Annals of the Institute of Statistical Mathematics (2022) 74:451–472
https://doi.org/10.1007/s10463-021-00804-4

Semiparametric inference on general functionals of two semicontinuous populations

Meng Yuan[1] · Chunlin Wang[2] · Boxi Lin[3] · Pengfei Li[1]

Received: 17 January 2021 / Revised: 11 June 2021 / Accepted: 14 June 2021 /
Published online: 5 July 2021

Abstract
In this paper, we propose new semiparametric procedures for inference on linear functionals in the context of two semicontinuous populations. The distribution of each semicontinuous population is characterized by a mixture of a discrete point mass at zero and a continuous skewed positive component. To utilize the information from both populations, we model the positive components of the two mixture distributions via a semiparametric density ratio model. Under this model setup, we construct the maximum empirical likelihood estimators of the linear functionals. The asymptotic normality of the proposed estimators is established and is used to construct confidence regions and perform hypothesis tests for these functionals. We show that the proposed estimators are more efficient than the fully nonparametric ones. Simulation studies demonstrate the advantages of our method over existing methods. Two real-data examples are provided for illustration.

Keywords Empirical likelihood · Density ratio model · Linear functional · Zero-excessive data

1 Introduction

Suppose that two independent samples are generated by the following mixture models:

✉ Chunlin Wang
wangc@xmu.edu.cn

[1] Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada

[2] Department of Statistics, School of Economics, Wang Yanan Institute for Studies in Economics, MOE Key Lab of Econometrics and Fujian Key Lab of Statistics, Xiamen University, 422 Siming South Road, Xiamen 361005, Fujian, China

[3] Dalla Lana School of Public Health, University of Toronto, 155 College Street, Toronto, ON M5T 3M7, Canada

$$X_{i1}, \ldots, X_{in_i} \sim F_i(x) = \nu_i I(x \geq 0) + (1 - \nu_i) I(x > 0) G_i(x), \quad \text{for } i = 0, 1, \tag{1}$$

where $\nu_i \in (0, 1)$, n_i is the sample size for the ith sample, $I(\cdot)$ is an indicator function, and the $G_i(\cdot)$'s are the cumulative distribution functions (CDFs) of the positive observations in the ith sample. We are interested in estimating linear functionals (Fernholz 1983, p. 6) of $F_0(x)$ and $F_1(x)$, defined as

$$\boldsymbol{\psi}_0 = \int_0^\infty \mathbf{a}(x) dF_0(x) \quad \text{and} \quad \boldsymbol{\psi}_1 = \int_0^\infty \mathbf{a}(x) dF_1(x) \tag{2}$$

for some given function $\mathbf{a}(x)$.

Many statistical applications naturally produce semicontinuous data with a mixture of excessive zero values and skewed positive outcomes. Examples include medical costs in public health research (Zhou and Tu 2000) and, in biological science, seasonal activity patterns for field mice (Koopmans 1981). More examples can be found in Wang et al. (2017) and in a special issue of the *Biometrical Journal* (Böhning and Alfò 2016) and the references therein. The functionals $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$ include the usual summary quantities like the centered and uncentered moments, and are widely used. For example, in public health research the mean ratio of two populations is a desirable summary quantity that characterizes the differences in medical costs between two groups (Zhou and Tu 2000). In business and economic studies, the moments and the generalized entropy class of inequality measures are important (Dufour et al. 2019).

Most existing procedures for inference on $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$ are either fully parametric or fully nonparametric. The parametric procedures are developed under a parametric model assumption, for example, a log-normal assumption, on G_i, $i = 0, 1$. Under this assumption, Tu and Zhou (1999) and Zhou and Tu (1999) developed a Wald-type test and a likelihood ratio test for the equality of two population means. Under the same assumption, Zhou and Tu (2000) proposed a maximum likelihood method and a two-stage bootstrap method to construct the confidence intervals (CIs) for the mean ratio; Chen and Zhou (2006) developed a set of approaches for constructing CIs for the mean ratio based on the generalized pivot and likelihood ratio statistic. The fully nonparametric methods usually first estimate $F_0(x)$ and $F_1(x)$ by the corresponding empirical CDFs, which are then used to construct the estimators for $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$. The asymptotic results for this type of estimator have been well studied in the literature; see Serfling (1980) for more details. The nonparametric Wald-type method (Brunner et al. 1997; Pauly et al. 2015; Dufour et al. 2019) and the empirical likelihood (EL) method (Kang et al. 2010; Wu and Yan 2012; Satter and Zhao 2021) were also used to construct CIs and perform hypothesis testing for $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$.

In general, the methods based on the parametric assumption on the G_i's are quite efficient. However, in many applications, this assumption, e.g., the log-normal assumption for G_i, may be violated. The corresponding parametric inference results may not be robust to model misspecification on the G_i's (Nixon and Thompson 2004). The fully nonparametric methods are generally quite robust to the model assumption on the G_i's. In the two-sample setting, the two populations may share certain characteristics. For example, the strengths of lumber produced in Canada in different years may follow

similar distributions (Chen and Liu 2013; Cai et al. 2017; Cai and Chen 2018). There is also a relationship between the distributions of biomarkers for the diagnosis of Duchenne muscular dystrophy in case and control groups (Yuan et al. 2021). The fully nonparametric methods, however, ignore such information.

In this paper, we propose new semiparametric procedures for estimating $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$ based on the semiparametric density ratio model (DRM; Anderson 1979; Qin 2017), which utilize the information from both populations effectively. Let $dG_i(x)$ be the probability density function of $G_i(x)$, $i = 0, 1$. The DRM links the two CDFs $G_0(x)$ and $G_1(x)$ in mixture model (1) via

$$dG_1(x) = \exp\{\alpha + \boldsymbol{\beta}^\top \mathbf{q}(x)\} dG_0(x) \tag{3}$$

for a prespecified, nontrivial, basis function $\mathbf{q}(x)$ of dimension d and unknown parameters α and $\boldsymbol{\beta}$. In the DRM (3), the baseline distribution $G_0(x)$ is not specified. Hence, the DRM is a semiparametric model and avoids distributional assumptions on $G_0(x)$ and $G_1(x)$. It is also quite flexible and has many important statistical models as special cases. For example, when $\boldsymbol{q}(x) = \log(x)$, the DRM includes the log-normal distribution of the same variance with respect to the log-scale, as well as the gamma distribution with the same scale parameters (Kay and Little 1987). Jiang and Tu (2012) observed that the DRM is actually broader than Cox proportional hazard models. It is also closely related to the well-studied logistic regression (Qin and Zhang 1997). Inference under the DRM can be converted to that under logistic regression (Wang et al. 2017).

The DRM has been proved to be a useful tool for inference when there is an excess of zeros in the data. Wang et al. (2017, 2018) developed the EL ratio (ELR) statistics for testing the homogeneity of distributions and the equality of population means, respectively. In the same setup, Lu et al. (2020) considered a test for the equality of the zero proportions and the equality of the means of two positive components jointly. Their simulation results showed that the proposed tests have great power advantages over existing nonparametric tests. To the best of our knowledge, semiparametric inference procedures such as point estimation and confidence regions for $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$ have not been explored under the mixture model (1) and the DRM (3). This paper aims to fill this gap.

Under the mixture model (1) and the DRM (3), we consider a class of general functionals $\boldsymbol{\psi}$ of dimension p, defined as

$$\boldsymbol{\psi} = \int_0^\infty \mathbf{u}(x; \boldsymbol{\nu}, \boldsymbol{\theta}) dG_0(x), \tag{4}$$

where $\boldsymbol{\nu} = (\nu_0, \nu_1)^\top$, $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top)^\top$, and $\mathbf{u}(x; \boldsymbol{\nu}, \boldsymbol{\theta}) = \big(u_1(x; \boldsymbol{\nu}, \boldsymbol{\theta}), \ldots, u_p(x; \boldsymbol{\nu}, \boldsymbol{\theta})\big)^\top$ is a given $(p \times 1)$-dimensional function. The parameters of interest are defined through $\mathbf{g}(\boldsymbol{\psi})$, where $\mathbf{g}(\cdot) : p \to q$ is a smooth function of $\boldsymbol{\psi}$. Note that $\boldsymbol{\psi}$ covers $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$, defined in (2), as special cases under $\mathbf{a}(0) = \mathbf{0}$, with

$$\mathbf{u}(X; \boldsymbol{\nu}, \boldsymbol{\theta}) = \begin{pmatrix} \mathbf{u}_0(X; \boldsymbol{\nu}, \boldsymbol{\theta}) \\ \mathbf{u}_1(X; \boldsymbol{\nu}, \boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} (1 - \nu_0)\mathbf{a}(x) \\ (1 - \nu_1)\mathbf{a}(x) \exp\{\alpha + \boldsymbol{\beta}^\top \mathbf{q}(x)\} \end{pmatrix}. \tag{5}$$

The parameters $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$ together cover many important summary quantities; see Sect. 2.4 for examples. Following Owen (2001), we construct the maximum EL estimator (MELE) of $\boldsymbol{\psi}$. We also establish the asymptotic normality of the MELE of $\boldsymbol{\psi}$. These results enable us to construct confidence regions for $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$ and perform hypothesis testing on $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$. We apply the results for general $\boldsymbol{\psi}$ to $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$, and then show that the asymptotic variances of the MELEs of $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$ are smaller than or equal to those of nonparametric estimators of $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$.

