To appear in the *Journal of Nonparametric Statistics*
Vol. 00, No. 00, Month 20XX, 1–24

Tests for stochastic ordering under biased sampling

Hsin-wen Chang[a] † and Hammou El Barmi[b] ‡ and Ian W. McKeague[c] §
[a]*Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.*; [b]*Department of Statistics and Computer Information Systems, Baruch College, The City University of New York, New York, NY 10010, U.S.A.*; [c]*Department of Biostatistics, Columbia University, New York, NY 10032, U.S.A.*

(*Received 00 Month 20XX; in final form 00 Month 20XX*)

In two-sample comparison problems it is often of interest to examine whether one distribution function majorizes the other, i.e., for the presence of stochastic ordering. This paper develops a nonparametric test for stochastic ordering from size-biased data, allowing the pattern of the size bias to differ between the two samples. The test is formulated in terms of a maximally-selected local empirical likelihood statistic. A Gaussian multiplier bootstrap is devised to calibrate the test. Simulation results show that the proposed test outperforms an analogous Wald-type test, and that it provides substantially greater power over ignoring the size bias. The approach is illustrated using data on blood alcohol concentration of drivers involved in car accidents, where the size bias is due to drunker drivers being more likely to be involved in accidents. Further, younger drivers tend to be more affected by alcohol, so in making comparisons with older drivers the analysis is adjusted for differences in the patterns of size bias.

Keywords: empirical likelihood; length bias; order-restricted inference; size bias; weighted distributions

AMS Subject Classification: 62G10; 62G20; 62G30

1. Introduction

Sampling bias is encountered in numerous biomedical settings, especially in genetics (see, e.g., Clark et al. 2005, ascertainment bias), wildlife population studies (Patil and Rao 1978), disease screening (Zelen and Feinleib 1969; Duffy et al. 2008), and vaccine efficacy trials (Gilbert et al. 1999). It is also an increasingly important issue in the analysis of "big data" (Harford 2014). Due to the bias, inference concerning the underlying distribution functions needs to be tailored to take into account known features of the sampling mechanism. This is especially challenging in two-sample settings when the bias affects the two groups in different ways.

In this paper we study the problem of testing for stochastic ordering between two lifetime distribution functions, F_1 and F_2, given independent samples from

$$G_j(x) = \int_0^x \frac{w_j(u)}{W_j} dF_j(u),$$

† Corresponding author. Email: hwchang@stat.sinica.edu.tw; phone: +886-2-66145636; fax: +886-2-27831523.
‡ Email: Hammou.Elbarmi@baruch.cuny.edu; phone: +1-646-312-3384.
§ Email: im2131@columbia.edu; phone: +1-212-342-1242; fax: +1-212-305-9408.

where the weight functions $w_j(\cdot) > 0$ are assumed to be known, and $W_j = \int_0^\infty w_j(u)\, dF_j(u) < \infty$ are the normalizing constants. Such data are called size-biased because the weights $w_j(x)$ depend on the size x of the datum. The comparison between F_1 and F_2 will be restricted to a given fixed interval $[t_1, t_2]$. The nonparametric maximum likelihood estimator (NPMLE) of a distribution function from a biased sample of this type and its asymptotic properties are known —Vardi (1982) and Gill et al. (1988) dealt with the uncensored case and Kvam et al. (1999) considered the censored case when the weight function takes the special form $w(x) = I(x > x_0)$ with some known x_0. The asymptotic distribution of the difference of the NPMLEs of F_1 and F_2 is therefore easily derived, but is not distribution free, and a suitable method for calibrating even an omnibus test for $F_1 = F_2$ versus $F_1 \neq F_2$ is not readily available. We will develop a Gaussian multiplier bootstrap approach for this simultaneous inference problem; our approach is also easily adapted to the case of an omnibus alternative.

The distribution function F_1 is said to be *stochastically larger* than F_2 on $[t_1, t_2]$ if $F_1(x) \leq F_2(x)$ for all $x \in [t_1, t_2]$; this ordering is denoted $F_1 \succeq F_2$. We investigate the problem of testing the two-sided alternative

$$H_0 : F_1 = F_2 \text{ versus } H_1 : F_1 \succ F_2 \text{ or } F_2 \succ F_1 \tag{1}$$

where $\succ$ denotes $\succeq$ with strict inequality for some $x \in [t_1, t_2]$. Our approach will first be developed for testing the one-sided alternative

$$H_0 : F_1 = F_2 \text{ versus } H_1 : F_1 \succ F_2 \tag{2}$$

and then extended to the two-sided alternative using the union-intersection principle.

The union of the null and alternative hypotheses in (1) excludes the possibility of crossing distribution functions: $F_1(u) > F_2(u)$ for some $u \geq 0$, $F_1(v) < F_2(v)$ for some $v \geq 0$. Denote the hypothesis of crossing distribution functions by H_c. An auxiliary test for the null hypothesis H_c (against the alternative that H_c is not true) is therefore recommended. Such a test is readily constructed using a simultaneous confidence band for the difference $F_2 - F_1$ (see Section 2.2.2). Reject H_c if the lower boundary of the confidence band is ≤ 0 or its upper boundary ≥ 0. It can be shown that the family-wise error rate of this auxiliary test combined with our test for stochastic ordering can be controlled at same alpha level as the individual tests, cf. Chang and McKeague (2014).

Our approach to testing (1) and (2) is based on the method of empirical likelihood (EL) (Owen 1988, 2001). This method has been adapted to various biased sampling problems (Qin 1993; El Barmi and Rothmann 1998; Davidov et al. 2010) and is more appealing than the Wald approach because it is self-studentized and provides more accurate Type-I error control. There is also evidence that EL-based tests have optimal power (see, e.g., Kitamura et al. 2012). As far as we know, however, the approach has not been used to develop tests for stochastic ordering under biased sampling.

