cdot 1} = n_{01} + \cdots + n_{m1}$ the total zero and nonzero sample sizes, and let $n = n_0 + \cdots + n_m$ denote the total sample size. Without loss of generality, we use the first n_{i1} observations $x_{i1}, \dots, x_{in_{i1}}$ to denote the positive observations in the ith group for $i \in \{0, \dots, m\}$.

Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_m)^\top$ be the mean vector of the $m + 1$ groups. The main goal of this section is to develop a test for the following general linear hypothesis about the means

$$\mathcal{H}_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{d}, \tag{3}$$

where the $p \times (m + 1)$ matrix $\mathbf{C}$ and $p \times 1$ vector $\mathbf{d}$ have real, non-random, entries that are completely specified under the null hypothesis and do not depend on sample sizes. We assume that $\mathbf{C}$ has full row rank so that $\text{rank}(\mathbf{C}) = p$, with $p \le m + 1$.

We comment that formulation (3) is very flexible and includes a variety of inference problems as special cases. For example, when

$$\mathbf{C} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \dots & 1 \end{pmatrix}_{m\times(m+1)}, \quad \mathbf{d} = \mathbf{0}_{m\times 1}, \tag{4}$$

then the hypothesis (3) becomes $\mathcal{H}_0^* : \mu_0 = \cdots = \mu_m$. In our numerical studies, in Section 3, we focus on this special, though important, hypothesis $\mathcal{H}_0^*$.

2.2. Empirical likelihood ratio

For a compact presentation, we use vector notation. Let $\boldsymbol{\nu} = (\nu_0, \dots, \nu_m)^\top$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_0^\top, \dots, \boldsymbol{\theta}_m^\top)^\top$ with $\boldsymbol{\theta}_i = (\alpha_i, \boldsymbol{\beta}_i^\top)^\top$ for $i \in \{0, \dots, m\}$. Further, let $\boldsymbol{\omega}(x; \boldsymbol{\theta}) = (\omega_1(x; \boldsymbol{\theta}_1), \dots, \omega_m(x; \boldsymbol{\theta}_m))^\top$ with $\omega_i(x; \boldsymbol{\theta}_i) = \exp\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x)\}$ for $i \in \{0, \dots, m\}$.

Under the DRM (2) for the G_is, we can refine the definition of the means based on the pooled positive samples

$$\forall_{i\in\{0,\dots,m\}} \quad \mu_i = (1 - \nu_i) \int_0^\infty x\omega_i(x; \boldsymbol{\theta}_i) dG_0(x).$$

Under the general null hypothesis $\mathcal{H}_0$ in (3), we have $\mathbf{C}\boldsymbol{\mu} - \mathbf{d} = \mathbf{0}_{p\times 1}$, or equivalently $\mathrm{E}_0\{\mathbf{g}(X; \boldsymbol{\nu}, \boldsymbol{\theta})\} = \mathbf{0}_{p\times 1}$, where $X \sim G_0(x)$ and E_0 means that the expectation is taken under $G_0(x)$, and

$$\mathbf{g}(x; \boldsymbol{\nu}, \boldsymbol{\theta}) = \begin{pmatrix} g_1(x; \boldsymbol{\nu}, \boldsymbol{\theta}) \\ \vdots \\ g_p(x; \boldsymbol{\nu}, \boldsymbol{\theta}) \end{pmatrix} = \mathbf{C} \begin{pmatrix} (1-\nu_0)x \\ (1-\nu_1)x\omega_1(x; \boldsymbol{\theta}_1) \\ \vdots \\ (1-\nu_m)x\omega_m(x; \boldsymbol{\theta}_m) \end{pmatrix} - \mathbf{d}. \tag{5}$$

Therefore, the information about the means in null hypothesis $\mathcal{H}_0$ can come in the form of an unbiased estimating equation (5). For parameters estimated through unbiased estimating equations, the empirical likelihood method has been shown to provide an effective inference platform [21]. Based on the construction of unbiased estimating equation (5), we proceed to develop inference procedures using the empirical likelihood method.

Along the lines of empirical likelihood [18], we restrict the form of baseline distribution G_0 to be

$$G_0(x) = \sum_{i=0}^{m} \sum_{j=1}^{n_{i1}} p_{ij} \mathbf{1}(x_{ij} \le x).$$

Given multiple groups of samples from (1) in which the G_is satisfy the DRM (2), the empirical log-likelihood function can be written as

$$\tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) = \sum_{i=0}^{m} \ln\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\} + \sum_{i=0}^{m} \sum_{j=1}^{n_{i1}} \{\alpha_i + \boldsymbol{\beta}_i^{\top} \mathbf{q}(x_{ij}) + \ln(p_{ij})\}.$$

We always have the following set of natural constraints:

$$C_1 = \left\{ (\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) : \nu_i \in (0,1), \quad p_{ij} > 0, \quad \sum_{i=0}^{m} \sum_{j=1}^{n_{i1}} p_{ij} = 1, \quad \sum_{i=0}^{m} \sum_{j=1}^{n_{i1}} p_{ij}\{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \boldsymbol{\iota}\} = \mathbf{0}_{m\times 1} \right\},$$

where $\boldsymbol{\iota}$ denotes a vector of 1s. Under the general null hypothesis $\mathcal{H}_0$ in (3), we also have the following set of constraints:

$$C_2 = \left\{ (\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) : \sum_{i=0}^{m} \sum_{j=1}^{n_{i1}} p_{ij} \mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta}) = \mathbf{0}_{p\times 1} \right\}.$$

The empirical likelihood ratio (ELR) statistic for testing the general null hypothesis given in (3) is then defined via

$$R_n = 2\left\{ \sup_{(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) \in C_1} \tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) - \sup_{(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) \in C_1 \cap C_2} \tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) \right\}. \tag{6}$$

For the convenience of presentation, the numerical evaluation of R_n is discussed in Appendix A.

2.3. Large-sample property

In this section, we study the asymptotic distribution of the ELR statistic, R_n, for the general hypothesis testing problem in (3) under (1) and (2).

Suppose that the true value of $(\boldsymbol{\nu}^{\top}, \boldsymbol{\theta}^{\top})^{\top}$ is $(\boldsymbol{\nu}^{*\top}, \boldsymbol{\theta}^{*\top})^{\top}$ under the null hypothesis $\mathcal{H}_0$. Deriving the asymptotic distribution of R_n relies on the following regularity conditions:

- R1. $\nu_i^* \in (0,1)$ for all $i \in \{0, \dots, m\}$.
- R2. $\lim_{\min(n_0,\dots,n_m)\to\infty} n_i/n \to \rho_i^*$, where $\rho_i^* \in (0,1)$ for all $i \in \{0, \dots, m\}$.
- R3. $\int (1, \mathbf{q}^{\top}(x))^{\top} (1, \mathbf{q}^{\top}(x))\, dG_i(x)$ exists and is positive definite for all $i \in \{0, \dots, m\}$.

R4. $\int \exp\{\boldsymbol{\beta}_i^\top \mathbf{q}(x)\} dG_i(x) < \infty$ in a neighborhood of $\boldsymbol{\beta}_i^*$, where $\boldsymbol{\beta}_i^*$ is the true value of $\boldsymbol{\beta}_i$ under the null hypothesis $\mathcal{H}_0$, for all $i \in \{0, \ldots, m\}$.

R5. The matrix $\mathbf{U}$ defined in (B.1), in Appendix B, is positive definite.

R6. $\left\|\partial \mathbf{g}(x; \boldsymbol{\nu}, \boldsymbol{\theta})/\partial \boldsymbol{\eta}\right\|$ and $\|\mathbf{g}(x; \boldsymbol{\nu}, \boldsymbol{\theta})\|^3$ are bounded by some integrable function of x with respect to $G_0(x)$ in a neighborhood of $(\boldsymbol{\nu}^{*\top}, \boldsymbol{\theta}^{*\top})^\top$, where $\boldsymbol{\eta} = (\boldsymbol{\nu}^\top, \boldsymbol{\theta}^\top)^\top$ and $\|\cdot\|$ denotes Euclidean norm.

Condition R1 states that the parameter $\boldsymbol{\nu}^*$ is an interior point of the parameter space of $\boldsymbol{\nu}$. Condition R2 assumes that the ratio of each group's sample size over n converges to a constant as $\min(n_0, \ldots, n_m) \to \infty$. For simplicity, and convenience of presentation, we write $\rho_i^* = n_i/n$ and assume that it is a constant. This does not affect our technical development. Under Conditions R1 and R2, there is no need to distinguish the stochastic orders with respect to n or n_i. Condition R3 is an identifiability condition, and it ensures that the components of $(1, \mathbf{q}^\top(x)\mathbf{1}(x > 0))$ are linearly independent under all $G_i(x)$s. Conditions R3–R6 guarantee that a quadratic approximation of R_n is applicable.

The following theorem defines the asymptotic null distribution of R_n under the general null hypothesis $\mathcal{H}_0$ in (3).

Theorem 1. *Suppose we have $m+1$ groups of samples of the form* (1) *and condition* (2) *is satisfied. Assume, also, that the regularity conditions R1–R6 hold. Under the null hypothesis $\mathcal{H}_0$, given in* (3)*, we have $R_n \rightsquigarrow \chi_p^2$ as $n \to \infty$, where χ_p^2 is a chi-squared random variable with p degrees of freedom, and $p = \mathrm{rank}(\mathbf{C})$ for some full rank $\mathbf{C}$ in $\mathcal{H}_0$.*

For convenience of presentation, proof of Theorem 1 is given in the Online Supplement. Here we make three remarks about Theorem 1.

(a) As a direct consequence of Theorem 1, the ELR test for the overall equality of $m+1$ group means, $\mathcal{H}_0^*$, has a limiting chi-squared distribution with m degrees of freedom, since the rank of $\mathbf{C}$ in (4) is m.