The rest of this paper is organized as follows. In Sect. 2, we study the asymptotic properties of the MELE of $(\boldsymbol{\nu}, \boldsymbol{\theta})$ as well as the MELE of $\boldsymbol{\psi}$. We further provide examples for $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$ which cover several important summary quantities. Simulation results are presented in Sect. 3, and two real-data applications are given in Sect. 4. We conclude the paper with some discussion in Sect. 5. All the technical details are provided in the supplementary material.

2 Main results

Let n_{i0} and n_{i1} be the (random) numbers of zero observations and positive observations, respectively, in each sample $i = 0, 1$. Clearly, $n_i = n_{i0} + n_{i1}$, for $i = 0, 1$. Without loss of generality, we assume that the first n_{i1} observations in group i, $X_{i1}, \ldots, X_{in_{i1}}$, are positive, and the remaining n_{i0} observations are 0. Let n be the total (fixed) sample size, i.e., $n = n_0 + n_1$.

2.1 Point estimation of $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$

We first discuss the maximum EL procedure for estimating the unknown parameters and functions in models (1) and (3).

With the two samples of observations from model (1), the full likelihood function is

$$\mathcal{L}_n = \prod_{i=0}^{1} \left\{ \nu_i^{n_{i0}} (1 - \nu_i)^{n_{i1}} \prod_{j=1}^{n_{i1}} dG_i(X_{ij}) \right\}.$$

Following the EL principle (Owen 2001), we model the baseline distribution $G_0(x)$ as

$$G_0(x) = \sum_{i=0}^{1} \sum_{j=1}^{n_{i1}} p_{ij} I(X_{ij} \le x), \tag{6}$$

where $p_{ij} = dG_0(X_{ij})$ for $i = 0, 1$ and $j = 1, \ldots, n_{i1}$. With (6) and under the DRM (3), the full likelihood function can be rewritten as

$$\mathcal{L}_n = \prod_{i=0}^{1} \nu_i^{n_{i0}} (1 - \nu_i)^{n_{i1}} \cdot \left\{ \prod_{i=0}^{1} \prod_{j=1}^{n_{i1}} p_{ij} \right\} \cdot \left[\prod_{j=1}^{n_{11}} \exp\left\{ \alpha + \boldsymbol{\beta}^\top \mathbf{q}(X_{1j}) \right\} \right],$$

where the p_{ij}'s satisfy the constraints

$$p_{ij} > 0, \quad \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} p_{ij} = 1, \quad \text{and} \quad \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} p_{ij}\exp\left\{\alpha + \boldsymbol{\beta}^{\top}\mathbf{q}(X_{ij})\right\} = 1. \tag{7}$$

These constraints ensure that $G_0(x)$ and $G_1(x)$ are CDFs.

Let $\boldsymbol{P} = \{p_{ij}\}$. The MELE of $(\boldsymbol{\nu}, \boldsymbol{\theta}, \boldsymbol{P})$ is then defined as

$$(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{P}}) = \arg\max_{\boldsymbol{\nu},\boldsymbol{\theta},\boldsymbol{P}} \mathcal{L}_n$$

subject to the constraints in (7). We write the logarithm of the EL function $\mathcal{L}_n$ as

$$\tilde{\ell}(\boldsymbol{\nu}, \boldsymbol{\theta}, G_0) = \ell_0(\boldsymbol{\nu}) + \tilde{\ell}_1(\boldsymbol{\theta}, \boldsymbol{P}),$$

where

$$\ell_0(\boldsymbol{\nu}) = \sum_{i=0}^{1}\log\left\{\nu_i^{n_{i0}}\left(1-\nu_i\right)^{n_{i1}}\right\} \text{ and } \tilde{\ell}_1(\boldsymbol{\theta}, \boldsymbol{P}) = \sum_{j=1}^{n_{11}}\left\{\alpha + \boldsymbol{\beta}^{\top}\mathbf{q}(X_{1j})\right\} + \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\log p_{ij}.$$

Here $\ell_0(\boldsymbol{\nu})$ is the binomial log-likelihood function corresponding to the number of zero observations, and $\tilde{\ell}_1(\boldsymbol{\theta}, \boldsymbol{P})$ represents the empirical log-likelihood function associated with the positive observations.

Following Wang et al. (2017), we have $\hat{\boldsymbol{\nu}} = \arg\max_{\boldsymbol{\nu}} \ell_0(\boldsymbol{\nu})$ and

$$(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{P}}) = \arg\max_{\boldsymbol{\theta},\boldsymbol{P}} \left\{\tilde{\ell}_1(\boldsymbol{\theta}, \boldsymbol{P}) : \; p_{ij} > 0, \; \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} p_{ij} = 1, \; \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} p_{ij}\exp\left\{\alpha + \boldsymbol{\beta}^{\top}\mathbf{q}(X_{ij})\right\} = 1\right\}.$$

By the method of Lagrange multipliers, $\hat{\boldsymbol{\theta}}$ can be obtained by maximizing the following dual empirical log-likelihood function (Cai et al. 2017):

$$\ell_1(\boldsymbol{\theta}) = -\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\log\left\{1 + \hat{\rho}[\exp\{\alpha + \boldsymbol{\beta}^{\top}\mathbf{q}(X_{ij})\} - 1]\right\} + \sum_{j=1}^{n_{11}}\{\alpha + \boldsymbol{\beta}^{\top}\mathbf{q}(X_{1j})\},$$

where $\hat{\rho} = n_{11}\{n_{01} + n_{11}\}^{-1}$. That is, $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell_1(\boldsymbol{\theta})$. Note that $\hat{\rho}$ is a random variable in our setup. This is fundamentally different from the case where there is no excess of zeros in the data (Qin and Zhang 1997), and it creates theoretical challenges for our asymptotic development in the next section.

Once $\hat{\boldsymbol{\theta}}$ is obtained, the MELEs of the $\hat{p}_{ij}$'s are (Wang et al. 2017)

$$\hat{p}_{ij} = \{n_{01} + n_{11}\}^{-1}\left\{1 + \hat{\rho}[\exp\{\hat{\alpha} + \hat{\boldsymbol{\beta}}^{\top}\mathbf{q}(X_{ij})\} - 1]\right\}^{-1},$$

and the MELEs of $G_0(x)$ and $G_1(x)$ are

$$\hat{G}_0(x) = \sum_{i=0}^{1} \sum_{j=1}^{n_{i1}} \hat{p}_{ij} I(X_{ij} \le x) \text{ and}$$

$$\hat{G}_1(x) = \sum_{i=0}^{1} \sum_{j=1}^{n_{i1}} \hat{p}_{ij} \exp\{\hat{\alpha} + \hat{\boldsymbol{\beta}}^\top \mathbf{q}(X_{ij})\} I(X_{ij} \le x).$$

By the definition of $\boldsymbol{\psi}$ in (4), $\boldsymbol{\psi}$ is a function of $(\boldsymbol{\nu}, \boldsymbol{\theta})$ and G_0. Replacing them with $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}})$ and $\hat{G}_0$, the MELE of $\boldsymbol{\psi}$ is

$$\hat{\boldsymbol{\psi}} = \sum_{i=0}^{1} \sum_{j=1}^{n_{i1}} \hat{p}_{ij} \mathbf{u}(X_{ij}; \hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\theta}}),$$

and the estimator of $\mathbf{g}(\boldsymbol{\psi})$ is $\mathbf{g}(\hat{\boldsymbol{\psi}})$.

When $\mathbf{u}(x; \boldsymbol{\nu}, \boldsymbol{\theta})$ takes the specific form of (5), we obtain the MELEs of $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$, defined in (2), as

$$\begin{aligned} \hat{\boldsymbol{\psi}}_0 &= \sum_{i=0}^{1} \sum_{j=1}^{n_{i1}} \hat{p}_{ij} (1 - \hat{\nu}_0) \mathbf{a}(X_{ij}) \text{ and} \\ \hat{\boldsymbol{\psi}}_1 &= \sum_{i=0}^{1} \sum_{j=1}^{n_{i1}} \hat{p}_{ij} (1 - \hat{\nu}_1) \mathbf{a}(X_{ij}) \exp\{\hat{\alpha} + \hat{\boldsymbol{\beta}}^\top \mathbf{q}(x)\}. \end{aligned} \tag{8}$$

2.2 Asymptotic properties

In this section, we first study the asymptotic properties of $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\nu}}^\top, \hat{\rho}, \hat{\boldsymbol{\theta}}^\top)^\top$ and then apply these results to establish the asymptotic properties of $\hat{\boldsymbol{\psi}}$ and $\mathbf{g}(\hat{\boldsymbol{\psi}})$. Without loss of generality, we assume that $\mathbf{a}(0) = \mathbf{0}$ throughout the paper; this assumption is satisfied by all the examples considered in Sect. 2.4.