First considering the one-sided alternative in (2), we devise a localized EL statistic for $H_0^t : F_1(t) = F_2(t)$ versus $H_1^t : F_1(t) < F_2(t)$ at each given $t \in [t_1, t_2]$. The proposed test rejects H_0 for large values of the maximally-selected EL statistic. Such a localization strategy has been used in Einmahl and McKeague (2003) and El Barmi and McKeague (2013) for testing various nonparametric hypotheses, except they considered an integral type test statistic and restricted attention to data without sampling bias. Various Kolmogorov–Smirnov type test statistics (not based on EL) for stochastic ordering have been proposed by El Barmi and Mukerjee (2005) and Davidov and Herman (2009), but these cannot deal with size-biased data either.

We find the limiting distribution of the resulting maximally-selected EL statistic, but it is not distribution-free (it depends on F_1 and F_2), so calibration of the test becomes a challenge. As mentioned before, we develop a Gaussian multiplier bootstrap approach to resolve this problem. The proposed calibration method is fast, because it avoids recalculation of several computationally expensive quantities in each bootstrap sample. The multiplier bootstrap has been been widely applied; for a recent example of its use to calibrate a two-sample nonparametric test, see Rémillard and Scaillet (2009).

The paper is organized as follows. In Section 2.1 we consider the one-sample case and in Section 2.2 we extend the theory to the two-sample case. Section 3 presents results from a simulation study: the proposed EL test performs better than the Wald test and the test ignoring size bias, in terms of both accuracy and power. Section 4 then provides an application of the proposed test to alcohol concentration records in fatal driving accidents. Finally, some concluding remarks are placed in Section 5.

2. EL tests for stochastic ordering in biased sampling model

2.1. *One-sample case*

In this section we drop the earlier subscript j, and just write F, G, w and W. The observed data $\{X_i, i = 1, \dots, n\}$ are iid copies of $X \sim G$. The nonparametric likelihood (see, e.g., Owen 2001, Ch. 6.1) for F is

$$L(F) = \prod_{i=1}^{n} dG(X_i) = \prod_{i=1}^{n} \frac{w(X_i)dF(X_i)}{W} = \prod_{i=1}^{n} \frac{w_i p_i}{W},$$

where $w_i = w(X_i)$ and p_i is the point mass that F places at X_i. The NPMLE is given by $\tilde{F}(t) \equiv \sum_{i=1}^{n} \tilde{p}_i I_{X_i \le t}$, where $\tilde{p}_i = \tilde{W}/(nw_i)$ and $\tilde{W} = n/\sum_{i=1}^{n}(1/w_i)$.

Consider testing $H_0 : F = F_0$ versus $H_1 : F \succ F_0$, where F_0 is a known distribution function. Our procedure is to first construct the statistic for testing the "local" hypotheses $H_0^t : F(t) = F_0(t)$ versus $H_1^t : F(t) < F_0(t)$ for a given t, and then to deal with the general hypotheses based on some functional of the local statistics.

To construct the local test statistic at t, consider the EL ratio

$$\mathcal{R}(t) = \frac{\sup\{L(F) : F(t) = F_0(t)\}}{\sup\{L(F) : F(t) \le F_0(t)\}}. \tag{3}$$

In defining $\mathcal{R}(t)$ we adopt the convention that $\sup \emptyset = 0$ and $0/x = 1$. A tractable form of the EL ratio can be obtained by comparing $F_0(t)$ and $\tilde{F}(t)$, the unconstrained maximizer of $L(F)$. When $\tilde{F}(t) \le F_0(t)$, the denominator of $\mathcal{R}(t)$ is the unconstrained maximum given by $\prod_{i=1}^{n}(w_i\tilde{p}_i)/\tilde{W} = \prod_{i=1}^{n}(1/n)$. When $\tilde{F}(t) > F_0(t)$, the constrained maximum in the denominator is attained on the boundary of the constraint set, and then $\mathcal{R}(t) = 1$. That is,

$$\mathcal{R}(t) = \begin{cases} 1 & \text{if } \tilde{F}(t) > F_0(t), \\ \dfrac{\sup\{L(F) : F(t) = F_0(t)\}}{n^{-n}} & \text{if } \tilde{F}(t) \le F_0(t). \end{cases}$$

To simplify the above expression, we follow a similar derivation to the literature on EL-based testing in biased sampling models (see, e.g., Qin 1993; El Barmi and Rothmann

1998). The numerator is seen to be $\prod_{i=1}^{n}(w_i\hat{p}_i)/\hat{W}$, where

$$\hat{p}_i = \frac{1}{n}\frac{1}{w_i/\hat{W} + \hat{\lambda}(I_{X_i \le t} - F_0(t))}$$

and $(\hat{W}, \hat{\lambda})$ satisfy the estimating equations $\sum_{i=1}^{n}\hat{p}_i(w_i - \hat{W}) = 0$ and $\sum_{i=1}^{n}\hat{p}_i(I_{X_i\le t} - F_0(t)) = 0$. This results in

$$\mathcal{R}(t) = \begin{cases} 1 & \text{if } \tilde{F}(t) > F_0(t), \\ \displaystyle\prod_{i=1}^{n}\frac{nw_i\hat{p}_i}{\hat{W}} & \text{if } \tilde{F}(t) \le F_0(t). \end{cases}$$

The derivation is omitted because it is similar to the two-sample case (presented in Appendix A).