(b) The mean differences and ratios are two quantities commonly used to measure the magnitudes of relative differences among the group means. The result of Theorem 1 is also useful for the construction of confidence intervals, or regions, for the mean differences and ratios. As an illustration, suppose we are interested in constructing a confidence interval for the mean difference $\delta = \mu_1 - \mu_0$ in the two-sample problem. The unbiased estimating equation (5) can be replaced by

$$d(x; \boldsymbol{\nu}, \boldsymbol{\theta}, \delta) = (1 - \nu_1)x\omega_1(x; \boldsymbol{\theta}_1) - (1 - \nu_0)x - \delta.$$

Then, the ELR, defined in (6), becomes a function of δ, since this parameter δ is incorporated through $d(x; \boldsymbol{\nu}, \boldsymbol{\theta}, \delta)$ in the constraint set C_2. We denote it as $R_n(\delta)$. It follows that the 95% ELR confidence interval for δ can be constructed as $\{\delta : R_n(\delta) \le \chi^2_{1,0.95}\}$, where $\chi^2_{1,0.95}$ denotes the 95th quantile of the χ^2_1 distribution.

(c) A Wald-type statistic may also be constructed based on the normal approximation to $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})$ defined in Appendix A. However, such a statistic is not invariant to transformations [8]. For example, mean differences and mean ratios are two different nonlinear transformations of $(\boldsymbol{\nu}, \boldsymbol{\theta})$. The Wald-type statistics for testing mean ratios equal to one, and for testing mean differences equal to zero, could lead to two different conclusions.

3. Simulation studies

In this section, we use Monte Carlo simulations to evaluate the finite-sample performance of the proposed ELR test for testing the overall mean equality, that is an important case covered by the hypothesis testing problem (3). The ELR test statistic is calculated by the procedure discussed in Appendix A with the forms of $\mathbf{C}$ and $\mathbf{d}$ given in (4).

We fix the number of groups under comparison to be $m+1 = 2$ or $m+1 = 3$. We compare the type I error rates and the power of the proposed ELR test with the ATS of [3] and WTPS of [19]. Note that the classical ANOVA F-test is designed for the case that the variances of F_i are homogenous, which may not be satisfied in our setup. Recall that both the ATS and WTPS do not require such an assumption. Hence in our comparison, we only include the ATS and WTPS.

For each test, the type I error rate and the power at the 5% significance level are calculated based on 10,000 and 2000 repetitions, respectively. The computations for the ATS and WTPS methods use the R package "GFD" [10].

Following the suggestion in [19], we use 10,000 permutation samples (the default number in "`GFD`") to calculate the type I error rate and the power of the WTPS.

The random observations are generated, conditional on all $\hat{\nu}_i$s $\neq 0$ or 1, from (1) with all the G_is being log-normal, or all the G_is being Gamma distributions. Note that if any $\hat{\nu}_i$s equals 0 or 1, then some test statistics may not be well defined. This is not a problem in practice. In the following, we use $\mathcal{LN}(a_i, b_i)$ to denote a log-normal distribution with mean a_i and variance b_i both with respect to the log scale (i.e., mean and variance of the associated normal random variable), and $\mathcal{GAM}(a_i, b_i)$ to denote a Gamma distribution with shape parameter a_i and scale parameter b_i.

The parameter settings under the null hypothesis (LN_1–LN_6 and GAM_1–GAM_6) that all the means are equal, and the alternative hypothesis (LN_7–LN_{15} and GAM_7–GAM_{15}) are given in Table 1. Note that in the following we use the same model notation for two- and three-sample comparisons when no confusion is caused. We consider the case with equal sample sizes by setting (n_0, n_1) to be $(50, 50)$ and $(100, 100)$ for the two-sample comparison, and (n_0, n_1, n_2) to be $(50, 50, 50)$ and $(100, 100, 100)$ for the three-sample comparison. We also consider the case with unequal sample sizes by setting (n_0, n_1) to be $(50, 150)$ and $(150, 50)$ for the two-sample comparison, and (n_0, n_1, n_2) to be $(50, 150, 100)$ and $(150, 50, 100)$ for the three-sample comparison. With the parameter settings in Table 1, these combinations of unequal sample sizes correspond to two cases where increasing the sample sizes is related with increasing variances (positive pairing) or with decreasing variances (negative pairing).

To evaluate the performance of the ELR test with respect to the choices of user-specified basis function $\mathbf{q}(x)$ in a DRM, we consider the following three scenarios that may be encountered in practice:

Scenario I: All the distributions G_is are homogenous, and thus any basis function is correctly specified under the models LN_1, LN_2, LN_7–LN_9, and GAM_1, GAM_2, GAM_7–GAM_9.

Table 1: Parameter settings for simulation studies. In the second column, each LN_1–LN_{15} and each GAM_1–GAM_{15} denote mixture models whose continuous parts follow the distributions $\mathcal{LN}(a_i, b_i)$ and $\mathcal{GAM}(a_i, b_i)$, respectively, for $i \in \{0, 1\}$ under two-sample comparison, or for $i \in \{0, 1, 2\}$ under three-sample comparison.

Scenario	Model	(ν_0, ν_1, ν_2)	(a_0, a_1, a_2)	(b_0, b_1, b_2)	Means	Variances
I (null)	LN_1	(0.3, 0.3, 0.3)	(0.00, 0.00, 0.00)	(1.00, 1.00, 1.00)	(1.15, 1.15, 1.15)	(3.84, 3.84, 3.84)
	LN_2	(0.7, 0.7, 0.7)	(0.00, 0.00, 0.00)	(1.00, 1.00, 1.00)	(0.49, 0.49, 0.49)	(1.97, 1.97, 1.97)
II (null)	LN_3	(0.3, 0.5, 0.4)	(0.33, 0.66, 0.48)	(1.00, 1.00, 1.00)	(1.60 ,1.60, 1.60)	(7.38, 11.36, 9.04)
	LN_4	(0.5, 0.7, 0.6)	(0.37, 0.89, 0.60)	(1.00, 1.00, 1.00)	(1.20, 1.20, 1.20)	(6.39, 11.61, 8.35)
III (null)	LN_5	(0.3, 0.5, 0.4)	(0.05, 0.29, 0.16)	(0.80, 1.00, 0.90)	(1.10, 1.10, 1.10)	(2.64, 5.37, 3.75)
	LN_6	(0.5, 0.7, 0.6)	(0.00, 0.50, 0.25)	(0.94, 0.96, 0.89)	(0.80, 0.80, 0.80)	(2.64, 4.94, 3.24)
I (alternative)	LN_7	(0.5, 0.3, 0.4)	(0.00, 0.00, 0.00)	(1.00, 1.00, 1.00)	(0.82, 1.15, 0.99)	(3.01, 3.84, 3.45)
	LN_8	(0.7, 0.5, 0.6)	(0.00, 0.00, 0.00)	(1.00, 1.00, 1.00)	(0.49, 0.82, 0.66)	(1.97, 3.01, 2.52)
	LN_9	(0.6, 0.4, 0.5)	(0.00, 0.00, 0.00)	(1.00, 1.00, 1.00)	(0.66, 0.99, 0.82)	(2.52, 3.45, 3.01)
II (alternative)	LN_{10}	(0.3, 0.3, 0.3)	(0.00, 0.50, 0.25)	(1.00, 1.00, 1.00)	(1.15, 1.90, 1.48)	(3.84, 10.44, 6.33)
	LN_{11}	(0.7, 0.7, 0.7)	(0.00, 0.75, 0.50)	(1.00, 1.00, 1.00)	(0.49, 1.05, 0.82)	(1.97, 8.84, 5.36)
	LN_{12}	(0.4, 0.6, 0.5)	(0.00, 1.00, 0.50)	(1.00, 1.00, 1.00)	(0.99, 1.79, 1.36)	(3.45, 18.63, 8.20)
III (alternative)	LN_{13}	(0.3, 0.3, 0.3)	(0.00, 0.50, 0.25)	(1.00, 0.80, 0.90)	(1.15, 1.72, 1.41)	(3.84, 6.46, 4.99)
	LN_{14}	(0.7, 0.7, 0.7)	(0.00, 0.75, 0.50)	(1.00, 0.80, 0.90)	(0.49, 0.95, 0.78)	(1.97, 5.76, 4.33)
	LN_{15}	(0.6, 0.4, 0.5)	(0.00, 0.50, 0.25)	(1.00, 0.60, 0.80)	(0.66, 1.34, 0.96)	(2.52, 3.63, 3.17)
I (null)	GAM_1	(0.3, 0.3, 0.3)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(0.70, 0.70, 0.70)	(0.91, 0.91, 0.91)
	GAM_2	(0.7, 0.7, 0.7)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(0.30, 0.30, 0.30)	(0.51, 0.51, 0.51)
II (null)	GAM_3	(0.3, 0.5, 0.4)	(1.43, 2.00, 1.67)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(1.43, 2.00, 1.67)
	GAM_4	(0.5, 0.7, 0.6)	(2.00, 3.33, 2.50)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(2.00, 3.33, 2.50)
III (null)	GAM_5	(0.3, 0.5, 0.4)	(1.71, 1.20, 1.33)	(1.00, 2.00, 1.50)	(1.20, 1.20, 1.20)	(1.82, 3.84, 2.76)
	GAM_6	(0.5, 0.7, 0.6)	(2.00, 1.50, 1.75)	(1.00, 2.22, 1.43)	(1.00, 1.00, 1.00)	(2.00, 4.56, 2.93)
I (alternative)	GAM_7	(0.5, 0.3, 0.4)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(0.50, 0.70, 0.60)	(0.75, 0.91, 0.84)
	GAM_8	(0.7, 0.5, 0.6)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(0.30, 0.50, 0.40)	(0.51, 0.75, 0.64)
	GAM_9	(0.6, 0.4, 0.5)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(0.40, 0.60, 0.50)	(0.64, 0.84, 0.75)
II (alternative)	GAM_{10}	(0.3, 0.3, 0.3)	(1.00, 2.00, 1.50)	(1.00, 1.00, 1.00)	(0.70, 1.40, 1.05)	(0.91, 2.24, 1.52)
	GAM_{11}	(0.7, 0.7, 0.7)	(1.00, 2.00, 1.50)	(1.00, 1.00, 1.00)	(0.30, 0.60, 0.45)	(0.51, 1.44, 0.92)
	GAM_{12}	(0.4, 0.6, 0.5)	(1.00, 2.50, 1.50)	(1.00, 1.00, 1.00)	(0.60, 1.00, 0.75)	(0.84, 2.50, 1.31)
III (alternative)	GAM_{13}	(0.3, 0.3, 0.3)	(1.50, 1.00, 1.25)	(1.00, 2.00, 1.50)	(1.05, 1.40, 1.31)	(1.52, 3.64, 2.71)
	GAM_{14}	(0.7, 0.7, 0.7)	(1.75, 1.25, 1.50)	(1.00, 2.00, 1.50)	(0.53, 0.75, 0.68)	(1.17, 2.81, 2.08)
	GAM_{15}	(0.6, 0.4, 0.5)	(2.00, 1.00, 1.50)	(1.00, 2.00, 1.50)	(0.80, 1.20, 1.13)	(1.76, 3.36, 2.95)

Scenario II: The parameters a_is are not equal while the parameters b_is are held constant, and thus the basis function $\mathbf{q}(x) = \ln(x)$ is correctly specified under the models LN_3, LN_4, LN_{10}–LN_{12}, and GAM_3, GAM_4, GAM_{10}–GAM_{12}.