For ease of presentation, we introduce some notation. We use $\boldsymbol{\nu}^*$ and $\boldsymbol{\theta}^*$ to denote the true values of $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$, respectively. Let $\mathbf{Q}(x) = (1, \mathbf{q}(x)^\top)^\top$ and

$$w = n_0/n,\ \Delta^* = w(1 - \nu_0^*) + (1 - w)(1 - \nu_1^*),\ \rho^* = \frac{(1 - w)(1 - \nu_1^*)}{\Delta^*},$$

$$\omega(x) = \exp\{\boldsymbol{\theta}^{*\top} \mathbf{Q}(x)\},\ h(x) = 1 + \rho^* \{\omega(x) - 1\},\ h_1(x) = \rho^* \omega(x)/h(x),$$

$$\mathbf{A}_{\boldsymbol{\nu}} = \operatorname{diag}\left\{ \frac{w}{\nu_0^*(1 - \nu_0^*)}, \frac{1 - w}{\nu_1^*(1 - \nu_1^*)} \right\},\ \mathbf{A}_{\boldsymbol{\theta}} = \Delta^*(1 - \rho^*) E_0\{h_1(X) \mathbf{Q}(X) \mathbf{Q}(X)^\top\},$$

where $E_0(\cdot)$ represents the expectation operator with respect to G_0 and X refers to a random variable from G_0. Note that although $\omega(\cdot)$, $h(\cdot)$, and $h_1(\cdot)$ also depend on $\boldsymbol{\theta}^*$ and/or ρ^*, we drop these redundant parameters for notational simplicity.

The asymptotic results in this section are developed under the following regularity conditions:

C1: The true value ν_i^* satisfies $0 < \nu_i^* < 1$ for $i = 0, 1$.

C2: As the total sample size n goes to infinity, $n_0/n = w$ for some constant $w \in (0, 1)$.
C3: The components of $\mathbf{Q}(x)$ are continuous and stochastically linearly independent.
C4: $\int_0^\infty \exp\{\boldsymbol{\beta}^\top \mathbf{q}(x)\} dG_0(x) < \infty$ for all $\boldsymbol{\beta}$ in a neighborhood of the true value $\boldsymbol{\beta}^*$.

Condition C1 ensures that the binomial likelihood $\ell_0(\boldsymbol{\nu})$ has regular properties. Condition C2 means that both n_0 and n_1 go to the infinity at the same rate. Conditions C1 and C2 imply that $\mathbf{A}_{\boldsymbol{\nu}}$ is positive definite. Condition C3 ensures that no linear combinations of any components of $\mathbf{Q}(x)$ can be 0 with probability 1 under G_0. Condition C4 guarantees the existence of finite moments of $\mathbf{q}(X)$ in a neighborhood of $\boldsymbol{\beta}^*$ under both $G_0(x)$ and $G_1(x)$. Conditions C3 and C4 together imply that $\mathbf{A}_{\boldsymbol{\theta}}$ is positive definite.

The following theorem establishes the asymptotic normality of $\hat{\boldsymbol{\eta}}$.

Theorem 1 *Let $\boldsymbol{\eta}^* = (\boldsymbol{\nu}^{*\top}, \rho^*, \boldsymbol{\theta}^{*\top})^\top$. Assume that Conditions* C1–C4 *are satisfied. As $n \to \infty$,*

$$n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Lambda}),$$

where $\xrightarrow{d}$ denotes "convergence in distribution" and

$$\boldsymbol{\Lambda} = \begin{pmatrix} \mathbf{A}_{\boldsymbol{\nu}}^{-1} & \rho^*(1-\rho^*)\mathbf{A}_{\boldsymbol{\nu}}^{-1}\mathbf{W}^\top & \mathbf{0} \\ \rho^*(1-\rho^*)\mathbf{W}\mathbf{A}_{\boldsymbol{\nu}}^{-1} & (\Delta^*)^{-1}\rho^*(1-\rho^*)\{\rho^* v_0^* + (1-\rho^*)v_1^*\} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)} \end{pmatrix}$$

with $\mathbf{W} = \left((1-v_0^*)^{-1}, -(1-v_1^*)^{-1}\right)$ and $\boldsymbol{e} = (1, \mathbf{0}_{d\times 1}^\top)^\top$.

Qin and Zhang (1997) considered the asymptotic normality of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ when there is no excess of zeros in the data. Theorem 1 generalizes their results to the case where the data contain excessive zeros. Furthermore, it establishes the joint limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$, $\sqrt{n}(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*)$, and $\sqrt{n}(\hat{\rho} - \rho^*)$, where the latter two are induced by the semicontinuous data structure. This joint limiting distribution plays an important role in deriving the asymptotic normality of $\hat{\boldsymbol{\psi}}$ in the following theorem.

Theorem 2 *Let $\boldsymbol{\psi}^*$ be the true value of $\boldsymbol{\psi}$. Under the conditions of Theorem 1, as $n \to \infty$,*

(a) $\sqrt{n}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma})$, *where*

$$\boldsymbol{\Gamma} = \frac{1}{\Delta^*} E_0\left\{ \frac{\mathbf{u}(X;\boldsymbol{\nu}^*, \boldsymbol{\theta}^*)\mathbf{u}(X;\boldsymbol{\nu}^*, \boldsymbol{\theta}^*)^\top}{h(X)} \right\} - \frac{\boldsymbol{\psi}^*\boldsymbol{\psi}^{*\top}}{\Delta^*} + \mathcal{M}_1\mathbf{A}_{\boldsymbol{\nu}}^{-1}\mathcal{M}_1^\top - \frac{\mathcal{M}_2\mathcal{M}_2^\top}{\Delta^*\rho^*(1-\rho^*)} + \mathcal{M}_3\mathbf{A}_{\boldsymbol{\theta}}^{-1}\mathcal{M}_3^\top,$$

with

$$\mathcal{M}_1 = E_0\left\{\frac{\partial \mathbf{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)}{\partial \boldsymbol{\nu}}\right\},$$
$$\mathcal{M}_2 = E_0\left[\left\{\partial \mathbf{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}\right\}\boldsymbol{e}\right] - \rho^*\boldsymbol{\psi}^*,$$
$$\mathcal{M}_3 = E_0\left\{\partial \mathbf{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta} - h_1(X)\mathbf{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)\mathbf{Q}(X)^\top\right\};$$

(b) *for some smooth function* $\mathbf{g}(\cdot) : p \to q$, $\sqrt{n}\{\mathbf{g}(\hat{\boldsymbol{\psi}}) - \mathbf{g}(\boldsymbol{\psi}^*)\} \xrightarrow{d} N\big(0, \boldsymbol{\Gamma}_{\mathbf{g}}\big)$, *where*

$$\boldsymbol{\Gamma}_{\mathbf{g}} = \left\{\frac{\partial \mathbf{g}(\boldsymbol{\psi}^*)}{\partial \boldsymbol{\psi}}\right\}\boldsymbol{\Gamma}\left\{\frac{\partial \mathbf{g}(\boldsymbol{\psi}^*)}{\partial \boldsymbol{\psi}}\right\}^\top.$$

Li et al. (2018) derived a similar result in their Theorem 2.1 for $\hat{\boldsymbol{\psi}}$ when there is no excess of zeros in the data and $p = 1$. Theorem 2 covers the case with excessive zeros. The two results complement each other to cover both cases.

We now apply the results for $\hat{\boldsymbol{\psi}}$ in Theorem 2 to $\hat{\boldsymbol{\psi}}_0$ and $\hat{\boldsymbol{\psi}}_1$ in (8), and then compare them with the fully nonparametric estimators $\tilde{\boldsymbol{\psi}}_0$ and $\tilde{\boldsymbol{\psi}}_1$:

$$\tilde{\boldsymbol{\psi}}_0 = \frac{1}{n_0}\sum_{j=1}^{n_0}\mathbf{a}(X_{0j}) \quad \text{and} \quad \tilde{\boldsymbol{\psi}}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1}\mathbf{a}(X_{1j}).$$

For $i = 0, 1$, let

$$\mathbf{V}_i = \int_0^\infty \mathbf{a}(x)\{\mathbf{a}(x)\}^\top dF_i(x) - \int_0^\infty \mathbf{a}(x)dF_i(x)\int_0^\infty \{\mathbf{a}(x)\}^\top dF_i(x).$$

Then, $\sqrt{n}\big(\tilde{\boldsymbol{\psi}}_0^\top - \boldsymbol{\psi}_0^\top, \tilde{\boldsymbol{\psi}}_1^\top - \boldsymbol{\psi}_1^\top\big)^\top$ has the asymptotic variance–covariance matrix

$$\boldsymbol{\Gamma}_{non} = \begin{pmatrix} w^{-1}\mathbf{V}_0 & \mathbf{0} \\ \mathbf{0} & (1-w)^{-1}\mathbf{V}_1 \end{pmatrix}.$$

In comparison with the asymptotic variance of the MELEs $\hat{\boldsymbol{\psi}}_0$ and $\hat{\boldsymbol{\psi}}_1$ given in (8), we have the following results.

Corollary 1 *Under the conditions of Theorem 1, as* $n \to \infty$,

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\psi}}_0 - \boldsymbol{\psi}_0 \\ \hat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_1 \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_{sem}),$$

where

$$\boldsymbol{\Gamma}_{sem} = \boldsymbol{\Gamma}_{non} - \Delta^*(1-\rho^*)E_0\left\{h_1(X)\begin{pmatrix} w^{-1}\mathbf{d}(X) \\ -(1-w)^{-1}\mathbf{d}(X) \end{pmatrix}\begin{pmatrix} w^{-1}\mathbf{d}(X) \\ -(1-w)^{-1}\mathbf{d}(X) \end{pmatrix}^\top\right\},$$

with

$$\mathbf{d}(X) = \mathbf{a}(X) - \Delta^*(1-\rho^*)E_0\left\{h_1(X)\mathbf{a}(X)\mathbf{Q}(X)^\top\right\}\mathbf{A}_{\boldsymbol{\theta}}^{-1}\mathbf{Q}(X).$$

Corollary 1 implies that $\mathbf{\Gamma}_{non} - \mathbf{\Gamma}_{sem}$ is positive semidefinite. Hence, the proposed MELEs of $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$ are more efficient than the corresponding nonparametric ones. The simulation studies in Sect. 3 confirm this property.