To derive the large sample properties of the local EL test statistic $-2\log\mathcal{R}(t)$, we will first approximate it by a Wald-type counterpart $U_n^2(t)I_{U_n(t)\ge 0}$, where

$$U_n(t) = \hat{\sigma}^{-\frac{1}{2}}(t,t)\left[\frac{\hat{W}}{\sqrt{n}}\sum_{i=1}^{n}\frac{F_0(t) - I_{X_i\le t}}{w_i}\right] \tag{4}$$

$$= \sqrt{n}\hat{\sigma}^{-\frac{1}{2}}(t,t)\frac{\hat{W}}{\tilde{W}}\left[F_0(t) - \tilde{F}(t)\right] \tag{5}$$

and $\hat{\sigma}(t,t) = \hat{W}^2\sum_{i=1}^{n}[(I_{X_i\le t} - F_0(t))/w_i]^2/n$ (The derivation is omitted because it is similar to the two-sample case presented in Appendix B.) Both of the above expressions are useful: (5) shows that maximizing the positive part of $U_n(t)$ (over $t \in [t_1, t_2]$) extends the one-sided Kolmogorov–Smirnov statistic by adjusting for size bias (see Remark 3 below), whereas (4) will be the basis for the multiplier bootstrap approach introduced later. It can be shown that $U_n^2(t)I_{U_n(t)\ge 0}$ has asymptotically a chi-bar square distribution under H_0^t, which is also the limiting null distribution of $-2\log\mathcal{R}(t)$. That is, for t such that $0 < F_0(t) < 1$,

$$-2\log\mathcal{R}(t)\xrightarrow{d}Z_+^2$$

under H_0^t, where $Z \sim N(0,1)$ and $Z_+ = \max(Z, 0)$. This result can be used to test the local H_0^t versus H_1^t.

To test for the alternative of stochastic ordering, we propose the following maximally-selected local EL statistic:

$$M_n = \sup_{t\in[t_1,t_2]}\left[-2\log\mathcal{R}(t)\right]. \tag{6}$$

Our first result gives the asymptotic null distribution of M_n. The proof is omitted because it is similar to the two-sample case (presented in Appendix B).

Theorem 2.1 *Suppose* $0 < F_0(t_1) < F_0(t_2) < 1$ *and* $\int_0^\infty w(u)^{-1}\,dF_0(u) < \infty$. *Then, under* H_0

$$M_n\xrightarrow{d}\sup_{t\in[t_1,t_2]}\left[U_+^2(t)\right],$$

where $U_+ = \max(U, 0)$*,* $U(t)$ *is a mean-zero pinned Gaussian process with covariance function*

$$\mathrm{cov}(U(s), U(t)) = \sigma(s,t)/\sqrt{\sigma(s,s)\sigma(t,t)}$$

and

$$\sigma(s,t) = W^2 E[(I_{X\leq s} - F_0(s))(I_{X\leq t} - F_0(t))/w^2(X)].$$

Remark 1. Gill et al. (1988) studied the problem of estimating a single underlying distribution function F_0 from multiple biased samples having different weight functions. The moment condition $\int_0^\infty w(u)^{-1}\, dF_0(u) < \infty$ we use in Theorem 2.1 matches the moment condition they imposed on each weight function. This moment condition is needed to check the Donsker property of the class of functions of the form $x \mapsto I_{x\leq t}/w(x)$. Our approach is based on bracketing entropy (rather than uniform entropy), which leads to a more direct proof.

As an example that the moment condition suffices, consider $w(\cdot) = 1$. Then our problem reduces to finding uniform convergence of a standardized version of $\sqrt{n}(F_n - F_0)$, where $F_n(t)$ denotes the empirical cdf (see Remark 3 below). In this case our moment condition simplifies to $\int_0^\infty dF_0(u) < \infty$, which always holds.

Remark 2. Given an additional unbiased sample directly from F_0 (see, e.g., El Barmi and Rothmann 1998), a moment condition is not needed since the denominator $w(x)$ of the relevant function class in that case is replaced by $1 - \kappa + \kappa w(x)/W$, which is bounded away from zero provided $\kappa < 1$, where κ is the proportion of the combined sample size in the biased sample.

Remark 3. Under H_0, when there is no size bias (i.e., $w(\cdot) \equiv 1$), $U_n(t)$ reduces to the familiar form

$$\frac{\sqrt{n}\,[F_0(t) - F_n(t)]}{\sqrt{F_n(t)\,(1 - F_n(t))}},$$

where $F_n(t)$ denotes the empirical cdf, and in this case

$$M_n \xrightarrow{d} \sup_{x\in[x_1,x_2]} \frac{B_+^2(x)}{x(1-x)},$$

where B is a standard Brownian bridge on $[0,1]$ and $x_l = F_0(t_l)$, $l = 1, 2$.

The limit in Theorem 2.1 is not distribution free. To obtain critical values for M_n, we propose a Gaussian multiplier bootstrap approach.

2.1.1. *Gaussian multiplier bootstrap*

Our proposed test statistic M_n is asymptotically equivalent to $\sup_{t\in[t_1,t_2]}[\, U_n^2(t) I_{U_n(t)\geq 0}]$, so in view of (4) it is reasonable to construct a Gaussian multiplier bootstrap in terms of

$$U_n^*(t) = \tilde{\sigma}^{*-\frac{1}{2}}(t,t)\left[\frac{\tilde{W}}{\sqrt{n}} \sum_{i=1}^n \xi_i \frac{\tilde{F}(t) - I_{X_i\leq t}}{w_i}\right],$$

where the ξ_i are i.i.d. $N(0,1)$ and independent of the data, $\tilde{\sigma}^*(t,t)$ is the sample variance of $V_i^* = \xi_i \tilde{W}[\tilde{F}(t) - I_{X_i \le t}]/w_i$, and $\tilde{F}(t)$ (the NPMLE) and $\tilde{W}$ are defined in the beginning of Section 2.1. The reason to use the NPMLE instead of $F_0(t)$ under the null is to avoid a loss of power due to imposing the null hypothesis on the bootstrap (see, e.g., Hall and Wilson 1991).

We will show bootstrap consistency of the process $U_n^*(t)$ for the distribution of $U_n(t)$, which leads to the same property for

$$M_n^* \equiv \sup_{t\in[t_1,t_2]} \left[U_n^{*2}(t) I_{U_n^*(t)\ge 0} \right],$$

as in the following theorem. The proof is omitted because it is similar to the two-sample case (see Appendix D).