Scenario III: All the parameters a_is and b_is are not equal, and thus the basis function $\mathbf{q}(x) = (\ln(x), \ln^2(x))^\top$ is correctly specified under the LN models LN_5, LN_6, LN_{13}–LN_{15}, and the basis function $\mathbf{q}(x) = (x, \ln(x))^\top$ is correctly specified under the GAM models GAM_5, GAM_6, GAM_{13}–GAM_{15}.

In the following comparisons, the simulation results are discussed by the above three scenarios.

3.1. Scenario I

The simulated type I error rates of the ELR, ATS, and WTPS under Scenario I are summarized in Tables 2–3, and the simulated power under the same scenario are plotted in Figure 1. Here, the DRM (2) is correctly specified with any form of $\mathbf{q}(x)$. After experimenting with several forms of $\mathbf{q}(x)$, the basis function $\mathbf{q}(x) = \ln(x)$ is recommended since the ELR with such basis function has the most accurate type I error and the largest power. Hence, we only present the results under the basis function $\mathbf{q}(x) = \ln(x)$.

Table 2: Scenario I: Simulated probabilities (%) of rejecting $\mathcal{H}_0^*$ at 5% significance level when data are generated from log-normal mixture models according to the parameter settings given in Table 1. Here the ELR is defined under basis function $\mathbf{q}(x) = \ln(x)$.

Model	Two-sample Comparison				Three-sample Comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
LN_1	(50, 50)	5.10	4.05	4.85	(50, 50, 50)	5.01	3.03	4.97
	(100, 100)	4.83	4.84	5.15	(100, 100, 100)	5.10	3.77	4.75
	(50, 150)	5.10	5.85	4.92	(50, 150, 100)	5.01	5.00	5.06
	(150, 50)	4.94	5.91	4.84	(150, 50, 100)	5.14	4.75	5.08
LN_2	(50, 50)	4.96	3.68	5.01	(50, 50, 50)	4.77	2.45	5.13
	(100, 100)	5.07	4.04	4.73	(100, 100, 100)	4.87	3.01	4.98
	(50, 150)	4.74	5.78	3.99	(50, 150, 100)	5.20	4.67	4.97
	(150, 50)	4.77	6.15	4.25	(150, 50, 100)	5.15	4.79	5.02

* NOTE: The Monte Carlo error is 0.218 (%) under the null models LN_1–LN_2.

Based on our simulation results, our major observations for both the two- and three-sample comparisons are summarized as follows.

(a) It can be seen from the results in Tables 2–3, that the proposed ELR test and the WTPS method well control the type I error rates close to their nominal level. However, the ATS method tends to be conservative for equal sample sizes; and when the sample sizes are unequal it tends to be liberal for the two-sample comparisons.

Table 3: Scenario I: Simulated probabilities (%) of rejecting $\mathcal{H}_0^*$ at 5% significance level when data are generated from Gamma mixture models according to the parameter settings given in Table 1. Here the ELR is defined under basis function $\mathbf{q}(x) = \ln(x)$.

Model	Two-sample Comparison				Three-sample Comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
GAM_1	(50, 50)	5.16	4.68	4.82	(50, 50, 50)	5.47	4.34	4.99
	(100, 100)	4.98	4.56	4.84	(100, 100, 100)	5.17	4.62	5.09
	(50, 150)	4.83	5.89	5.18	(50, 150, 100)	5.28	4.82	4.94
	(150, 50)	5.13	5.74	5.08	(150, 50, 100)	5.32	5.11	5.07
GAM_2	(50, 50)	5.28	4.50	4.99	(50, 50, 50)	5.22	3.20	4.84
	(100, 100)	5.02	4.94	5.16	(100, 100, 100)	5.10	4.15	5.15
	(50, 150)	5.28	6.06	4.74	(50, 150, 100)	5.14	5.30	4.78
	(150, 50)	5.06	6.12	4.88	(150, 50, 100)	5.18	5.15	4.97

* NOTE: The Monte Carlo error is 0.218 (%) under the null models GAM_1–GAM_2.

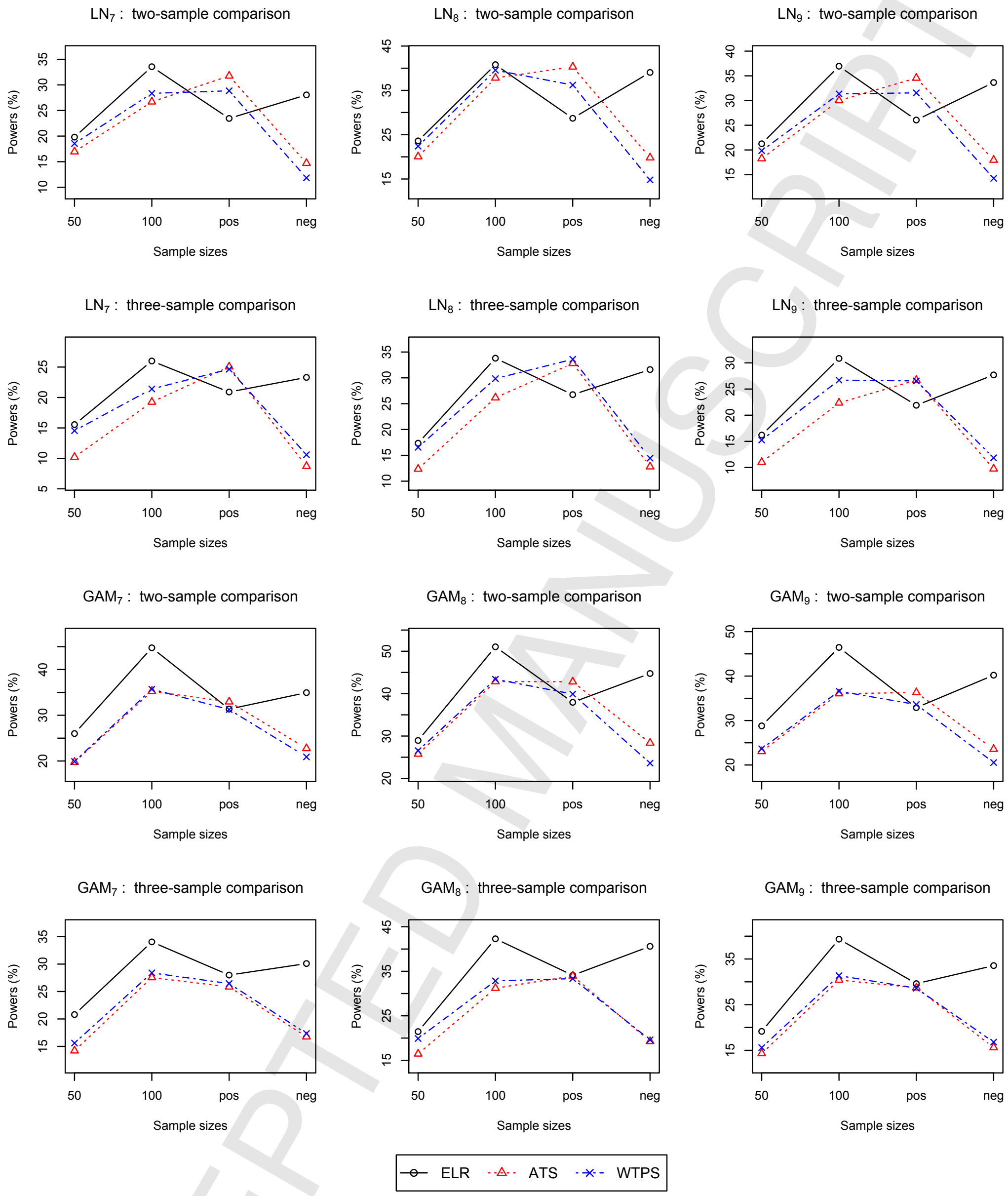

Figure 1: Scenario I: Simulated power (%) of rejecting $\mathcal{H}_0^*$ at 5% significance level when data are generated from log-normal mixture models or Gamma mixture models with parameter settings given in Table 1. The ELR test is defined under $\mathbf{q}(x) = \ln(x)$. The horizontal axis denotes combinations of sample sizes (n_0, n_1) equal to $(50, 50)$, $(100, 100)$, $(50, 150)$ and $(150, 50)$ for two-sample comparisons; and (n_0, n_1, n_2) equal to $(50, 50, 50)$, $(100, 100, 100)$, $(50, 150, 100)$ and $(150, 50, 100)$ for three-sample comparisons, from left to right.

(b) In terms of power, it can be observed from Figure 1, that the performance of the proposed ELR test seems to be less sensitive to the sample sizes than the other two tests. Although there is no uniformly dominant method for all the settings in Figure 1, the ELR test seems to be the most powerful for equal sample sizes and unequal sample sizes with a negative pairing. For the unequal sample sizes with positive pairing, the WTPS method may have some advantage over the ELR test for the LN models, and thus it may be favorable in this scenario, while recalling that the ATS method may have inflated type I error.