2.3 Confidence regions and hypothesis tests for $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$

The two variance–covariance matrices $\mathbf{\Gamma}$ and $\mathbf{\Gamma}_{\mathbf{g}}$ may depend on $\boldsymbol{\psi}^*$ and G_0. Replacing them by $\hat{\boldsymbol{\psi}}$ and $\hat{G}_0$, we get the corresponding estimators $\hat{\mathbf{\Gamma}}$ and $\hat{\mathbf{\Gamma}}_{\mathbf{g}}$. With the results of Theorem 1, it can easily be shown that both $\hat{\mathbf{\Gamma}}$ and $\hat{\mathbf{\Gamma}}_{\mathbf{g}}$ are consistent; the details are omitted.

Theorem 3 *Under the conditions of Theorem 1, as* $n \to \infty$, $\hat{\mathbf{\Gamma}} \xrightarrow{p} \mathbf{\Gamma}$ *and* $\hat{\mathbf{\Gamma}}_{\mathbf{g}} \xrightarrow{p} \mathbf{\Gamma}_{\mathbf{g}}$, *where* $\xrightarrow{p}$ *denotes "convergence in probability."*

Theorems 2 and 3 together imply that, as $n \to \infty$,

$$n(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}^*)^\top \hat{\mathbf{\Gamma}}^{-1}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}^*) \quad \text{and} \quad n\{\mathbf{g}(\hat{\boldsymbol{\psi}}) - \mathbf{g}(\boldsymbol{\psi}^*)\}^\top \hat{\mathbf{\Gamma}}_{\mathbf{g}}^{-1}\{\mathbf{g}(\hat{\boldsymbol{\psi}}) - \mathbf{g}(\boldsymbol{\psi}^*)\}$$

converge in distribution to χ_p^2 and χ_q^2, respectively. Hence, both of them are asymptotically pivotal and can be used to construct Wald-type confidence regions for $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$ and perform hypothesis tests on $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$. For illustration, we consider the case where the dimension q of $\mathbf{g}(\cdot)$ is 1, which is perhaps the most common situation in applications. Let $\phi = \mathbf{g}(\boldsymbol{\psi})$. Next, we explain how to apply the results to construct a $100(1-\gamma)\%$ CI for ϕ and perform the hypothesis test for $H_0 : \phi = 0$. For general $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$, similar procedures are available.

Let $\hat{\phi} = \mathbf{g}(\hat{\boldsymbol{\psi}})$ and $\hat{\sigma}_\phi^2 = \hat{\mathbf{\Gamma}}_{\mathbf{g}}$. Then, a $100(1-\gamma)\%$ CI for ϕ is

$$\mathcal{I}_\phi = \left\{ \phi : n(\hat{\phi} - \phi)^2/\hat{\sigma}_\phi^2 \le \chi_{1,\gamma}^2 \right\} = \left[\hat{\phi} - z_{\gamma/2}\hat{\sigma}_\phi/\sqrt{n}, \hat{\phi} + z_{\gamma/2}\hat{\sigma_\phi}/\sqrt{n} \right], \quad (9)$$

where $\chi_{1,\gamma}^2$ and $z_{\gamma/2}$ denote the $(1-\gamma)$ quantile of the χ_1^2 distribution and the $(1-\gamma/2)$ quantile of the $N(0, 1)$ distribution, respectively. When testing $H_0 : \phi = 0$, we reject the null hypothesis if

$$n\hat{\phi}^2/\hat{\sigma}_\phi^2 > \chi_{1,\gamma}^2 \text{ or equivalently } |\sqrt{n}\hat{\phi}/\hat{\sigma}_\phi| > z_{\gamma/2}, \quad (10)$$

for the given significance level γ.

2.4 Examples of $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$

In this section, we provide some examples to demonstrate that $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$ cover many important summary quantities. The proposed methods and the general results in Sections 2.1–2.3 can readily be applied to these quantities.

Example 1 (Uncentered moments) Let $\mu_i^{(k)} = \int_0^\infty x^k dF_i(x)$ be the kth (uncentered) moments of $F_i(x)$, $i = 0, 1$. When $k = 1$, we write $\mu_i = \mu_i^{(1)}$. Clearly, if

$$u_1(x;\boldsymbol{\nu}, \boldsymbol{\theta}) = (1 - \nu_0)x^k \text{ and } u_2(x;\boldsymbol{\nu}, \boldsymbol{\theta}) = (1 - \nu_1)x^k \exp\{\alpha + \boldsymbol{\beta}^\top \mathbf{q}(x)\}, \quad (11)$$

then $\boldsymbol{\psi} = (\mu_0^{(k)}, \mu_1^{(k)})^\top$.

Example 2 (Mean ratio) Let $\delta = \mu_1/\mu_0$ denote the mean ratio of two populations. Setting $k = 1$ in (11), we obtain $\boldsymbol{\psi} = (\mu_0, \mu_1)^\top$. Further, let $g(x_1, x_2) = x_2/x_1$, then we get $\delta = g(\boldsymbol{\psi})$. We can directly construct a CI for δ using the result given in (9).

An alternative way is to consider $g(x_1, x_2) = \log(x_2) - \log(x_1)$; then $g(\boldsymbol{\psi}) = \log \delta$. We can use the form of (9) to first construct a CI for $\log \delta$ and then transform it to a CI for δ. Our simulation indicates that this approach leads to a CI with better coverage accuracy.

Example 3 (Centered moments) Let $C_i^{(k)} = \int_0^\infty (x - \mu_i)^k dF_i(x)$ be the kth centered moments of $F_i(x)$, $i = 0, 1$. When $k = 2$, we write $\sigma_i^2 = C_i^{(2)}$. As demonstrated in Serfling (1980), centered moments $C_i^{(k)}$ can be written as functions of $\mu_i^{(1)}, \ldots, \mu_i^{(k)}$. For illustration, we concentrate on $k = 2$ and consider the variances of the two populations, σ_0^2 and σ_1^2. Let

$$\mathbf{u}(x;\boldsymbol{\nu}, \boldsymbol{\theta}) = \left((1 - \nu_0)x, (1 - \nu_0)x^2, (1 - \nu_1)x \exp\{\alpha + \boldsymbol{\beta}^\top \mathbf{q}(x)\}, (1 - \nu_1)x^2 \exp\{\alpha + \boldsymbol{\beta}^\top \mathbf{q}(x)\}\right)^\top.$$

Then, $\boldsymbol{\psi} = (\mu_0, \mu_0^{(2)}, \mu_1, \mu_1^{(2)})^\top$. Define $\mathbf{g}(\cdot)$ as

$$\mathbf{g}(x_1, x_2, x_3, x_4) = (x_2 - x_1^2, x_4 - x_3^2)^\top.$$

We have $\mathbf{g}(\boldsymbol{\psi}) = (\sigma_0^2, \sigma_1^2)^\top$. The results of Theorem 2 can be used to obtain the joint limiting distribution of $\sqrt{n}(\hat{\sigma}_0^2 - \sigma_0^2, \hat{\sigma}_1^2 - \sigma_1^2)^\top$, where $\hat{\sigma}_0^2$ and $\hat{\sigma}_1^2$ are the MELEs of σ_0^2 and σ_1^2, respectively. If we choose

$$\mathbf{g}(x_1, x_2, x_3, x_4) = (x_4 - x_3^2) - (x_2 - x_1^2),$$

then $\mathbf{g}(\boldsymbol{\psi}) = \sigma_1^2 - \sigma_0^2$, and the procedure described in (10) can be used to test $H_0 : \sigma_0^2 = \sigma_1^2$.

Example 4 (Coefficient of variation) Let $CV_i = \sigma_i/\mu_i$ be the coefficient of variation of the ith population, $i = 0, 1$. If we choose

$$\mathbf{g}(x_1, x_2, x_3, x_4) = (\sqrt{x_2}/x_1, \sqrt{x_4}/x_3)^\top,$$

then $\mathbf{g}(\boldsymbol{\psi}) = (CV_0, CV_1)^\top$. If we choose $\mathbf{g}(x_1, x_2, x_3, x_4) = \sqrt{x_4}/x_3 - \sqrt{x_2}/x_1$, then $\mathbf{g}(\boldsymbol{\psi}) = CV_1 - CV_0$, and the procedure described in (10) can be used to test $H_0 : CV_0 = CV_1$.