Theorem 2.2 *Under H_0 and the conditions of Theorem 2.1, conditionally on the data almost surely,*

$$M_n^* \xrightarrow{d} \sup_{t\in[t_1,t_2]} \left[U_+^2(t) \right].$$

Based on Theorem 2.2, to calibrate the test we simulate M_n^* by repeatedly generating samples of Gaussian random multipliers $\{\xi_i\}$. We then compare the empirical quantiles of these bootstrapped values M_n^* with our test statistic M_n.

2.2. *Two-sample case*

The notation is similar to the one-sample case, with the further subscript j now indicating the j-th sample. The nonparametric likelihood $L(F_1, F_2)$ is the product of the two one-sample likelihoods, $L(F_1)L(F_2)$. We assume that the sample proportion $\kappa_j \equiv n_j/n > 0$ remains fixed as the total sample size $n = n_1 + n_2 \to \infty$.

The "local" hypotheses are $H_0^t : F_1(t) = F_2(t)$ versus $H_1^t : F_1(t) < F_2(t)$ for a given t. Denote the common cdf under H_0^t by $F_0(t)$. The local EL ratio at t is defined to be

$$\mathcal{R}(t) = \frac{\sup\{L(F_1,F_2) : F_1(t) = F_2(t)\}}{\sup\{L(F_1,F_2) : F_1(t) \le F_2(t)\}}. \tag{7}$$

By Lagrange multipliers (see Appendix A) we obtain

$$\mathcal{R}(t) = \begin{cases} 1 & \text{if } \tilde{F}_1(t) > \tilde{F}_2(t), \\ \displaystyle\prod_{j=1}^{2} \prod_{i=1}^{n_j} \frac{n_j w_{ij} \hat{p}_{ij}}{\hat{W}_j} & \text{if } \tilde{F}_1(t) \le \tilde{F}_2(t), \end{cases} \tag{8}$$

where $\tilde{F}_1$ and $\tilde{F}_2$ are the unconstrained NPMLEs, and $\hat{p}_{ij}, \hat{W}_j, \hat{\lambda}$, and $\hat{F}_0(t)$ satisfy the system of equations

$$\hat{p}_{ij} = \frac{1}{n} \frac{1}{(\kappa_j w_{ij})/\hat{W}_j + \hat{\lambda}(-1)^{j-1}(I_{X_{ij}\le t} - \hat{F}_0(t))},$$

$$\sum_{i=1}^{n_j} \hat{p}_{ij} \left(w_{ij} - \hat{W}_j \right) = 0, \quad \sum_{i=1}^{n_j} \hat{p}_{ij} \left(I_{X_{ij} \le t} - \hat{F}_0(t) \right) = 0. \tag{9}$$

Under H_0^t, $\hat{F}_0(t)$ is the maximum EL estimate of the common distribution function at t. The local EL test statistic $-2 \log \mathcal{R}(t)$ is shown to converge weakly to a chi-bar square distribution. The derivation involves approximating $-2 \log \mathcal{R}(t)$ by its Wald-type counterpart $U_n^2(t) I_{U_n(t) \ge 0}$, where

$$U_n(t) = \hat{\sigma}^{-\frac{1}{2}}(t,t) \left[\frac{\hat{W}_2}{\sqrt{n_2}\sqrt{\kappa_2}} \sum_{i=1}^{n_2} \frac{I_{X_{i2} \le t} - \hat{F}_0(t)}{w_{i2}} - \frac{\hat{W}_1}{\sqrt{n_1}\sqrt{\kappa_1}} \sum_{i=1}^{n_1} \frac{I_{X_{i1} \le t} - \hat{F}_0(t)}{w_{i1}} \right] \tag{10}$$

and $\hat{\sigma}(t,t) = \sum_{j=1}^{2} (\hat{W}_j^2/\kappa_j) \sum_{i=1}^{n_j} [(I_{X_{ij} \le t} - \hat{F}_0(t))/w_{ij}]^2/n_j$ (see Appendix B for more details).

To test H_0 vs. H_1, we propose the maximally selected EL statistic M_n as in (6), except $\mathcal{R}(t)$ is now given in (8). The following result gives the asymptotic null distribution of M_n (see Appendix B for the proof).

Theorem 2.3 *Suppose* $0 < F_0(t_1) < F_0(t_2) < 1$ *and* $\int_0^\infty w_j(u)^{-1}\, dF_0(u) < \infty$. *Then, under* H_0

$$M_n \xrightarrow{d} \sup_{t \in [t_1, t_2]} \left[U_+^2(t) \right],$$

where $U(t)$ *is a mean-zero pinned Gaussian process with covariance function*

$$\text{cov}(U(s), U(t)) = \sigma(s,t)/\sqrt{\sigma(s,s)\sigma(t,t)}$$

and

$$\sigma(s,t) = \sum_{j=1}^{2} (W_j^2/\kappa_j) E[(I_{X_j \le s} - F_0(s))(I_{X_j \le t} - F_0(t))/w_j^2(X_j)].$$

Remark 1. When there is no size bias (i.e., $w_j(\cdot) \equiv 1$), $U_n(t)$ reduces to

$$\sigma_n^{-1/2}(t,t)\sqrt{n}[F_{n_2 2}(t) - F_{n_1 1}(t)],$$

where $F_{n_j j}(t)$ is the empirical cdf of the j-th sample and

$$\sigma_n(t,t) = \sum_{j=1}^{2} (1/\kappa_j) \sum_{i=1}^{n_j} (I_{X_{ij} \le t} - \hat{F}_0(t))^2/n_j.$$

This implies that M_n is asymptotically equivalent to

$$\sup_{t \in [t_1, t_2]} \left\{ \sigma_n^{-1}(t,t) n \left[F_{n_2 2}(t) - F_{n_1 1}(t) \right]_+^2 \right\},$$

which is the square of the one-sided scaled version of the commonly used two-sample Kolmogorov–Smirnov statistic, $\sup_{t \in [t_1, t_2]} \left[F_{n_2 2}(t) - F_{n_1 1}(t) \right]_+$.