3.2. Scenario II

The simulated type I error rates for Scenario II are summarized in Tables 4–5, and the simulated power for Scenario II are plotted in Figure 2.

Table 4: Scenario II: Simulated probabilities (%) of rejecting $\mathcal{H}_0^*$ at 5% significance level when data are generated from log-normal mixture models according to the parameter settings given in Table 1. Here the ELR is defined under basis function $\mathbf{q}(x) = \ln(x)$.

Model	Two-sample Comparison				Three-sample Comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
LN_3	(50, 50)	5.27	4.26	5.08	(50, 50, 50)	5.01	3.29	4.97
	(100, 100)	5.12	5.31	5.74	(100, 100, 100)	5.19	3.74	5.08
	(50, 150)	5.13	5.48	4.32	(50, 150, 100)	5.37	4.17	4.37
	(150, 50)	5.11	7.37	6.22	(150, 50, 100)	5.28	6.27	6.51
LN_4	(50, 50)	5.19	5.16	5.53	(50, 50, 50)	5.10	3.60	4.87
	(100, 100)	5.34	4.89	5.42	(100, 100, 100)	5.41	3.85	5.79
	(50, 150)	5.42	4.82	3.06	(50, 150, 100)	5.23	3.61	3.67
	(150, 50)	5.36	8.88	7.52	(150, 50, 100)	5.47	7.18	7.67

* NOTE: The Monte Carlo error is 0.218 (%) under the null models LN_3–LN_4.

Table 5: Scenario II: Simulated probabilities (%) of rejecting $\mathcal{H}_0^*$ at 5% significance level when data are generated from Gamma mixture models according to the parameter settings given in Table 1. Here the ELR is defined under basis function $\mathbf{q}(x) = \ln(x)$.

Model	Two-sample Comparison				Three-sample Comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
GAM_3	(50, 50)	5.00	4.64	4.83	(50, 50, 50)	5.36	4.41	4.96
	(100, 100)	4.84	4.57	4.69	(100, 100, 100)	5.16	4.79	5.10
	(50, 150)	5.03	5.60	5.14	(50, 150, 100)	4.91	4.84	4.85
	(150, 50)	5.11	5.43	5.16	(150, 50, 100)	5.09	5.10	4.83
GAM_4	(50, 50)	5.10	5.21	5.39	(50, 50, 50)	5.04	4.29	4.91
	(100, 100)	5.28	5.26	5.33	(100, 100, 100)	5.25	4.82	5.10
	(50, 150)	5.16	5.02	4.38	(50, 150, 100)	5.30	4.84	4.69
	(150, 50)	4.92	6.05	5.58	(150, 50, 100)	4.93	5.65	5.29

* NOTE: The Monte Carlo error is 0.218 (%) under the null models GAM_3–GAM_4.

Based on our simulation results, our major observations for both the two- and three-sample comparisons are summarized as follows.

(c) In terms of type I error control, it can be seen from the results in Tables 4–5, that the ELR test under $\mathbf{q}(x) = \ln(x)$ always retains error rates close to the nominal level, and the results seem insensitive to whether sample sizes are equal or not. On the other hand, the type I error rates of both ATS and WTPS methods seem to be sensitive to whether the sample sizes are equal or not. For equal sample sizes, both ATS and WTPS methods control the type I error satisfactory, although the ATS method may be conservative in some settings. For unequal sample sizes with positive pairing, both the ATS and WTPS methods are conservative in their type I error rates; however, with the negative pairing, they tend to have inflated type I error rates, particularly for the LN models.

(d) In terms of power, it can be observed from Figure 2, that the performance of the proposed ELR test is, again, not as sensitive to unequal sample sizes as the other two tests. Further, in all the settings in Figure 2, the ELR test is the most, or one of the most, powerful tests, for both the equal and unequal sample sizes under comparisons. In some special cases, the gain in power could be over 50%. Together with the observations from the type I error rates, the proposed ELR test under $\mathbf{q}(x) = \ln(x)$ is much preferred in this scenario.

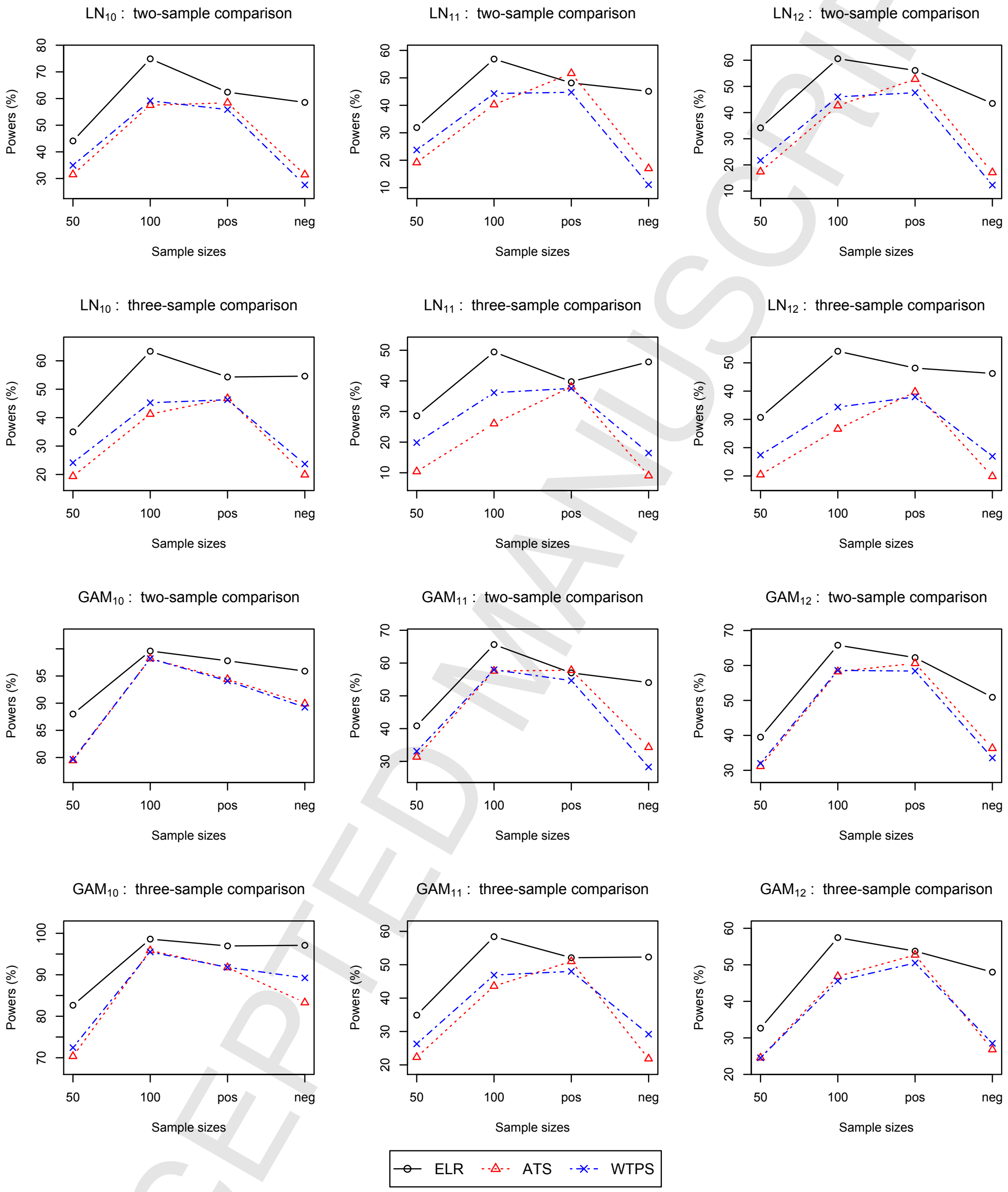

Figure 2: Scenario II: Simulated power (%) of rejecting $\mathcal{H}_0^*$ at 5% significance level when data are generated from log-normal mixture models or Gamma mixture models with parameter settings given in Table 1. The ELR test is defined under $\mathbf{q}(x) = \ln(x)$. The horizontal axis denotes combinations of sample sizes (n_0, n_1) equal to $(50, 50)$, $(100, 100)$, $(50, 150)$ and $(150, 50)$ for two-sample comparisons; and (n_0, n_1, n_2) equal to $(50, 50, 50)$, $(100, 100, 100)$, $(50, 150, 100)$ and $(150, 50, 100)$ for three-sample comparisons, from left to right.

3.3. Scenario III

The simulated type I error rates for Scenario III are summarized in Tables 6–7, and the simulated power for Scenario III are plotted in Figure 3.

Table 6: Scenario III: Simulated probabilities (%) of rejecting $\mathcal{H}_0^*$ at 5% significance level when data are generated from log-normal mixture models according to the parameter settings given in Table 1. Here the ELR is defined under basis function $\mathbf{q}(x) = (\ln(x), \ln^2(x))^\top$.

Model	Two-sample Comparison				Three-sample Comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
LN_5	(50, 50)	6.00	4.97	5.53	(50, 50, 50)	5.82	3.89	5.89
	(100, 100)	5.64	5.17	5.49	(100, 100, 100)	5.44	4.11	5.41
	(50, 150)	5.28	4.49	3.31	(50, 150, 100)	4.93	3.58	3.68
	(150, 50)	6.05	8.11	7.16	(150, 50, 100)	5.91	6.16	6.40
LN_6	(50, 50)	5.68	4.74	5.57	(50, 50, 50)	5.36	3.43	5.33
	(100, 100)	5.48	5.22	5.79	(100, 100, 100)	5.33	3.90	5.62
	(50, 150)	4.91	4.70	3.10	(50, 150, 100)	5.23	3.85	3.73
	(150, 50)	5.79	8.19	6.72	(150, 50, 100)	6.10	7.24	7.08

* NOTE: The Monte Carlo error is 0.218 (%) under the null models LN_5–LN_6.