Example 5 (Generalized entropy class of inequality measures) Let

$$GE_i^{(\xi)} = \begin{cases} \frac{1}{\xi^2-\xi}\left\{\int_0^\infty \left(\frac{x}{\mu_i}\right)^\xi dF_i(x) - 1\right\}, & \text{if } \xi \neq 0, 1, \\ -\int_0^\infty \log\left(\frac{x}{\mu_i}\right) dF_i(x), & \text{if } \xi = 0, \\ \int_0^\infty \frac{x}{\mu_i}\log\left(\frac{x}{\mu_i}\right) dF_i(x), & \text{if } \xi = 1 \end{cases}$$

be the generalized entropy class of inequality measures of the ith population, $i = 0, 1$. Note that the $GE_i^{(\xi)}$ are not well defined for the population with excessive zeros when $\xi = 0$. In our setup, $(GE_0^{(\xi)}, GE_1^{(\xi)})^\top$ can also be written as $\mathbf{g}(\boldsymbol{\psi})$ with certain $\mathbf{u}(\cdot)$ and $\mathbf{g}(\cdot)$ functions provided $\xi \neq 0$. For illustration, we consider $\xi = 1$. Let

$$\mathbf{u}(x;\boldsymbol{\nu},\boldsymbol{\theta}) = \Big((1-\nu_0)x, (1-\nu_0)x\log(x), (1-\nu_1)x\exp\{\alpha + \boldsymbol{\beta}^\top\mathbf{q}(x)\}, (1-\nu_1)x\log(x)\exp\{\alpha + \boldsymbol{\beta}^\top\mathbf{q}(x)\}\Big)^\top$$

and

$$\mathbf{g}(x_1, x_2, x_3, x_4) = (x_2/x_1 - \log x_1, x_4/x_3 - \log x_3)^\top.$$

Then $\mathbf{g}(\boldsymbol{\psi}) = (GE_0^{(1)}, GE_1^{(1)})^\top$. Similarly to Examples 3 and 4, we can choose an appropriate $\mathbf{g}(\cdot)$ function to construct a testing procedure for $H_0 : GE_0^{(1)} = GE_1^{(1)}$.

3 Simulation study

In this section, we conduct simulations to compare the finite-sample performance of the proposed estimators and CIs with existing methods. We consider three parameters, the mean ratio δ, discussed in Example 2, and the population variances σ_0^2 and σ_1^2, discussed in Example 3, for the performance comparison of the point estimators. For comparison of the CIs, we mainly focus on the mean ratio δ.

3.1 Simulation setup

In our simulations, the random observations are generated from the mixture model (1), with G_i being the log-normal distribution. We use the log-normal distribution because it has positive support and is highly skewed to the right. These properties allow us to check that the proposed method is applicable to skewed data, which often occur in reality. We use $\mathcal{LN}(a, b)$ to denote the log-normal distribution, where a and b are, respectively, the mean and variance in the log scale. Table 1 gives the parameter settings for the simulation studies.

For all the models listed in Table 1, the DRM (3) is satisfied with $\boldsymbol{q}(x) = \log x$. For each model, we consider four combinations of sample sizes (n_0, n_1): (50, 50), (100, 100), (50, 150), and (150, 50). The number of replications is 10,000 for each configuration of the parameter settings.

3.2 Comparison of point estimators

We first study the finite-sample performance of the point estimators. Under model (1) and the DRM (3), our estimators for δ, σ_0^2, and σ_1^2 are

$$\hat{\delta} = \frac{\hat{\mu}_1}{\hat{\mu}_0},\ \hat{\sigma}_0^2 = (1-\hat{\nu}_0)\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\hat{p}_{ij}X_{ij}^2 - \hat{\mu}_0^2,$$

and

$$\hat{\sigma}_1^2 = (1-\hat{\nu}_1)\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\hat{p}_{ij}\exp\left\{\hat{\alpha}+\hat{\boldsymbol{\beta}}^\top\mathbf{q}(X_{ij})\right\}X_{ij}^2 - \hat{\mu}_1^2,$$

with

$$\hat{\mu}_0 = (1-\hat{\nu}_0)\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\hat{p}_{ij}X_{ij} \text{ and } \hat{\mu}_1 = (1-\hat{\nu}_1)\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\hat{p}_{ij}\exp\left\{\hat{\alpha}+\hat{\boldsymbol{\beta}}^\top\mathbf{q}(X_{ij})\right\}X_{ij}.$$

We compare $\hat{\delta}$, $\hat{\sigma}_0^2$, and $\hat{\sigma}_1^2$ with the fully nonparametric estimators

$$\tilde{\delta} = \frac{\tilde{\mu}_1}{\tilde{\mu}_0},\ \tilde{\sigma}_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij}-\tilde{\mu}_i)^2 \text{ with } \tilde{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}X_{ij},$$

for $i = 0, 1$.

Table 2 presents the bias and mean square error (MSE) of these estimators. It shows that the biases of $\hat{\delta}$ and $\tilde{\delta}$ are consistently quite negligible, and $\hat{\delta}$ usually has a smaller bias. Moreover, $\hat{\delta}$ outperforms $\tilde{\delta}$ in terms of the MSE; this is expected since $\hat{\delta}$ uses more information to estimate the population means μ_0 and μ_1. The biases of $(\hat{\sigma}_0^2, \hat{\sigma}_1^2)$ and $(\tilde{\sigma}_0^2, \tilde{\sigma}_1^2)$ are quite small, and the MSEs of $(\hat{\sigma}_0^2, \hat{\sigma}_1^2)$ are significantly smaller than those of $(\tilde{\sigma}_0^2, \tilde{\sigma}_1^2)$. In some settings, e.g., Model 8 with

Table 1 Parameter settings for simulation studies: $G_0 = \mathcal{LN}(a_0, b_0)$ and $G_1 = \mathcal{LN}(a_1, b_1)$

Model	(ν_0, ν_1)	(a_0, a_1)	(b_0, b_1)	(μ_0, μ_1)	(σ_0^2, σ_1^2)	δ
1	(0.30, 0.30)	(0.00, 0.00)	(1.00, 1.00)	(1.15, 1.15)	(3.84, 3.84)	1.00
2	(0.70, 0.70)	(0.00, 0.00)	(1.00, 1.00)	(0.49, 0.49)	(1.97, 1.97)	1.00
3	(0.30, 0.50)	(0.33, 0.66)	(1.00, 1.00)	(1.61, 1.59)	(7.43, 11.29)	0.99
4	(0.50, 0.70)	(0.37, 0.89)	(1.00, 1.00)	(1.19, 1.20)	(6.32, 11.69)	1.01
5	(0.50, 0.30)	(0.00, 0.00)	(1.00, 1.00)	(0.82, 1.15)	(3.02, 3.84)	1.40
6	(0.70, 0.50)	(0.00, 0.00)	(1.00, 1.00)	(0.49, 0.82)	(1.97, 3.02)	1.67
7	(0.60, 0.40)	(0.00, 0.00)	(1.00, 1.00)	(0.66, 0.99)	(2.52, 3.45)	1.50
8	(0.30, 0.30)	(0.00, 0.50)	(1.00, 1.00)	(1.15, 1.90)	(3.84, 10.44)	1.65
9	(0.70, 0.70)	(0.00, 0.75)	(1.00, 1.00)	(0.49, 1.05)	(1.97, 8.84)	2.12
10	(0.40, 0.60)	(0.00, 1.00)	(1.00, 1.00)	(0.99, 1.79)	(3.45, 18.63)	1.81

sample sizes $(n_0, n_1) = (100, 100)$, the MSE of $\hat{\sigma}_0^2$ is less than 20% of the MSE of $\tilde{\sigma}_0^2$.

3.3 Comparison of confidence intervals

We now examine the finite-sample behavior of the following 95% CIs of the mean ratio δ:

- $\mathcal{I}_1$: Wald-type CI based on $\log \tilde{\delta}$ using the quantile of $N(0, 1)$;
- $\mathcal{I}_{1B}$: bootstrap Wald-type CI based on $\log \tilde{\delta}$ using the quantile from the nonparametric bootstrap method;
- $\mathcal{I}_2$: ELR-based CI using the quantile of the χ_1^2 distribution (Wu and Yan 2012);
- $\mathcal{I}_{2B}$: bootstrap ELR-based CI using the quantile from the nonparametric bootstrap method (Wu and Yan 2012);
- $\mathcal{I}_3$: ELR-based CI under the DRM (3) using the quantile of the χ_1^2 distribution (Wang et al. 2018);
- $\mathcal{I}_4$: proposed Wald-type CI based on $\hat{\delta}$;
- $\mathcal{I}_{4L}$: proposed Wald-type CI based on $\log \hat{\delta}$.

We note that the normal and χ_1^2 distributions may not provide good approximations to $\log \tilde{\delta}$ in $\mathcal{I}_1$ and the ELR statistic in $\mathcal{I}_2$, respectively, especially when n is not large enough. This may be because of the specific features of the two-sample semicontinuous data from model (1): excessive zeros and severe positive/negative skewness of the positive observations. Hence, we employ the nonparametric bootstrap method (Efron and Tibshirani 1993; Shao and Tu 1995) to approximate the quantiles of the target asymptotic distributions, which leads to $\mathcal{I}_{1B}$ and $\mathcal{I}_{2B}$. The number of bootstrap samples is set to 999.

We construct the first four CIs without the DRM (3) and the remaining three CIs with the DRM. We evaluate the performance of a CI in terms of the coverage probability (CP) and average length (AL), which are calculated as follows:

$$\mathrm{CP}(\%) = 100 \times \frac{\sum_{h=1}^{10000} I(\delta_L^{(h)} < \delta < \delta_U^{(h)})}{10000}, \ \mathrm{AL} = \frac{\sum_{h=1}^{10000} \left(\delta_U^{(h)} - \delta_L^{(h)}\right)}{10000}.$$

Here $[\delta_L^{(h)}, \delta_U^{(h)}]$ denotes a CI for δ calculated from the hth model. Table 3 summarizes the simulation results.