Remark 2. As an example that the moment condition $\int_0^\infty w_j(u)^{-1}\,dF_0(u) < \infty$ suffices, consider $w_1(\cdot) = 1$ and $w_2(\cdot) = 1$. Then our problem reduces to finding uniform convergence of a standardized version of $\sqrt{n}(F_{n_2 2}(t) - F_{n_1 1}(t))$ (see Remark 1 above). In this case our moment condition simplifies to $\int_0^\infty dF_0(u) < \infty$, which always holds.

Remark 3. The condition $0 < F_0(t_1) < F_0(t_2) < 1$ suggests a data-driven rule for t_1 and t_2: $t_1 = \inf\{t : \tilde{F}_1(t) > 0 \text{ and } \tilde{F}_2(t) > 0\}$ and $t_2 = \sup\{t : \tilde{F}_1(t) < 1 \text{ and } \tilde{F}_2(t) < 1\}$. This is what we use in the later simulation runs and data analysis.

As in the one-sample case, we use a Gaussian multiplier bootstrap to calibrate the test.

2.2.1. *Gaussian multiplier bootstrap*

Similar to the one-sample case, it suffices to bootstrap $U_n(t)$ since the test statistic M_n is asymptotically equivalent to $\sup_{t\in[t_1,t_2]}[U_n^2(t) I_{U_n(t)\geq 0}]$. Define a Gaussian multiplier bootstrap for $U_n(t)$ by

$$U_n^*(t) = \tilde{\sigma}^{*-\frac{1}{2}}(t,t)\left[\frac{\tilde{W}_2}{\sqrt{n_2}\sqrt{\kappa_2}}\sum_{i=1}^{n_2}\xi_{i2}\frac{I_{X_{i2}\leq t} - \tilde{F}_2(t)}{w_{i2}} - \frac{\tilde{W}_1}{\sqrt{n_1}\sqrt{\kappa_1}}\sum_{i=1}^{n_1}\xi_{i1}\frac{I_{X_{i1}\leq t} - \tilde{F}_1(t)}{w_{i1}}\right],$$

where the ξ_{ij} are i.i.d. $N(0,1)$ and independent of the data, and $\tilde{\sigma}^*(t,t)$ is the sum of the sample variance of $V_{ij}^* = \xi_{ij}\tilde{W}_j[I_{X_{ij}\leq t} - \tilde{F}_j(t)]/(\sqrt{\kappa_j}w_{ij})$ over $j = 1,2$. Similarly to what was noted in Section 2.1.1, for estimates of unknown quantities $F_j(t)$ and W_j $(j = 1,2)$ in our bootstrap, we use the NPMLEs (the unconstrained maximizers) instead of the constrained maximizer under the null, to avoid a loss of power.

We show bootstrap consistency of $U_n^*(t)$, thereby establishing consistency of

$$M_n^* \equiv \sup_{t\in[t_1,t_2]}\left[U_n^{*2}(t) I_{U_n^*(t)\geq 0}\right].$$

The result is provided in the following theorem (see Appendix D for the proof).

THEOREM 2.4 *Under H_0 and the conditions of Theorem 2.3, conditionally on the data almost surely,*

$$M_n^* \xrightarrow{d} \sup_{t\in[t_1,t_2]}\left[U_+^2(t)\right].$$

Based on Theorem 2.4, to calibrate the test, we simulate M_n^* by repeatedly generating samples of Gaussian random multipliers $\{\xi_{ij}\}$ while holding the observed data fixed. We then compare the empirical quantiles of these bootstrapped values M_n^* with our test statistic M_n.

2.2.2. *A simultaneous confidence band constructed by Gaussian multiplier bootstrap*

As mentioned in the Introduction, a simultaneous confidence band for $F_2 - F_1$ is needed for an auxiliary test that F_1 and F_2 do not cross. Such a confidence band can be constructed using the aforementioned Gaussian multiplier bootstrap approach. Here we briefly explain how a suitable band can be constructed based on the standardized difference between the NPMLEs and the true cdfs,

$$\sqrt{n}\tilde{\sigma}^{-\frac{1}{2}}(t,t)[\tilde{F}_2(t) - \tilde{F}_1(t) - \{F_2(t) - F_1(t)\}], \tag{11}$$

where $\tilde{\sigma}(t,t) = \sum_{j=1}^{2} (\tilde{W}_j^2/\kappa_j) \sum_{i=1}^{n_j} [(I_{X_{ij} \le t} - \tilde{F}_j(t))/w_{ij}]^2/n_j$. This is the counterpart of the equal precision confidence band in censored data (Nair 1984). It is more challenging to construct an EL-type band, as it requires solving for two roots at each t for the upper and lower bounds (see, e.g., McKeague and Zhao 2002); this is beyond the scope of the present article.

The asymptotic distribution of (11) is given as follows (see Appendix E for the proof).

THEOREM 2.5 *Suppose* $0 < F_j(t_1) < F_j(t_2) < 1$ *and* $\int_0^\infty w_j(u)^{-1}\, dF_j(u) < \infty$ *for* $j = 1, 2$. *Then*

$$\sup_{t \in [t_1, t_2]} \sqrt{n} \tilde{\sigma}^{-\frac{1}{2}}(t,t) \left| \tilde{F}_2(t) - \tilde{F}_1(t) - \{F_2(t) - F_1(t)\} \right| \xrightarrow{d} \sup_{t \in [t_1, t_2]} \left| U^B(t) \right|,$$

where $U^B(t)$ *is a mean-zero pinned Gaussian process with covariance function* $\mathrm{cov}(U^B(s), U^B(t)) = \sigma^B(s,t)/\sqrt{\sigma^B(s,s)\sigma^B(t,t)}$ *and* $\sigma^B(s,t) = \sum_{j=1}^{2} (W_j^2/\kappa_j) E[(I_{X_j \le s} - F_j(s))(I_{X_j \le t} - F_j(t))/w_j^2(X_j)]$.