Table 7: Scenario III: Simulated probabilities (%) of rejecting $\mathcal{H}_0^*$ at 5% significance level when data are generated from Gamma mixture models according to the parameter settings given in Table 1. Here the ELR is defined under basis function $\mathbf{q}(x) = \{x, \ln(x)\}^\top$.

Model	Two-sample Comparison				Three-sample Comparison			
	(n_0, n_1)	ELR	ATS	WTPS	(n_0, n_1, n_2)	ELR	ATS	WTPS
GAM_5	(50, 50)	5.86	5.18	5.29	(50, 50, 50)	5.76	4.75	5.44
	(100, 100)	5.40	4.92	5.03	(100, 100, 100)	5.21	4.61	4.85
	(50, 150)	5.15	4.92	4.17	(50, 150, 100)	5.20	4.69	4.32
	(150, 50)	5.76	6.20	6.01	(150, 50, 100)	5.59	5.96	5.83
GAM_6	(50, 50)	5.91	5.46	5.63	(50, 50, 50)	5.93	4.60	5.53
	(100, 100)	5.82	5.51	5.61	(100, 100, 100)	5.46	4.99	5.50
	(50, 150)	5.42	4.96	3.90	(50, 150, 100)	5.22	4.45	4.02
	(150, 50)	6.24	7.15	6.67	(150, 50, 100)	5.83	6.96	6.53

* NOTE: The Monte Carlo error is 0.218 (%) under the null models GAM_5–GAM_6.

Based on our simulation results, our major observations for both two-sample and three-sample comparisons are summarized as follows.

(e) In terms of type I error control, it can be seen from the results in Tables 6–7 that the ELR test, under basis functions $\mathbf{q}(x) = (\ln(x), \ln^2(x))^\top$ for LN models, or $\mathbf{q}(x) = (x, \ln(x))^\top$ for GAM models, may fail to keep the error under control, except for the unequal sample sizes with positive pairing. In this scenario, the WTPS method may also lead to inflated type I error rates in some settings. On the other hand, the ATS method seems to have overall good control of the type I error, except for the case of unequal sample sizes with negative pairing. Indeed, when the sample sizes are unequal with negative pairing, all three methods under comparison lead to inflated type I error rates. In general, for the cases of equal sample sizes and unequal sample sizes with positive pairing, the ATS is a suitable testing method in this scenario. In this scenario with more complex data distributions, the ELR and WTPS methods seem to need larger sample sizes than those considered to give adequate approximations.

(f) In terms of power, it can be observed from Figure 3 that the ATS method still has consistent performance in the settings of equal sample sizes and unequal sample sizes with positive pairing. However, keep in mind that the ELR and WTPS methods both may fail to control the type I error rates. Therefore, no fair comparison can be made with ELR and WTPS methods.

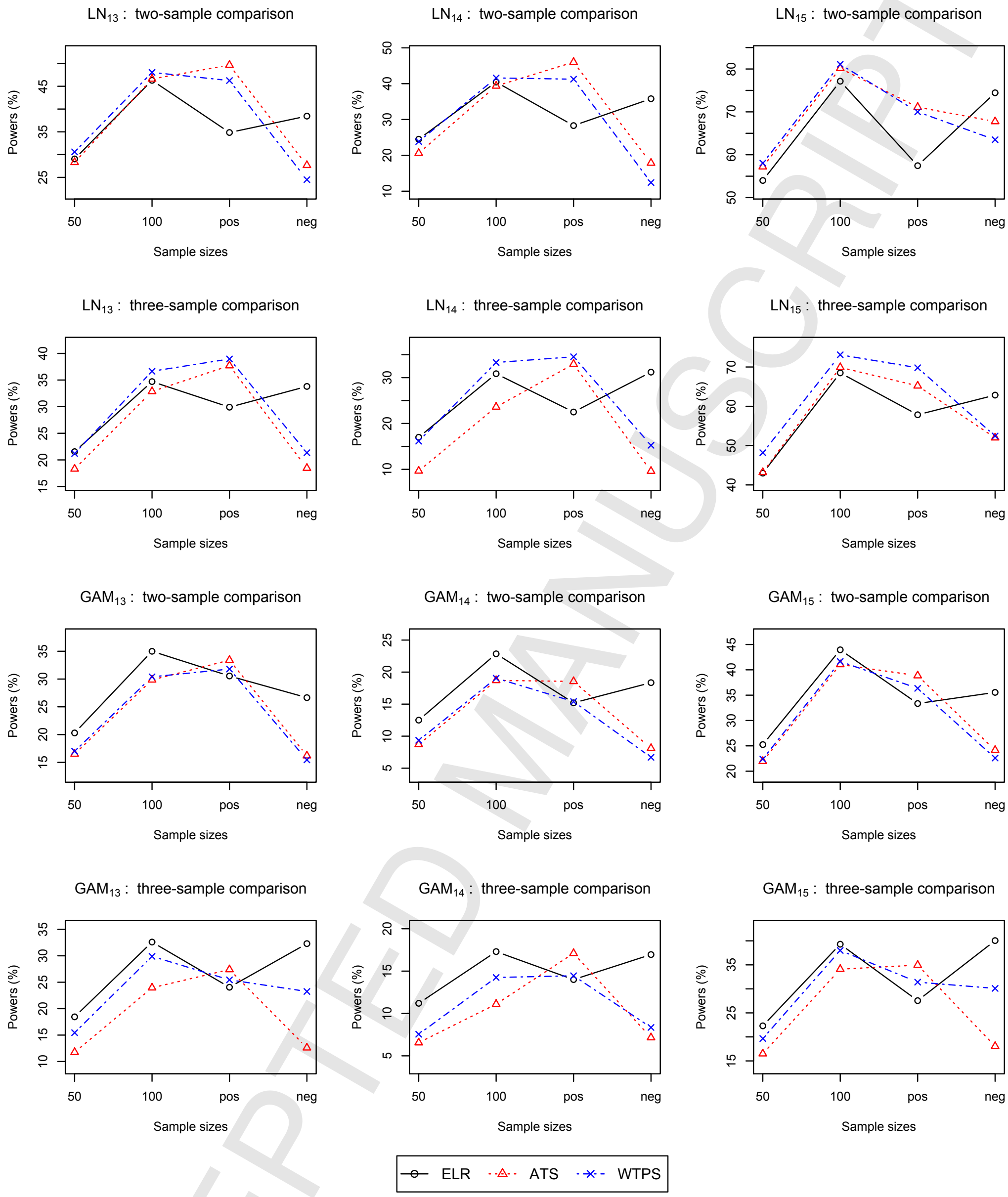

Figure 3: Scenario III: Simulated power (%) of rejecting $\mathcal{H}_0^*$ at 5% significance level when data are generated from log-normal mixture models or Gamma mixture models with parameter settings given in Table 1. The ELR test is defined under $\mathbf{q}(x) = (\ln(x), \ln^2(x))^\top$. The horizontal axis denotes combinations of sample sizes (n_0, n_1) equal to $(50, 50)$, $(100, 100)$, $(50, 150)$ and $(150, 50)$ for two-sample comparisons; and (n_0, n_1, n_2) equal to $(50, 50, 50)$, $(100, 100, 100)$, $(50, 150, 100)$ and $(150, 50, 100)$ for three-sample comparisons, from left to right.

In summary, for testing mean equality in the three scenarios considered, the proposed ELR test is robust to the assumption of parametric models that generate the data. The ELR test generally performs best in Scenario II with basis function $\mathbf{q}(x) = \ln(x)$ correctly specified in a DRM. As we may expect, the failure of selecting an appropriate form of $\mathbf{q}(x)$ in a DRM may affect the control of type I error of the ELR test in some settings, and potential convergence problems in computation can also be an issue.

In our simulation studies, the numerical procedure for calculating ELR, discussed in Appendix A, converges fast and is in general not sensitive to the selection of initial values when the basis function $\mathbf{q}(x) = \ln(x)$ is used in Scenarios I and II. However, when the basis function of increasing dimension are specified in Scenario III, we found that the computation may not be very stable and a good choice of initial value can be helpful. Besides, in our experience, the ELR with correctly specified basis function is also important for convergence in Scenarios II and III.

In practice, a comprehensive DRM selection strategy, as described in [9], would be encouraged at a preliminary stage of data analysis. The performance of Akaike's information criterion (AIC) was evaluated in [9], and was shown to be robust for moderate sample sizes. We comment that such a preliminary DRM selection procedure is not too restrictive. Given the difficulty in identifying a suitable parametric model for many practical situations, the proposed ELR test can be an attractive and robust semiparametric alternative approach.

Combining the conclusions of our simulation results and numerical experience, we recommend using the ELR test when the basis function $\mathbf{q}(x) = \ln(x)$ is selected in the DRM.

We also comment that for the three-sample comparisons, the unequal sample sizes combinations presented in this section correspond to two extreme cases that are perfectly positive, or negative, pairing with the unequal variances. For these cases with positive and negative paring, the findings for ATS and WTPS methods are consistent with those in [28] and [19]. We do not report all possible patterns of unequal sample sizes and variances combinations in simulation. For other combinations that are somewhat in between those reported, the pattern of results do not exhibit large difference from those reported. In general, the ELR test has been observed to be robust against unequal sample sizes when compared with the ATS and WTPS methods. In the next section, we examine a real data example in which the relation between sample sizes and variances are not dramatically positively or negatively paired.

4. An illustrative real data example

In this section, we analyze a real data set from Koopmans [15, p. 107]. It arises in a biological study of the seasonal activity patterns of a species of field mice. The measurements are the average distances (in meters) travelled between captures by those mice at least twice in a given month. One of the objectives in this study is to discover if the mean measurements differ between the four seasons. In addition to the continuous positive measurements, there are substantial proportions of zero values, especially in the fall and winter data. Some summary statistics are:

- ✓ the sample estimate of $\boldsymbol{\nu}^\top$ is $(0.18, 0.11, 0.37, 0.29)$ with sample sizes $(17, 27, 27, 34)$;
- ✓ the sample means are $(26.94, 30.81, 13.04, 15.03)$;
- ✓ the sample variances are $(679.31, 1118.93, 178.88, 293.54)$.