From Table 3, we observe that the bootstrap Wald-type CI $\mathcal{I}_{1B}$ and bootstrap ELR-based CI $\mathcal{I}_{2B}$ have much better coverage accuracy than $\mathcal{I}_1$ and $\mathcal{I}_2$, respectively. Comparing $\mathcal{I}_{1B}$ and $\mathcal{I}_{2B}$, we see that $\mathcal{I}_{1B}$ has slightly more accurate CPs in most cases, but $\mathcal{I}_{2B}$ has shorter ALs in most cases. The behavior of $\mathcal{I}_3$ and $\mathcal{I}_{4L}$ is comparable and satisfactory in terms of both CP and AL in every case, while $\mathcal{I}_4$ gives shorter ALs and has lower coverage rates compared with $\mathcal{I}_3$ and $\mathcal{I}_{4L}$, especially for small sample sizes.

Table 2 Bias and mean square error of point estimates for δ, σ_0^2, and σ_1^2

Model	(n_0, n_1)	$\tilde{\delta}$		$\hat{\delta}$		$\tilde{\sigma}_0^2$		$\hat{\sigma}_0^2$		$\tilde{\sigma}_1^2$		$\hat{\sigma}_1^2$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
1	(50, 50)	0.06	0.13	0.04	0.09	0.05	39.62	− 0.03	21.72	0.03	34.87	− 0.01	22.25
	(50, 150)	0.06	0.09	0.03	0.06	− 0.04	23.11	0.01	8.38	0.02	9.23	− 0.03	7.07
	(150, 50)	0.01	0.08	0.02	0.05	− 0.01	10.74	− 0.08	8.09	− 0.02	50.72	0.06	18.71
	(100, 100)	0.03	0.06	0.02	0.04	− 0.01	17.81	− 0.05	10.12	− 0.03	14.53	− 0.06	8.14
2	(50, 50)	0.17	0.59	0.13	0.39	0.09	61.65	0.03	39.30	− 0.01	20.32	− 0.02	15.92
	(50, 150)	0.18	0.41	0.11	0.26	− 0.01	9.91	0.09	6.71	0.02	6.63	− 0.03	4.58
	(150, 50)	0.06	0.25	0.06	0.18	0.00	3.62	− 0.04	2.86	0.04	14.89	0.10	6.11
	(100, 100)	0.09	0.21	0.06	0.15	− 0.01	5.85	− 0.01	3.72	0.01	5.51	− 0.03	3.19
3	(50, 50)	0.06	0.17	0.03	0.11	− 0.05	104.16	− 0.03	54.97	0.19	426.17	− 0.15	292.11
	(50, 150)	0.06	0.10	0.03	0.07	− 0.05	118.37	0.19	36.01	0.11	120.63	− 0.08	100.58
	(150, 50)	0.01	0.11	0.02	0.08	0.02	42.75	− 0.12	23.86	− 0.26	193.29	− 0.16	135.72
	(100, 100)	0.03	0.08	0.02	0.05	− 0.06	57.48	− 0.10	21.76	− 0.09	147.69	− 0.21	108.06
4	(50, 50)	0.09	0.33	0.07	0.24	0.08	147.42	0.03	54.06	− 0.04	458.52	− 0.31	398.77
	(50, 150)	0.09	0.19	0.05	0.14	− 0.02	72.97	0.24	26.55	− 0.08	160.82	− 0.27	138.70
	(150, 50)	0.03	0.21	0.04	0.16	0.00	32.66	− 0.09	18.79	0.02	438.73	− 0.02	289.92
	(100, 100)	0.04	0.14	0.03	0.11	0.03	57.22	0.01	19.11	0.03	275.65	− 0.11	233.74
5	(50, 50)	0.13	0.38	0.08	0.24	0.01	32.45	− 0.02	13.00	− 0.04	25.04	− 0.12	19.46
	(50, 150)	0.13	0.28	0.07	0.18	− 0.05	16.82	0.03	6.33	− 0.07	8.81	− 0.13	6.94
	(150, 50)	0.04	0.18	0.04	0.12	0.00	7.68	− 0.05	5.76	− 0.04	41.11	0.02	16.64
	(100, 100)	0.06	0.16	0.04	0.10	0.02	10.98	0.03	6.97	0.03	18.57	− 0.03	9.81
6	(50, 50)	0.05	0.13	0.04	0.08	− 0.02	27.01	− 0.08	17.19	− 0.07	25.68	− 0.14	13.10
	(50, 150)	0.05	0.09	0.02	0.06	0.06	28.59	0.06	12.45	− 0.07	8.70	− 0.11	6.01
	(150, 50)	0.02	0.08	0.02	0.05	0.00	9.55	− 0.05	8.88	0.06	53.31	0.07	13.47
	(100, 100)	0.03	0.06	0.02	0.04	0.04	17.30	− 0.02	9.59	0.00	14.84	− 0.02	8.38

Table 2 (continued)

Model	(n_0, n_1)	$\tilde{\delta}$		$\hat{\delta}$		$\tilde{\sigma}_0^2$		$\hat{\sigma}_0^2$		$\tilde{\sigma}_1^2$		$\hat{\sigma}_1^2$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
7	(50, 50)	0.17	0.61	0.11	0.39	0.09	33.66	0.02	14.45	− 0.10	21.87	− 0.14	17.33
	(50, 150)	0.19	0.48	0.10	0.30	0.01	25.25	0.10	6.54	− 0.01	8.87	− 0.07	7.44
	(150, 50)	0.06	0.28	0.06	0.19	0.00	5.25	− 0.05	3.62	− 0.05	21.97	0.00	10.64
	(100, 100)	0.09	0.26	0.05	0.17	− 0.02	7.74	0.02	4.52	0.07	18.22	− 0.02	10.97
8	(50, 50)	0.10	0.37	0.06	0.27	0.02	33.28	0.04	9.29	− 0.14	190.88	− 0.41	160.50
	(50, 150)	0.09	0.24	0.05	0.17	− 0.01	27.55	0.15	6.82	0.11	106.76	− 0.03	97.86
	(150, 50)	0.03	0.22	0.02	0.15	− 0.05	8.84	− 0.05	4.72	− 0.12	173.09	− 0.35	109.84
	(100, 100)	0.04	0.17	0.03	0.12	0.02	18.88	0.00	3.54	− 0.12	211.21	− 0.22	196.79
9	(50, 50)	0.37	2.66	0.26	2.09	0.12	37.11	0.16	8.87	0.13	389.19	− 0.11	352.51
	(50, 150)	0.36	1.82	0.23	1.36	− 0.08	7.39	0.19	3.65	0.06	135.93	− 0.10	128.85
	(150, 50)	0.13	1.16	0.10	0.90	− 0.01	4.89	0.03	3.01	0.23	381.40	− 0.08	279.96
	(100, 100)	0.17	0.96	0.11	0.74	− 0.01	6.76	0.04	1.85	− 0.19	124.13	− 0.34	114.59
10	(50, 50)	0.12	0.76	0.08	0.62	− 0.04	18.98	0.13	7.76	0.12	1461.27	− 0.47	1334.57
	(50, 150)	0.13	0.45	0.08	0.35	− 0.03	22.09	0.16	4.31	− 0.09	356.85	− 0.29	345.40
	(150, 50)	0.03	0.48	0.02	0.40	− 0.03	9.00	0.03	3.78	− 0.37	743.33	− 0.95	640.08
	(100, 100)	0.06	0.33	0.04	0.28	0.00	14.03	0.09	3.61	− 0.04	654.22	− 0.34	603.20