Remark When $F_1 = F_2 = F_0$, $U^B(t) = U(t)$ and $\sigma^B(s,t) = \sigma(s,t)$, where $U(t)$ and $\sigma(s,t)$ are defined in Theorem 2.3.

The derivation utilizes the equality between (11) and

$$\tilde{\sigma}^{-\frac{1}{2}}(t,t) \left[\frac{\tilde{W}_2}{\sqrt{n_2}\sqrt{\kappa_2}} \sum_{i=1}^{n_2} \frac{I_{X_{i2} \le t} - F_2(t)}{w_{i2}} - \frac{\tilde{W}_1}{\sqrt{n_1}\sqrt{\kappa_1}} \sum_{i=1}^{n_1} \frac{I_{X_{i1} \le t} - F_1(t)}{w_{i1}} \right], \tag{12}$$

as noted in Appendix E. Based on (12), we can use a Gaussian multiplier bootstrap to construct the confidence band, as in the case of one-sample and two-sample tests for stochastic ordering. In particular, we can directly use $U_n^*(t)$ from Section 2.2.1. Bootstrap consistency of $U_n^*(t)$ can be established along similar lines using the proofs in Appendix D, except that $F_0(t)$ is replaced by $F_j(t)$ in quantities related the j-th sample. Then by the continuous mapping theorem, we have the following theorem of bootstrap consistency.

THEOREM 2.6 *Under the conditions of Theorem 2.5, conditionally on the data almost surely,*

$$\sup_{t \in [t_1, t_2]} |U_n^*(t)| \xrightarrow{d} \sup_{t \in [t_1, t_2]} \left| U^B(t) \right|.$$

From Theorem 2.6 we can construct an asymptotic $100(1-\alpha)\%$ confidence band for $F_2(t) - F_1(t)$ as follows: simulate $\sup_{t \in [t_1, t_2]} |U_n^*(t)|$ by repeatedly generating samples of Gaussian random multipliers $\{\xi_{ij}\}$ while holding the observed data fixed. We then denote the upper α quantile of these bootstrapped values as c_α^*, and the band is given by

$$\tilde{F}_2(t) - \tilde{F}_1(t) \pm n^{-1/2} c_\alpha^* \sqrt{\tilde{\sigma}(t,t)}$$

for $t \in [t_1, t_2]$.

3. Simulation study

In this section, we report the result of a simulation study for the one-sided tests. Results for the two-sided tests are similar. We investigate the performance of M_n in terms of accuracy and power.

3.1. *Accuracy*

We consider $F_1 = F_2 = Beta(4,3)$ and two scenarios for the weight functions: (A) $w_1(x) = x$ and $w_2(x) = \sqrt{x}$, (B) $w_1(x) = \sqrt{x}$ and $w_2(x) = x$. The scenarios are illustrated in Figure 1. From the figure, we can see that although the underlying F_1 and F_2 are the same, the observed distributions G_1 and G_2 are different due to size bias.

[Figure 1 here]

It would be interesting to see what happens if size bias is ignored; that is, one mistakes G_j as F_j. To this end, we compare M_n with its counterpart that sets $w_{ij} \equiv 1$ (see Remark 1 after Theorem 2.3), which is related to the one-sided two-sample Kolmogorov–Smirnov statistic. We denote this statistic by M_n^{ign}. Another statistic for comparison is the one-sided Wald-type statistic $\sup_{t\in[t_1,t_2]}[U_n^2(t)I_{U_n(t)\geq 0}]$ (see Section 2.2).

The size simulation results are given in Table 1. Note that the empirical significance levels of our EL test are close to the nominal level in all the cases considered. On the other hand, the Wald test is too conservative in Scenario B. As for the test ignoring size bias, its empirical significance levels are too large in Scenario A and too small in Scenario B. We conclude that the proposed EL test has better accuracy than the other tests.

[Table 1 here]

We have also compared our EL test with the Wald test in terms of computational cost, and we have found that they have very similar run times. This is because the major computational task lies in solving the estimating equations in (9) for each t, and both our EL and Wald statistics involve quantities from those estimating equations.

3.2. *Power comparisons*

In this section, we compare the small sample power of the proposed test with its counterpart ignoring size bias and the Wald test. Two models of underlying distribution functions are considered: (C) $F_1 = Beta(4,3)$ versus $F_2 = Beta(4,4)$, (D) $F_1 = Beta(3,5)$ versus $F_2 = Beta(3,7)$. For both models, we set $w_1(x) = \sqrt{x}$ and $w_2(x) = x$. The weight functions make the difference between G_1 and G_2 smaller than the difference between F_1 and F_2, as illustrated in Figure 2. As a result, the test ignoring size bias (i.e., comparing G_j instead of F_j) is expected to have lower power.

[Figure 2 here]

The power simulation results are summarized in Table 2. M_n outperforms the other tests in all the cases considered. The Wald test tends to have lower power. The much lower power of M_n^{ign} shows the importance of taking sampling bias into account. Similar results are obtained for the Gamma distribution (results available upon request). The proposed EL test has considerable advantages over competing approaches for testing stochastically ordered alternatives.

[Table 2 here]

4. Real data example

In this section we apply our test to the question of whether the distribution of blood alcohol concentration (BAC) of drivers involved in fatal car accidents depends on age. Size bias arises because drivers with higher BAC are more likely to be involved in accidents and have their BAC recorded.