As discussed in Section 3, we need to select a basis function $\mathbf{q}(x)$ in a DRM that provides a reasonable fit to this data. We apply the AIC, discussed in [9], to select a basis function in the DRM for the positive data in this example. The results are given in Table 8. It can be seen that the DRM with $\mathbf{q}(x) = \ln(x)$ has the smallest AIC among five commonly used basis functions, and hence it is recommended in this example. Furthermore, we have also applied the ELR test for homogeneity in distributions developed in [31] to this real data set; their bootstrap ELR test for homogeneity based on $\mathbf{q}(x) = \ln(x)$ gives a p-value of 0.0402, with $10,000$ times bootstrap resampling. With this preliminary data analysis, it seems reasonable to categorize this real data example into the Scenario II considered in Section 3.2.

Table 8: AIC for five commonly used basis functions $\mathbf{q}(x)$ in a DRM for the positive field mice data.

$\mathbf{q}(x)$	$\ln(x)$	x	$(\ln(x), \ln^2(x))^\top$	$(x, \ln(x))^\top$	$(x, \ln(x), \ln^2(x))^\top$
AIC	219.63	220.45	224.19	224.05	228.86

Table 9: Fitted parameters for log-normal mixture models and Gamma mixture models under the null and alternative hypotheses for field mice data. The models LN_{16} and GAM_{16} are fitted under the null hypothesis; and the models LN_{17} and GAM_{17} are fitted under the alternative hypothesis. The last two columns are the means and variances corresponding to each model.

Model	$(\nu_0, \nu_1, \nu_2, \nu_3)$	(a_0, a_1, a_2, a_3)	(b_0, b_1, b_2, b_3)	Means	Variances
LN_{16}	$(0.28, 0.21, 0.28, 0.23)$	$(3.11, 3.03, 3.12, 3.05)$	$(0.41, 0.41, 0.41, 0.41)$	$(19.95, 19.95, 19.95, 19.95)$	$(430.69, 362.72, 433.63, 376.22)$
LN_{17}	$(0.18, 0.11, 0.37, 0.29)$	$(3.29, 3.25, 2.91, 2.87)$	$(0.37, 0.37, 0.37, 0.37)$	$(26.68, 27.49, 13.84, 14.99)$	$(539.26, 474.92, 248.75, 235.92)$
GAM_{16}	$(0.27, 0.20, 0.28, 0.23)$	$(2.32, 2.13, 2.35, 2.20)$	$(12.22, 12.22, 12.22, 12.22)$	$(20.68, 20.68, 20.68, 20.68)$	$(410.53, 362.76, 417.95, 381.00)$
GAM_{17}	$(0.18, 0.11, 0.37, 0.29)$	$(2.88, 2.77, 2.10, 2.04)$	$(11.24, 11.24, 11.24, 11.24)$	$(26.65, 27.67, 14.87, 16.22)$	$(451.68, 406.71, 297.12, 291.97)$

We further fit this data by log-normal mixture models and Gamma mixture models, under the null and alternative hypotheses, by the parametric maximum likelihood method. Here the null hypothesis is that the means of measurements for all four seasons are the same. The details are provided in Table 9. These fitted models will be used in our confirmative simulation later on.

We then applied the proposed ELR test, and other testing methods discussed in Section 3, for the mean equality in this real data. The observed test statistics and their corresponding p-values are reported in Table 10.

Table 10: Test statistics, corresponding p-values, and confirmative simulation for field mice data. The ELR test is under basis function $\mathbf{q}(x) = \ln(x)$ in a DRM.

Pair	Method	Field Mice Data		Confirmative Simulation			
		Test statistic	p-value	LN_{16}	LN_{17}	GAM_{16}	GAM_{17}
All four seasons	ELR	12.40	0.00613	5.20	86.36	5.21	79.11
	ATS	3.17	0.0435	3.92	68.11	3.99	62.07
	WTPS	9.71	0.0388	4.83	76.58	4.67	66.88
Spring vs. Autumn	ELR	5.65	0.0175	5.05	59.35	5.29	53.19
	ATS	4.15	0.0542	4.36	47.32	4.61	42.50
	WTPS	4.15	0.0431	4.56	49.07	4.61	43.05
Spring vs. Winter	ELR	4.95	0.0261	4.91	57.24	5.23	45.97
	ATS	2.92	0.1009	4.69	42.15	4.90	34.24
	WTPS	2.92	0.0991	4.62	43.61	4.80	34.56
Summer vs. Autumn	ELR	6.56	0.0105	5.16	79.56	5.29	74.43
	ATS	6.58	0.0149	4.40	74.16	4.72	68.31
	WTPS	6.58	0.0071	4.88	74.92	4.91	68.41
Summer vs. Winter	ELR	6.62	0.0101	5.37	78.91	5.31	70.61
	ATS	4.98	0.0319	4.50	71.83	4.77	62.69
	WTPS	4.98	0.0262	4.80	72.96	4.92	62.91

From the results in Table 10, the proposed ELR test as well as the other two tests all produce significant p-values at 5% significance level. It is worth emphasizing that the proposed ELR test provides the strongest evidence in terms of its p-value. Therefore, a natural followup concern is to detect any potential pairwise mean differences. We further test for the two-sample mean equality for six pairwise combinations of this data. The testing results for the significant pairs are given in Table 10. We particularly highlight one pairwise comparison, that is, Spring vs. Winter. The proposed ELR test in this two-sample mean comparison gives a significant p-value of 0.0261 at 5% significance level. In comparison, both the ATS and WTPS methods fail to detect the mean difference for this particular pair at 5% significance level.

These results may be further verified by the simulation according to model settings in Table 9. The simulation results based on 10,000 repetitions are summarized in Table 10. It can be seen that all the simulation results demonstrate very good agreement with our real data analysis.

5. Concluding remarks

In this paper, we discussed the problem of making statistical inferences on the means of multiple distributions with excess zero observations. Under the semiparametric framework developed in [31], we proposed an ELR statistic and derived its limiting distribution under a fairly general linear null hypothesis about the means. We illustrated the good behaviour of the proposed ELR in finite sample simulation studies, and also with a real data example, where the emphasis was on testing mean equality. In particular, we identified a scenario, where the ELR showed a clear advantage, when $\mathbf{q}(x) = \ln(x)$ is the correctly specified basis function in the DRM, compared with other existing tests for mean equality.

It would be interesting to further consider the model selection problem for density ratios, and to approximate the power function of the test by exploring the limiting distribution of the ELR statistic under an alternative. We leave this for future work. The proposed ELR based inference framework on the means can also be extended to more general settings. For example, given that we have prior knowledge that all, or part of, the group means are equal [11, 12, 26], then the proposed framework can be used to obtain refined inference results.

Acknowledgments

The authors thank the Editor-in-Chief, Prof. Christian Genest, the Associate Editor, and three referees for their very careful reading and valuable comments and suggestions that led to an improved version of the article. This work was supported by the Natural Sciences and Engineering Research Council of Canada, for Dr. Marriott (Grant RGPIN-2014-05424) and for Dr. Li (Grant RGPIN-2015-06592), and by the Fundamental Research Funds for the Central Universities for Dr. Wang (Grants 20720181043 and 20720181003).

Online Supplement

The Online Supplement contains a proof of Theorem 1.

Appendix A. Numerical implementation

In this Appendix, we discuss numerical calculation of R_n, defined in (6), for some given $\mathbf{C}$ and $\mathbf{d}$. In what follows, $\mathbf{I}_m$ denotes a $m \times m$ unit diagonal matrix, and $\boldsymbol{\iota}$ denotes a vector with all entries being 1.

We first discuss how to calculate $\sup_{(\boldsymbol{\nu},\boldsymbol{\theta},G_0)\in C_1} \tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)$ in R_n. Note that the optimization problem of maximizing $\tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)$ subject to C_1 for given $(\boldsymbol{\nu},\boldsymbol{\theta})$ is identical to the one discussed in [31]. Following a similar profiling procedure as used in [21] and the results of [4, 14] on the *dual* empirical log-likelihood, we have

$$\sup_{(\boldsymbol{\nu},\boldsymbol{\theta},G_0)\in C_1} \tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0) = \sup_{(\boldsymbol{\nu},\boldsymbol{\theta})} \ell_A(\boldsymbol{\nu},\boldsymbol{\theta}) - n_{\cdot 1}\ln(n_{\cdot 1}), \tag{A.1}$$

where

$$\ell_A(\boldsymbol{\nu},\boldsymbol{\theta}) = \sum_{i=0}^{m} \ln\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\} + \sum_{i=1}^{m}\sum_{j=1}^{n_{i1}}\{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij})\} - \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} \ln\left[\rho_0 + \sum_{r=1}^{m}\rho_r \exp\{\alpha_r + \boldsymbol{\beta}_r^\top \mathbf{q}(x_{ij})\}\right],$$

with $\rho_r = n_{r1}/n_{\cdot 1}$ for $r \in \{0,\ldots,m\}$. The numerical calculation of $(\hat{\boldsymbol{\nu}},\hat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{\nu},\boldsymbol{\theta})} \ell_A(\boldsymbol{\nu},\boldsymbol{\theta})$ and $\ell_A(\hat{\boldsymbol{\nu}},\hat{\boldsymbol{\theta}}) = \sup_{(\boldsymbol{\nu},\boldsymbol{\theta})} \ell_A(\boldsymbol{\nu},\boldsymbol{\theta})$ can be solved straightforwardly via the connection with logistic regression, as discussed in [31].