Table 3 Coverage probability (%) and average length of 95% CIs for δ

Model	(n_0, n_1)	$\mathcal{I}_1$		$\mathcal{I}_{1B}$		$\mathcal{I}_2$		$\mathcal{I}_{2B}$		$\mathcal{I}_3$		$\mathcal{I}_4$		$\mathcal{I}_{4L}$	
		CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
1	(50, 50)	92.6	1.37	94.6	1.73	91.7	1.34	94.1	1.59	94.6	1.18	93.7	1.09	94.4	1.15
	(50, 150)	92.6	1.09	94.0	1.26	91.8	1.08	93.7	1.22	94.9	0.94	94.2	0.90	94.8	0.93
	(150, 50)	92.5	1.08	94.3	1.86	91.7	1.08	93.6	1.31	94.9	0.92	94.0	0.88	94.7	0.91
	(100, 100)	93.9	0.94	95.1	1.04	93.2	0.95	94.7	1.05	94.9	0.79	94.6	0.76	95.0	0.78
2	(50, 50)	91.5	2.73	94.4	3.70	89.8	2.70	93.3	3.63	94.5	2.48	91.5	2.10	94.9	2.45
	(50, 150)	91.7	2.13	94.0	2.65	90.7	2.16	93.6	2.82	94.5	1.99	92.9	1.74	94.9	1.94
	(150, 50)	91.6	1.92	93.6	2.78	90.6	1.88	93.6	2.58	94.9	1.72	92.2	1.58	94.7	1.75
	(100, 100)	92.3	1.72	94.5	2.11	92.5	1.71	94.5	2.00	94.9	1.52	93.5	1.40	95.3	1.51
3	(50, 50)	92.3	1.55	94.5	1.92	91.5	1.52	94.1	1.84	94.3	1.36	92.8	1.25	94.8	1.34
	(50, 150)	92.6	1.17	94.3	1.36	93.0	1.15	94.7	1.30	94.8	1.04	93.5	0.97	95.0	1.01
	(150, 50)	92.5	1.27	94.2	1.66	91.3	1.25	93.7	1.56	94.8	1.11	93.3	1.06	95.2	1.11
	(100, 100)	94.2	1.06	95.4	1.19	92.7	1.06	93.9	1.18	94.8	0.92	94.2	0.88	95.2	0.91
4	(50, 50)	91.5	2.20	94.0	2.99	90.7	2.09	93.9	2.76	93.7	1.96	90.4	1.73	94.5	1.94
	(50, 150)	92.6	1.60	94.4	1.88	92.0	1.60	93.8	1.86	93.9	1.47	92.2	1.35	94.3	1.45
	(150, 50)	91.8	1.78	93.9	2.66	90.4	1.68	93.1	2.30	94.5	1.57	91.5	1.46	94.6	1.59
	(100, 100)	93.0	1.45	94.7	1.71	92.5	1.45	94.3	1.70	94.7	1.30	92.4	1.21	94.2	1.29
5	(50, 50)	92.9	2.23	95.1	2.69	91.5	2.22	93.9	2.65	95.1	1.96	93.2	1.80	94.6	1.92
	(50, 150)	91.8	1.88	93.4	2.21	91.0	1.86	93.4	2.18	94.7	1.69	93.8	1.57	94.7	1.65
	(150, 50)	92.9	1.64	94.6	1.97	92.5	1.61	93.9	1.88	94.8	1.40	94.0	1.33	94.8	1.38
	(100, 100)	93.4	1.52	94.7	1.68	93.5	1.52	94.7	1.67	94.9	1.30	94.4	1.24	95.0	1.28
6	(50, 50)	91.6	3.95	94.3	5.08	90.9	3.89	93.6	5.06	94.5	3.63	92.5	3.09	94.9	3.49
	(50, 150)	91.4	3.31	93.6	4.18	90.5	3.32	93.3	4.45	94.8	3.13	93.2	2.75	94.9	3.03
	(150, 50)	92.6	2.58	94.3	3.39	91.9	2.55	94.2	3.06	94.9	2.26	93.2	2.12	94.7	2.26
	(100, 100)	92.5	2.50	94.2	2.85	92.4	2.49	94.3	2.83	94.5	2.22	94.0	2.05	94.9	2.17

Table 3 (continued)

Model	(n_0, n_1)	$\mathcal{I}_1$		$\mathcal{I}_{1B}$		$\mathcal{I}_2$		$\mathcal{I}_{2B}$		$\mathcal{I}_3$		$\mathcal{I}_4$		$\mathcal{I}_{4L}$	
		CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
7	(50, 50)	92.1	2.82	94.3	3.53	91.4	2.81	94.1	3.45	94.6	2.52	92.9	2.27	95.0	2.48
	(50, 150)	92.3	2.37	94.1	2.86	90.9	2.38	93.2	2.90	94.6	2.18	93.8	1.99	95.0	2.13
	(150, 50)	92.5	2.00	94.1	2.52	91.6	1.96	93.7	2.31	94.4	1.73	93.7	1.63	94.7	1.71
	(100, 100)	93.3	1.87	94.8	2.10	92.4	1.86	93.9	2.08	94.8	1.63	94.0	1.54	95.0	1.61
8	(50, 50)	92.6	2.23	94.5	2.66	91.4	2.24	93.8	2.65	94.0	1.98	92.6	1.85	94.0	1.95
	(50, 150)	92.6	1.79	94.2	2.07	91.5	1.77	93.5	2.00	94.8	1.61	93.8	1.53	94.5	1.58
	(150, 50)	92.7	1.77	94.5	2.68	92.2	1.78	94.2	2.14	94.4	1.54	92.9	1.47	94.6	1.52
	(100, 100)	93.5	1.56	95.0	1.75	92.8	1.56	94.4	1.72	94.3	1.37	93.5	1.30	94.5	1.34
9	(50, 50)	91.3	5.78	93.8	7.90	90.8	5.70	94.5	7.74	93.3	5.41	89.3	4.60	93.5	5.42
	(50, 150)	92.0	4.56	94.0	5.72	90.9	4.63	93.6	6.49	94.2	4.45	91.8	3.86	94.0	4.33
	(150, 50)	91.7	4.10	94.0	8.56	91.1	3.89	93.6	5.23	93.3	3.65	90.2	3.35	94.0	3.71
	(100, 100)	92.5	3.66	94.3	4.31	91.9	3.64	94.0	4.26	93.7	3.39	91.7	3.08	94.3	3.34
10	(50, 50)	92.3	3.26	94.5	4.10	91.3	3.19	94.1	4.00	93.0	3.00	89.6	2.74	93.3	3.01
	(50, 150)	92.8	2.46	94.5	2.83	91.8	2.42	93.8	2.75	94.1	2.33	92.6	2.16	94.0	2.28
	(150, 50)	92.4	2.68	94.1	3.83	90.8	2.63	93.3	3.43	93.0	2.41	90.4	2.27	93.7	2.44
	(100, 100)	93.7	2.22	95.1	2.53	92.6	2.22	94.5	2.52	94.0	2.06	91.8	1.94	93.0	2.04

In general, the CIs for the DRM are better than those without the DRM. In conclusion, $\mathcal{I}_3$ and $\mathcal{I}_{4L}$ give the best results in terms of CP and AL. However, $\mathcal{I}_{4L}$ has lower computational complexity and uses shorter computation time than $\mathcal{I}_3$, and thus it may be preferred.

Remark 1 We would like to provide some comments on the choice of $\mathbf{q}(x)$ in the DRM (3). To apply the proposed method, the basis function $\mathbf{q}(x)$ is required to be prespecified. We have conducted an additional small simulation to examine the impact of the basis function misspecification on our inference results. The additional simulation results and discussion can be found in the supplementary material. In application, the basis function $\mathbf{q}(x)$ is always unknown. Prior belief and information could be useful for choosing an appropriate $\mathbf{q}(x)$ before employing the proposed method. For example, if one observes that underlying populations have the features of lognormal distributions, the DRM with $\mathbf{q}(x) = \log(x)$ or $\mathbf{q}(x) = (\log(x), \log^2(x))^\top$ could be used instead of a fully parametric model to achieve robustness of inferences. The choice of $\mathbf{q}(x)$ can be further checked by the goodness-of-fit test proposed by Qin and Zhang (1997). Nonparametric methods may be preferable when no prior belief or information is available.

4 Real-data analysis

In this section, we illustrate the performance of our method by analyzing two real datasets. We estimate the mean ratio δ and the population variances σ_0^2, σ_1^2, and we construct the CIs for δ.

The first dataset (Koopmans 1981) is from a biological study of the seasonal activity patterns of a species of field mice. The measurements are the average distances (in meters) traveled between captures by those mice at least twice in a given month. Table 4 summarizes this dataset.

Table 4 shows that there are many zero measurements, especially in Autumn and Winter. Wang et al. (2018) conducted hypothesis tests to determine if the mean traveled distance differs among the four seasons; they found no significant difference between Spring and Summer. Hence, we combine the Spring and Summer measurements into one sample and refer to this as sample 0. Similarly, we combine the Autumn and Winter measurements into sample 1.

Table 4 Summary of mice dataset

Season	Sample size	Proportion (number) of zeros
Spring	17	0.176 (3)
Summer	27	0.111 (3)
Autumn	27	0.370 (10)
Winter	34	0.294 (10)

To analyze the dataset with our method, we need to choose an appropriate $\mathbf{q}(x)$ in the DRM (3). To balance model fitting and model complexity, we choose $\mathbf{q}(x) = \log(x)$. For this choice, the goodness-of-fit test proposed by Qin and Zhang (1997) gives a p-value of 0.64 for the mice data. This may indicate that $\mathbf{q}(x) = \log(x)$ is suitable for this dataset.

We apply all the methods explored in our simulation study. Our estimate $\hat{\delta} = 0.487$, and the fully nonparametric estimate $\tilde{\delta} = 0.483$. Our semiparametric estimates of the two-sample variances are $\hat{\sigma}_0^2 = 869.583$ for sample 0 and $\hat{\sigma}_1^2 = 268.774$ for sample 1; the fully nonparametric estimates are $\tilde{\sigma}_0^2 = 932.966$ for sample 0 and $\tilde{\sigma}_1^2 = 239.961$ for sample 1. Given the simulation results in Table 2, our point estimates are expected to be more accurate.

The 95% CIs for δ are presented in Table 5: $\mathcal{I}_4$ is the shortest and $\mathcal{I}_{1B}$ the longest. The lower and upper bounds of $\mathcal{I}_4$ tend to be smaller than those of the other CIs. The results for $\mathcal{I}_{2B}$, $\mathcal{I}_3$, and $\mathcal{I}_{4L}$ are similar. The CIs do not include 1, which indicates a significant mean difference between the two samples.

The second dataset (Neuhauser 2011) is from a study of the methylation of DNA, which is a common method for gene regulation. The methylation patterns of tumor cells can be compared to those of normal cells; there are also differences between different types of cancer. DNA methylation can serve as a biomarker in cancer diagnosis. The dataset consists of two samples of methylation measurements: small-cell lung cancer (sample 0) and non-small-cell lung cancer (sample 1). When methylation is undetectable or only partially present, the result of the measurement is negative, which is treated as a zero value. Fully present methylation gives a positive value. Sample 0 contains 41 measurements, of which 25 are zero. Sample 1 contains 46 measurements, of which 16 are zero.