We compare the BAC of two age groups: younger (< 30 years) and older (≥ 30 years) drivers. The pattern of sampling bias is different between these two groups, as discussed by Ramírez and Vidakovic (2010). This is due to the tendency of younger drivers to be more affected by alcohol, resulting in upweighted sampling at lower levels of BAC in the younger group. Ramírez and Vidakovic take the weight functions to be $w_y(x) = \sqrt{x}$ and $w_o(x) = x$ for the younger and older groups, respectively, although they admit that this choice is subjective (it is similar in spirit to the selection of the prior distribution in a Bayesian setting). We specify the younger-group weight function to take the more general form $w_y(x) = x^r$, where $r \in [0, 1]$.

The BAC data are available online from the Fatality Analysis Reporting System (FARS) of the U.S. National Highway Traffic Safety Administration. To ensure sample homogeneity, we restrict our analysis to whole blood test results of male drives involved in interstate highway accidents in California during 2009, criteria satisfied by 125 subjects. There are 67 younger and 58 older drivers. Although the empirical cdfs (see top panel of Figure 3) show there is no obvious difference between the two biased distributions (from which the data are observed), the NPMLEs for the underlying distribution functions (based on the weight functions used by Ramírez and Vidakovic) suggest that BAC in the younger group is stochastically larger than in the older group, see the bottom panel of Figure 3. We specified t_1 and t_2 to be the smallest and largest observations in the pooled sample, respectively.

[Figure 3 here]

Applying the one-sided EL test, the p-value is found to be an increasing function of r, see Figure 4 (solid line). For $r \leq 0.4$, the test shows significance at the 0.05 level; for $r \leq 0.5$, it shows significance at the 0.1 level. This indicates that the younger group has stochastically larger BAC than the older group over a reasonable range of weight functions. The one-sided Wald test (dashed line), on the other hand, is much more conservative, only showing significance at the 0.05 level for $r \leq 0.25$. The test ignoring size bias yields a very large p-value of 0.841, reflecting the fact that the two empirical cdfs almost overlap.

[Figure 4 here]

5. Discussion

We have developed a new test for stochastically ordered alternatives based on size-biased data, allowing the pattern of size bias to differ between the two samples. The test is formulated in terms of a maximally-selected local empirical likelihood statistic. A simulation study shows that the new test is more powerful than its counterpart ignoring size bias and an analogous Wald-type test. We applied our test to blood alcohol measurements in fatal driving accidents and found a more significant result than the Wald-type test and the test ignoring sampling bias.

We calibrate the proposed test using a Gaussian multiplier bootstrap approach. Other exchangeable bootstrap procedures (see, e.g., van der Vaart and Wellner 1996, Ch. 3.6) could also be considered. A computationally feasible adaptation of the nonparametric bootstrap for $U_n(t)$ can be defined by replacing ξ_{ij} in $U_n^*(t)$ (see Section 2.2.1) with $M_{n_j i} - 1$ ($i = 1, \ldots, n_j$, $j = 1, 2$), where $M_{n_j i}$ is the number of times that X_{ij} is redrawn from the original sample. Here note that despite resampling of the original observations, we keep $\hat{\sigma}^{-\frac{1}{2}}(t, t)$, $\hat{W}_j$ and $\hat{F}_0(t)$ intact. This is because computing them requires solving the estimating equations for each t, which could be time consuming if we repeat such computation for each bootstrap sample. The proposed multiplier bootstrap approach also avoids such recomputation. Although the aforementioned bootstrap procedures are asymptotically first-order equivalent, a detailed comparison of their higher-order properties (e.g., Hall 1992) would be needed for further insight.

Our key contribution is the development of the first EL-based test for ordered underlying distribution functions in biased sampling models. We envision the test to be useful in numerous applications involving length/size bias, especially in the biostatistical settings mentioned in the Introduction, but also in reliability engineering (Oluyede and George 2002), and marketing research (Nowell and Stanley 1991). One future direction is to extend our test to the multiplicative censorship model, which can be applied to prevalence cohort studies (Ning et al. 2013) with different rates of diagnoses among the groups (Walker et al. 2014). Another direction is to deal with the situation where, in addition to the samples observed from the G_j, we also have random samples observed from the F_j. One possible test for stochastic ordering in this case is to use a convex combination of the two statistics M_n and M_m^{ign}, where M_n is computed based on samples (of total size n) from G_1 and G_2 and M_m^{ign} based on samples (of total size m) from F_1 and F_2. Still another direction is to extend our approach to allow multiple samples with different weight functions acting on F_1 and F_2, just as Gill et al. (1988) do for a single underlying distribution.

Acknowledgements

The research of Hsin-wen Chang was partially supported by Ministry of Science and Technology of Taiwan under grant 104-2118-M-001-001. The work of Hammou El Barmi was supported by The City University of New York (PSC-CUNY 45 grant). The research of Ian McKeague was partially supported by NSF Grant DMS-1307838 and NIH Grant R01GM095722. The authors thank Cheng-Chen Tsai for computational support.

Appendix A. Numerator of the local EL ratio

We only consider the two-sample case; the one-sample case is similar. We first maximize

$$\log L(F_1, F_2) \cong \sum_{j=1}^{2}\sum_{i=1}^{n_j} \log p_{ij} - \sum_{j=1}^{2} n_j \log\left(\sum_{i=1}^{n_j} w_{ij} p_{ij}\right), \tag{A1}$$

where $p_{ij} = dF_j(X_{ij})$ and $w_{ij} = w_j(X_{ij})$, subject to the constraints

$$\sum_{i=1}^{n_j} p_{ij} = 1, \sum_{i=1}^{n_j} p_{ij}\left(I_{X_{ij}\le t} - F_0(t)\right) = 0, \text{ and } \sum_{i=1}^{n_j} p_{ij}\left(w_{ij} - W_j\right) = 0, \tag{A2}$$