We next discuss how to calculate $\sup_{(\boldsymbol{\nu},\boldsymbol{\theta},G_0)\in C_1\cap C_2} \tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)$ in R_n. We start with the profiling procedure of $\tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0)$ by profiling out the infinite-dimensional parameter G_0. First, we set up the Lagrangian function. For given $(\boldsymbol{\nu},\boldsymbol{\theta})$, define

$$\Psi(G_0,\boldsymbol{\lambda},\mathbf{t}) = \tilde{\ell}(\boldsymbol{\nu},\boldsymbol{\theta},G_0) + \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} p_{ij}\boldsymbol{\lambda}^\top\{\boldsymbol{\omega}(x_{ij},\boldsymbol{\theta}) - \boldsymbol{\iota}\} + \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} p_{ij}\mathbf{t}^\top \mathbf{g}(x_{ij};\boldsymbol{\nu},\boldsymbol{\theta}),$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^\top$ and $\mathbf{t} = (t_1, \ldots, t_p)^\top$ are corresponding Lagrangian multipliers. The point $(p_{i1}, \ldots, p_{in_i} : i \in \{0, \ldots, m\})$ that maximize $\tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0)$ must be a stationary point of Ψ satisfying

$$\partial\Psi(G_0, \boldsymbol{\lambda}, \mathbf{t})/\partial p_{ij} = 0, \quad \partial\Psi(G_0, \boldsymbol{\lambda}, \mathbf{t})/\partial\lambda_i = 0, \quad \text{and} \quad \partial\Psi(G_0, \boldsymbol{\lambda}, \mathbf{t})/\partial t_i = 0. \tag{A.2}$$

It follows from (A.2) that, for fixed $(\boldsymbol{\nu}, \boldsymbol{\theta})$, $\tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0)$ attains its maximum at

$$p_{ij} = \frac{1}{n_{\cdot 1}} \times \frac{1}{1 + \boldsymbol{\lambda}^\top\{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \boldsymbol{\iota}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta})}, \tag{A.3}$$

where the Lagrange multipliers, $\boldsymbol{\lambda}$ and $\mathbf{t}$, solve following equations

$$\frac{1}{n_{\cdot 1}} \sum_{i=0}^{m} \sum_{j=1}^{n_{i1}} \frac{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \boldsymbol{\iota}}{1 + \boldsymbol{\lambda}^\top\{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \boldsymbol{\iota}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta})} = \mathbf{0}_{m\times 1}, \tag{A.4}$$

and

$$\frac{1}{n_{\cdot 1}} \sum_{i=0}^{m} \sum_{j=1}^{n_{i1}} \frac{\mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta})}{1 + \boldsymbol{\lambda}^\top\{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \boldsymbol{\iota}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta})} = \mathbf{0}_{p\times 1}. \tag{A.5}$$

Therefore, using (A.3) to profile out p_{ij}, the profile empirical log-likelihood function of $(\boldsymbol{\nu}, \boldsymbol{\theta})$ under the null hypothesis $\mathcal{H}_0$ given in (3) can be written as $\ell_N(\boldsymbol{\nu}, \boldsymbol{\theta}) - n_{\cdot 1}\ln(n_{\cdot 1})$ with

$$\ell_N(\boldsymbol{\nu}, \boldsymbol{\theta}) = \sum_{i=0}^{m} \ln\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\} + \sum_{i=1}^{m}\sum_{j=1}^{n_{i1}} \{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij})\} - \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} \ln\left[1 + \boldsymbol{\lambda}^\top\{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \boldsymbol{\iota}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta})\right].$$

Then, it follows that

$$\sup_{(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) \in C_1 \cap C_2} \tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) = \sup_{(\boldsymbol{\nu}, \boldsymbol{\theta})} \ell_N(\boldsymbol{\nu}, \boldsymbol{\theta}) - n_{\cdot 1}\ln(n_{\cdot 1}). \tag{A.6}$$

Let

$$(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{\nu}, \boldsymbol{\theta})} \ell_N(\boldsymbol{\nu}, \boldsymbol{\theta})$$

be the maximum EL estimate of $(\boldsymbol{\nu}, \boldsymbol{\theta})$ under $\mathcal{H}_0$ in (3). Hence, to calculate $\sup_{(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) \in C_1 \cap C_2} \tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0)$, it is sufficient to obtain $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})$.

Unfortunately, the numerical calculation of $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})$ may not be an easy task since there are no analytical solutions for $\boldsymbol{\lambda}$ and $\mathbf{t}$ in the definition of $\ell_N(\boldsymbol{\nu}, \boldsymbol{\theta})$.

Let $\boldsymbol{\psi} = (\boldsymbol{\nu}^\top, \boldsymbol{\theta}^\top, \boldsymbol{\lambda}^\top, \mathbf{t}^\top)^\top$ and define

$$\ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t}) = \sum_{i=0}^{m} \ln\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\} + \sum_{i=1}^{m}\sum_{j=1}^{n_{i1}} \{\alpha_i + \boldsymbol{\beta}_i^\top \mathbf{q}(x_{ij})\} - \sum_{i=0}^{m}\sum_{j=1}^{n_{i1}} \ln\left[1 + \boldsymbol{\lambda}^\top\{\boldsymbol{\omega}(x_{ij}, \boldsymbol{\theta}) - \boldsymbol{\iota}\} + \mathbf{t}^\top \mathbf{g}(x_{ij}; \boldsymbol{\nu}, \boldsymbol{\theta})\right]. \tag{A.7}$$

Then $\ell_N(\boldsymbol{\nu}, \boldsymbol{\theta}) = \ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t})$ with $\boldsymbol{\lambda}$ and $\mathbf{t}$ being the solutions of (A.4) and (A.5). Note that, $\boldsymbol{\lambda}$ and $\mathbf{t}$ in $\ell_N(\boldsymbol{\nu}, \boldsymbol{\theta})$ are actually functions of $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$.

Let $\tilde{\boldsymbol{\lambda}}$ and $\tilde{\mathbf{t}}$ be the solutions of (A.4) and (A.5) when $(\boldsymbol{\nu}, \boldsymbol{\theta})$ is replaced by $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})$. Hence, $\ell(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}}) = \ell_N(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})$. Further let $\tilde{\boldsymbol{\psi}} = (\tilde{\boldsymbol{\nu}}^\top, \tilde{\boldsymbol{\theta}}^\top, \tilde{\boldsymbol{\lambda}}^\top, \tilde{\mathbf{t}}^\top)^\top$.

We summarize some key properties of using $\ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t})$ to find $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{\nu}, \boldsymbol{\theta})} \ell_N(\boldsymbol{\nu}, \boldsymbol{\theta})$ and $\ell_N(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}) = \sup_{(\boldsymbol{\nu}, \boldsymbol{\theta})} \ell_N(\boldsymbol{\nu}, \boldsymbol{\theta})$ in the following proposition.

Proposition 1. *For $\ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t})$, defined in* (A.7)*, and $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{\nu}, \boldsymbol{\theta})} \ell_N(\boldsymbol{\nu}, \boldsymbol{\theta})$, then $\tilde{\boldsymbol{\psi}}$ is a stationary point of $\ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t})$. That is, $\tilde{\boldsymbol{\psi}}$ is a solution of $\partial\ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t})/\partial\boldsymbol{\psi} = \mathbf{0}$.*

Proof. We sketch some key steps. First, when $\ell_N(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})$ is maximized, the following equations are satisfied:

$$\partial\ell_N(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})/\partial\boldsymbol{\nu} = \mathbf{0}, \quad \partial\ell_N(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})/\partial\boldsymbol{\theta} = \mathbf{0}.$$

Further, note that since $\boldsymbol{\lambda}$ and $\mathbf{t}$ in $\ell_N(\boldsymbol{\nu}, \boldsymbol{\theta})$ are the solutions of (A.4) and (A.5), they are actually functions of $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$. For illustration, we only show

$$\begin{aligned}\mathbf{0} = \frac{\partial\ell_N(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}_i} &= \frac{\partial\ell(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}})}{\partial\boldsymbol{\beta}_i} + \sum_{r=1}^{m} \frac{\partial\ell(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}})}{\partial\lambda_r} \frac{\partial\lambda_r}{\partial\boldsymbol{\beta}_i}\Big|_{(\boldsymbol{\nu},\boldsymbol{\theta})=(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})} + \sum_{s=1}^{p} \frac{\partial\ell(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}})}{\partial t_s} \frac{\partial t_s}{\partial\boldsymbol{\beta}_i}\Big|_{(\boldsymbol{\nu},\boldsymbol{\theta})=(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})}\\ &= \frac{\partial\ell(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}})}{\partial\boldsymbol{\beta}_i} - 0 \times \frac{\partial\lambda_r}{\boldsymbol{\beta}_i}\Big|_{(\boldsymbol{\nu},\boldsymbol{\theta})=(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})} - 0 \times \frac{\partial t_s}{\boldsymbol{\beta}_i}\Big|_{(\boldsymbol{\nu},\boldsymbol{\theta})=(\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\theta}})} = \frac{\partial\ell(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}})}{\partial\boldsymbol{\beta}_i},\end{aligned}$$

where the second last line is followed from (A.4) and (A.5). One can similarly verify that

$$\partial\ell_N(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})/\partial\boldsymbol{\eta} = \partial\ell(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{t}})/\partial\boldsymbol{\eta} = \mathbf{0}.$$

Hence $\tilde{\boldsymbol{\psi}}$ is a stationary point of $\ell(\boldsymbol{\theta})$. This completes the proof. □

With the result of Proposition 1, $(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})$ can be obtained by solving for a stationary point, in fact a saddlepoint of $\ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t})$, over the space of $\boldsymbol{\psi}$. To numerically calculate $\tilde{\boldsymbol{\psi}}$, we minimize the sum of squares of $\partial\ell(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{t})/\partial\boldsymbol{\psi}$ by using the built-in `nlminb` function in R [24]. Once we obtain $\tilde{\boldsymbol{\psi}}$, we calculate $\sup_{(\boldsymbol{\nu},\boldsymbol{\theta},G_0)\in C_1 \cap C_2} \tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0)$ by (A.6).

Combining (A.1) and (A.6), we finish the numeric calculation of R_n by

$$R_n = 2\{\ell_A(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}) - \ell_N(\tilde{\boldsymbol{\nu}}, \tilde{\boldsymbol{\theta}})\}.$$

Appendix B. Definition of the U matrix

This Appendix gives the definition of the $\mathbf{U}$ matrix, which is involved in R5 of the regularity conditions in Section 2.3. The inverse of this $\mathbf{U}$ matrix is required in a quadratic approximation of R_n, as shown in the Online Supplement.