Satter and Zhao (2021) argued that this dataset is highly skewed. This may suggest that it can be fitted by the DRM with $\mathbf{q}(x) = \log(x)$. The goodness-of-fit test of Qin and Zhang (1997) gives a p-value of 0.133. Therefore, there is no strong evidence for rejecting the DRM with $\mathbf{q}(x) = \log(x)$.

We apply all the methods explored in our simulation study. Our estimate $\hat{\delta} = 2.906$, and the fully nonparametric estimate $\tilde{\delta} = 3.679$. For the two-sample variances, our semiparametric estimates are $\hat{\sigma}_0^2 = 388.562$ for sample 0 and $\hat{\sigma}_1^2 = 1028.079$ for sample 1; the fully nonparametric estimates are $\tilde{\sigma}_0^2 = 406.796$ for sample 0 and $\tilde{\sigma}_1^2 = 1017.072$ for sample 1. There are differences between our estimates and the fully nonparametric estimates, especially for δ. We trust our estimates because our simulations have demonstrated the performance of our estimators.

Table 6 presents the 95% CIs for δ. According to the simulation results in Table 3, $\mathcal{I}_{1B}$, $\mathcal{I}_{2B}$, $\mathcal{I}_3$, and $\mathcal{I}_{4L}$ have better coverage accuracy. The CIs $\mathcal{I}_{1B}$ and $\mathcal{I}_{2B}$ contain 1,

Table 5 95% confidence intervals for δ (mice data)

	$\mathcal{I}_1$	$\mathcal{I}_{1B}$	$\mathcal{I}_2$	$\mathcal{I}_{2B}$	$\mathcal{I}_3$	$\mathcal{I}_4$	$\mathcal{I}_{4L}$
Lower bound	0.319	0.314	0.318	0.322	0.325	0.295	0.328
Upper bound	0.729	0.741	0.726	0.716	0.721	0.679	0.722
Length	0.410	0.427	0.408	0.393	0.396	0.383	0.393

Table 6 95% confidence intervals for δ (methylation data)

	$\mathcal{I}_1$	$\mathcal{I}_{1B}$	$\mathcal{I}_2$	$\mathcal{I}_{2B}$	$\mathcal{I}_3$	$\mathcal{I}_4$	$\mathcal{I}_{4L}$
Lower bound	1.056	0.650	1.158	0.568	1.278	0.362	1.211
Upper bound	12.814	20.838	12.306	27.631	7.527	5.451	6.975
Length	11.758	20.189	11.148	27.063	6.249	5.089	5.764

whereas $\mathcal{I}_3$ and $\mathcal{I}_{4L}$ do not. This indicates that $\mathcal{I}_3$ and $\mathcal{I}_{4L}$ provide more evidence than $\mathcal{I}_{1B}$ and $\mathcal{I}_{2B}$ for rejecting $H_0 : \delta = 1$. We note that $\mathcal{I}_{4L}$ is slightly shorter than $\mathcal{I}_3$.

5 Concluding remarks

We have proposed new statistical procedures for semiparametric inference on the general functional $\boldsymbol{\psi}$ defined in (4) and their functions $\mathbf{g}(\boldsymbol{\psi})$ with two samples of semicontinuous observations. The functional $\boldsymbol{\psi}$ includes the linear functionals $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$ as special cases. Under the semiparametric DRM (3), we have constructed the MELE of $\boldsymbol{\psi}$ and established the asymptotic normality of the MELE of $\boldsymbol{\psi}$. The MELEs of $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$ were shown to be more efficient than the fully nonparametric alternatives both theoretically and via simulation. We have applied the asymptotic results to construct confidence regions and perform hypothesis tests for $\boldsymbol{\psi}$ and $\mathbf{g}(\boldsymbol{\psi})$. We note that our methods and the general results can be applied to many important summary quantities, such as the uncentered and centered moments, the mean ratio, the coefficient of variation, and the generalized entropy class of inequality measures. As an illustration, we have considered the construction of CIs for the mean ratio of two such populations. Simulation results showed that the proposed Wald-type CIs have performance similar to that of the ELR-based CI under the DRM, and the computational cost is lower. We have implemented our methods in `R`; the code is available upon request.

It would be interesting to extend the current framework to general expectation functionals and their functions, e.g., the receiver operating characteristic (ROC) curve, the area under the ROC curve, and the Gini index. The associated theoretical development may be challenging.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-021-00804-4.

Acknowledgements The authors thank the Chief Editor, the Associate Editor, and two reviewers for their very careful reading and a number of helpful comments. The authors are grateful to Dr. Changbao Wu for his constructive and helpful comments. Dr. Wang's work is supported in part by National Natural Science Foundation of China Grants 12001454, 11971404, Humanities and Social Sciences Foundation of the Ministry of Education of China Grant 19YJC910005, and Natural Science Foundation of Fujian Province Grant 2020J01031. Dr. Li's work is supported in part by the Natural Sciences and Engineering Research Council of Canada Grant RGPIN-2020-04964.

References

Anderson, J. A. (1979). Multivariate logistic compounds. *Biometrika, 66,* 17–26.

Böhning, D., Alfò, M. (2016). Editorial: Special issue on models for continuous data with a spike at zero. *Biometrical Journal, 58,* 255–258.

Brunner, E., Dette, H., Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association, 92,* 1494–1502.

Cai, S., Chen, J. (2018). Empirical likelihood inference for multiple censored samples. *The Canadian Journal of Statistics, 46*(2), 212–232.

Cai, S., Chen, J., Zidek, J. V. (2017). Hypothesis test in the presence of multiple samples under density ratio models. *Statistica Sinica, 27,* 761–783.

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

Chen, Y.-H., Zhou, X.-H. (2006). *Generalized confidence intervals for the ratio or difference of two means for lognormal populations with zeros*. Working Paper 296, UW Biostatistics Working Paper Series. https://biostats.bepress.com/uwbiostat/paper296.

Dufour, J.-M., Flachaire, E., Khalaf, L. (2019). Permutation tests for comparing inequality measures. *Journal of Business and Economic Statistics, 37,* 457–470.

Efron, B., Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.

Fernholz, L. T. (1983). *von Mises calculus for statistical functionals*. New York: Springer.

Jiang, S., Tu, D. (2012). Inference on the probability $P(T_1 < T_2)$ as a measurement of treatment effect under a density ratio model and random censoring. *Computational Statistics and Data Analysis, 56,* 1069–1078.

Kang, L., Vexler, A., Tian, L., Cooney, M., Louis, G. M. B. (2010). Empirical and parametric likelihood interval estimation for populations with many zero values: Application for assessing environmental chemical concentrations and reproductive health. *Epidemiology, 21,* S58–S63.

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

Koopmans, L. H. (1981). *Introduction to contemporary statistical methods*. Boston: Duxbury Press.

Li, H., Liu, Y., Liu, Y., Zhang, R. (2018). Comparison of empirical likelihood and its dual likelihood under density ratio model. *Journal of Nonparametric Statistics, 30,* 581–597.

Lu, Y.-H., Liu, A.-Y., Jiang, M.-J., Jiang, T. (2020). A new two-part test based on density ratio model for zero-inflated continuous distributions. *Applied Mathematics-A Journal of Chinese Universities, 35,* 203–219.

Neuhauser, M. (2011). *Nonparametric statistical tests*: *A computational approach*. Boca Raton: CRC Press.

Nixon, R. M., Thompson, S. G. (2004). Parametric modelling of cost data in medical studies. *Statistics in Medicine, 23,* 1311–1331.

Owen, A. (2001). *Empirical likelihood*. New York: CRC Press.

Pauly, M., Brunner, E., Konietschke, F. (2015). Asymptotic permutation tests in general factorial designs. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)*, *77,* 461–473.

Qin, J. (2017). *Biased sampling, over-identified parameter problems and beyond*. Singapore: Springer.

Qin, J., Zhang, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika, 84,* 609–618.

Satter, F., Zhao, Y. (2021). Jackknife empirical likelihood for the mean difference of two zero-inflated skewed populations. *Journal of Statistical Planning and Inference, 211,* 414–422.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

Shao, J., Tu, D. (1995). *The jackknife and bootstrap*. New York: Springer.

Tu, W., Zhou, X.-H. (1999). A Wald test comparing medical costs based on log-normal distributions with zero valued costs. *Statistics in Medicine, 18,* 2749–2761.

Wang, C., Marriott, P., Li, P. (2017). Testing homogeneity for multiple nonnegative distributions with excess zero observations. *Computational Statistics and Data Analysis, 114,* 146–157.

Wang, C., Marriott, P., Li, P. (2018). Semiparametric inference on the means of multiple nonnegative distributions with excess zero observations. *Journal of Multivariate Analysis, 166,* 182–197.

Wu, C., Yan, Y. (2012). Empirical likelihood inference for two-sample problems. *Statistics and Its Interface, 5,* 345–354.

Yuan, M., Li, P., Wu, C. (2021). Semiparametric inference of the Youden index and the optimal cut-off point under density ratio models. *The Canadian Journal of Statistics*. https://doi.org/10.1002/cjs.11600.

Zhou, X.-H., Tu, W. (1999). Comparison of several independent population means when their samples contain log-normal and possibly zero observations. *Biometrics, 55,* 645–651.

Zhou, X.-H., Tu, W. (2000). Interval estimation for the ratio in means of log-normally distributed medical costs with zero values. *Computational Statistics and Data Analysis, 35,* 201–210.