We first need some notation before we proceed. Recall that the true value of $\boldsymbol{\eta} = (\boldsymbol{\nu}^\top, \boldsymbol{\theta}^\top)^\top$ is $\boldsymbol{\eta}^* = (\boldsymbol{\nu}^{*\top}, \boldsymbol{\theta}^{*\top})^\top$ under the null hypothesis $\mathcal{H}_0$. For simplicity, recall that we write $\rho_i^* = n_i/n$, for $i \in \{0, \ldots, m\}$, and assume that it is a constant. Further, we denote $\Delta^* = \sum_{i=0}^m \rho_i^*(1-\nu_i^*)$, and $\lambda_r^* = \rho_r^*(1-\nu_r^*)/\Delta^*$, for all $r \in \{1, \ldots, m\}$. Let $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_m^*)^\top$. For a compact presentation, we further denote

$$\lambda_0^* = 1 - \sum_{i=1}^m \lambda_i^*, \quad h(X; \boldsymbol{\eta}^*) = \sum_{i=0}^m \lambda_i^* \omega_i(X; \boldsymbol{\theta}_i^*).$$

Define, for each $i \in \{0, \ldots, m\}$,

$$h_i(X; \boldsymbol{\eta}^*) = \lambda_i^* \omega_i(X; \boldsymbol{\theta}_i^*)/h(X; \boldsymbol{\eta}^*).$$

Finally, let $\mathbf{Q}(X) = (1, \mathbf{q}^\top(X))^\top$, $\mathbf{h} = (h_1(X; \boldsymbol{\eta}^*), \ldots, h_m(X; \boldsymbol{\eta}^*))^\top$, $\boldsymbol{\omega} = (\omega_1(X; \boldsymbol{\theta}_0^*), \ldots, \omega_m(X; \boldsymbol{\theta}_m^*))^\top$, as well as $\mathbf{g} = (g_1(X; \boldsymbol{\eta}^*), \ldots, g_p(X; \boldsymbol{\eta}^*))^\top$.

Now we can define the form of the $\mathbf{U}$ matrix as follows. Let $\otimes$ denote the Kronecker product. Define

$$\mathbf{U} = A_{44} + A_{41}A_{11}^{-1}A_{14} + A_{42}A_{22}^{-1}A_{24} - V(A^{33})^{-1}V^\top, \tag{B.1}$$

where

$$\begin{aligned} A_{11} &= \text{diag}\left\{\frac{\rho_0^*}{\nu_0^*(1-\nu_0^*)}, \ldots, \frac{\rho_m^*}{\nu_m^*(1-\nu_m^*)}\right\}, \quad A_{14}^\top = A_{41} = -\Delta^*\mathbf{C}\text{diag}(\boldsymbol{\mu})\text{diag}\{(\boldsymbol{\iota} - \boldsymbol{\nu}^*)^{-1}\},\\ A_{22} &= \Delta^*\text{E}_0\left[h(X; \boldsymbol{\eta}^*)\{\text{diag}(\mathbf{h}) - (\mathbf{h}\mathbf{h}^\top)\} \otimes \{\mathbf{Q}(X)\mathbf{Q}^\top(X)\}\right],\\ A_{24} &= A_{42}^\top = \Delta^*\text{E}_0\Big[\Big\{\text{diag}(\boldsymbol{\omega})(\mathbf{0}_{m\times1}, \mathbf{I}_m)\text{diag}(\boldsymbol{\iota} - \boldsymbol{\nu}^*)\mathbf{C}^\top X - (\mathbf{h}\mathbf{g}^\top)\Big\} \otimes \mathbf{Q}(X)\Big],\\ A_{44} &= \Delta^*\text{E}_0\left\{\frac{\mathbf{g}\mathbf{g}^\top}{h(X; \boldsymbol{\eta}^*)}\right\}, \quad A^{33} = \frac{1}{\Delta^*}\{\text{diag}(\boldsymbol{\lambda}^*) - \boldsymbol{\lambda}^*\boldsymbol{\lambda}^{*\top}\}, \quad V = \mathbf{C}\text{diag}(\boldsymbol{\iota} - \boldsymbol{\nu}^*)(\mathbf{0}_{m\times1}, \mathbf{I}_m)^\top\text{E}_0\{\text{diag}(\boldsymbol{\omega})X\}.\end{aligned}$$

References

[1] J.A. Anderson, Multivariate logistic compounds, Biometrika 66 (1979) 17–26.
[2] D. Böhning, M. Alfò, Editorial: Special Issue on Models for Continuous Data With a Spike at Zero, Biometrical J. 58 (2016) 255–258.
[3] E. Brunner, H. Dette, A. Munk, Box-type approximations in nonparametric factorial designs, J. Amer. Statist. Assoc. 92 (1997) 1494–1502.
[4] S. Cai, J. Chen, J.V. Zidek, Hypothesis test in the presence of multiple samples under density ratio models, Statist. Sinica 27 (2017) 761–783.
[5] J. Chen, S.-Y. Chen, J.N.K. Rao, Empirical likelihood confidence intervals for the mean of a population containing many zero values, Canad. J. Statist. 31 (2003) 53–68.
[6] S.X. Chen, J. Qin, Empirical likelihood-based confidence intervals for data with possible zero observations, Statist. Probab. Lett. 65 (2003) 29–37.
[7] Y.-H. Chen, X.-H. Zhou, Generalized Confidence Intervals for the Ratio or Difference of Two Means for Lognormal Populations with Zeros, Working Paper 296, University of Washington Biostatistics Working Paper Series. Unpublished manuscript, 2006.
[8] F. Critchley, P. Marriott, M. Salmon, On the differential geometry of the Wald test with nonlinear restrictions, Econometrica 64 (1996) 1213–1222.
[9] K. Fokianos, Density ratio model selection, J. Statist. Comput. Simul. 77 (2007) 805–819.
[10] S. Friedrich, F. Konietschke, M. Pauly, GFD: Tests for General Factorial Designs, R Package Version 0.1.4, 2016.
[11] Y. Fu, X. Wang, C. Wu, Weighted empirical likelihood inference for multiple samples, J. Stat. Plann. Inf. 139 (2009) 1462–1473.
[12] R.C. Gupta, X. Li, Statistical inference for the common mean of two log-normal distributions and some applications in reliability, Comput. Statist. Data Anal. 50 (2006) 3141–3164.
[13] L. Kang, A. Vexler, L. Tian, M. Cooney, G.M.B. Louis, Empirical and parametric likelihood interval estimation for populations with many zero values: application for assessing environmental chemical concentrations and reproductive health, Epidemiology 21 (2010) S58–S63.
[14] A. Keziou, S. Leoni-Aubin, On empirical likelihood for semiparametric two-sample density ratio models, J. Stat. Plann. Inf. 138 (2008) 915–928.
[15] L.H. Koopmans, Introduction to Contemporary Statistical Methods, Duxbury Press, Boston, MA, 1981.
[16] A. Marazzi, F. Paccaud, C. Ruffieux, C. Beguin, Fitting the distributions of length of stay by parametric models, Medical Care 36 (1998) 915–927.
[17] R.M. Nixon, S.G. Thompson, Parametric modelling of cost data in medical studies, Stat. Med. 23 (2004) 1311–1331.
[18] A.B. Owen, Empirical Likelihood, Chapman and Hall/CRC Press, New York, 2001.
[19] M. Pauly, E. Brunner, F. Konietschke, F. (2015). Asymptotic permutation tests in general factorial designs, J. R. Stat. Soc. Ser. B (Stat. Method.) 77 (2015) 461–473.
[20] M. Pennington, Efficient estimators of abundance, for fish and plankton surveys, Biometrics 39 (1983) 281–286.
[21] J. Qin, J.F. Lawless, Empirical likelihood and general estimating equations, Ann. Statist. 22 (1994) 300–325.
[22] J. Qin, B. Zhang, A goodness-of-fit test for logistic regression models based on case-control data, Biometrika 84 (1997) 609–618.
[23] J. Qin, H. Zhang, P. Li, D. Albanes, K. Yu, Using covariate-specific disease prevalence information to increase the power of case-control studies, Biometrika 102 (2015) 169–180.
[24] R Development Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, 2014.
[25] S. Taylor, K. Pollard, Hypothesis tests for point-mass mixture data with application to 'omics data with many zero values, Statistical Applications in Genetics and Molecular Biology 8 (2009) 1–43.
[26] M. Tsao, C. Wu, Empirical likelihood inference for a common mean in the presence of heteroscedasticity, Canad. J. Statist. 34 (2006) 45–59.
[27] W. Tu, X.-H. Zhou, A Wald test comparing medical costs based on log-normal distributions with zero valued costs, Stat. Med. 18 (1999) 2749–2761.
[28] G. Vallejo, M.P. Fernández, P.E. Livacic-Rojas, Analysis of unbalanced factorial designs with heteroscedastic data, J. Stat. Comput. Simul. 80 (2010) 75–88.
[29] S. Wan, B. Xu, B. Zhang, Using logistic regression for semiparametric comparison of population means and variances, Comm. Statist. Theory Meth. 45 (2016) 2485–2503.
[30] C. Wang, Empirical Likelihood and Bootstrap Inference with Constraints, Doctoral dissertation, University of Waterloo, Waterloo, ON, Canada, 2017.
[31] C. Wang, P. Marriott, P. Li, Testing homogeneity for multiple nonnegative distributions with excess zero observations, Comput. Statist. Data Anal. 114 (2017) 146–157.
[32] B.L. Welch, The significance of the difference between two means when the population variances are unequal, Biometrika 29 (1938) 350–362.
[33] C. Wu, Y. Yan, Empirical likelihood inference for two-sample problems, Statistics and Its Interface 5 (2012) 345–354.
[34] X.-H. Zhou, W. Tu, Comparison of several independent population means when their samples contain log-normal and possibly zero observations, Biometrics 55 (1999) 645–651.
[35] X.-H. Zhou, W. Tu, Interval estimation for the ratio in means of log-normally distributed medical costs with zero values, Comput. Statist. Data Anal. 35 (2000) 201–210.