for fixed W_j and $F_0(t)$, $j = 1, 2$. Here $\cong$ means up to a constant that does not depend on the p_{ij}s. This is similar to the usual empirical likelihood for estimating equations, except that now the likelihood is of a weighted form. By Lagrange multipliers, the optimum is found to occur at

$$p_{ij}\left(W_j, F_0(t)\right) = \frac{1}{n}\frac{1}{\kappa_j + \lambda_{1j}\left(I_{X_{ij}\le t} - F_0(t)\right) + \lambda_{2j}\left(w_{ij} - W_j\right)},$$

where $\lambda_{1j}, \lambda_{2j}$ satisfy

$$\sum_{i=1}^{n_j} p_{ij}\left(W_j, F_0(t)\right)\left(I_{X_{ij}\le t} - F_0(t)\right) = 0, \quad \sum_{i=1}^{n_j} p_{ij}\left(W_j, F_0(t)\right)\left(w_{ij} - W_j\right) = 0.$$

A profile log-likelihood can be obtained by plugging $p_{ij}(W_j, F_0(t))$ into (A1):

$$\sum_{j=1}^{2}\sum_{i=1}^{n_j} \log p_{ij}\left(W_j, F_0(t)\right) - \sum_{j=1}^{2} n_j \log W_j.$$

This profile log-likelihood is then maximized over $(W_1, W_2, F_0(t))$. The optimal solution $(\hat{W}_1, \hat{W}_2, \hat{F}_0(t))$ satisfies the relation $\lambda_{2j} = \kappa_j/\hat{W}_j$ and $\lambda_{11} = -\lambda_{12} \equiv \hat{\lambda}$. As a result, the optimal p_{ij}s are given by

$$\begin{aligned}\hat{p}_{ij} =& \frac{1}{n}\frac{1}{\hat{\lambda}\Delta_j\left(I_{X_{ij}\le t} - \hat{F}_0(t)\right) + \frac{\kappa_j w_{ij}}{\hat{W}_j}}\\ =& \frac{1}{n\hat{\eta}_{ij}}\frac{1}{1 + \hat{\lambda}\Delta_j g_{1ij}(\hat{F}_0(t), \hat{W}_j)},\end{aligned} \tag{A3}$$

where $\Delta_j = 1$ for $j = 1$ and -1 for $j = 2$, $\hat{\eta}_{ij} = (\kappa_j w_{ij})/\hat{W}_j$, $g_{1ij}(\hat{F}_0(t), \hat{W}_j) = (I_{X_{ij}\le t} - \hat{F}_0(t))/\hat{\eta}_{ij}$, and $(\hat{W}_1, \hat{W}_2, \hat{\lambda}, \hat{F}_0(t))$ satisfy (9). Here the dependence of the solution on t is suppressed.

Appendix B. Proof of Theorem 2.3

The events $\hat{\lambda} \le 0$ and $\tilde{F}_1(t) \le \tilde{F}_2(t)$ coincide, which can be seen using the relation between the sign of the Lagrange multiplier and the location of the global optimum. Then

$$-2\log\mathcal{R}(t) = 2\sum_{j=1}^{2}\sum_{i=1}^{n_j} \log\left(1 + \hat{\lambda}\Delta_j g_{1ij}(\hat{F}_0(t), \hat{W}_j)\right) I_{\hat{\lambda}\le 0}, \tag{B1}$$

so we can re-write the estimating equations in (9) as

$$Q_{1j}(\hat{F}_0(t), \hat{W}_j, \hat{\lambda}) = 0, \qquad Q_{2j}(\hat{F}_0(t), \hat{W}_j, \hat{\lambda}) = 0, \tag{B2}$$

where

$$Q_{1j}(\hat{F}_0(t), \hat{W}_j, \hat{\lambda}) \equiv \frac{1}{n}\sum_{i=1}^{n_j}\{g_{1ij}(\hat{F}_0(t), \hat{W}_j)/[1 + \Delta_j\hat{\lambda}g_{1ij}(\hat{F}_0(t), \hat{W}_j)]\},$$
$$Q_{2j}(\hat{F}_0(t), \hat{W}_j, \hat{\lambda}) \equiv \frac{1}{n}\sum_{i=1}^{n_j}\{g_{2ij}(\hat{F}_0(t), \hat{W}_j)/[1 + \Delta_j\hat{\lambda}g_{1ij}(\hat{F}_0(t), \hat{W}_j)]\},$$

and $g_{2ij}(\hat{F}_0(t), \hat{W}_j) = (w_{ij} - \hat{W}_j)/\hat{\eta}_{ij}$.

By (B2), we can show that $\hat{F}_0(t) - F_0(t)$, $\hat{\lambda}$ and $\hat{W}_j - W_j$ are $O_p(n^{-1/2})$ (see Appendix C). Here and in the sequel, the asymptotic o_p, o, O_p and O terms hold uniformly for $t \in [t_1, t_2]$. Based on these asymptotic orders, we apply Taylor's theorem to (B1) and get

$$-2\log \mathcal{R}(t) = 2\sum_{j=1}^{2}\left[\hat{\lambda}\Delta_j \sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t), \hat{W}_j) - \frac{\hat{\lambda}^2}{2}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)\right] I_{\hat{\lambda}\le 0} + o_p(1). \tag{B3}$$

We also expand $Q_{1j}(\hat{F}_0(t), \hat{W}_j, \hat{\lambda})$ (for $j = 1, 2$) around $(\hat{F}_0(t), \hat{W}_j, 0)$ and get

$$\begin{aligned} 0 =& Q_{1j}(\hat{F}_0(t), \hat{W}_j, 0) + \frac{\partial Q_{1j}(\hat{F}_0(t), \hat{W}_j, 0)}{\partial \lambda}(\hat{\lambda} - 0) + o_p(|\hat{\lambda}|) \\ =& \frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t), \hat{W}_j) - \frac{\Delta_j}{n}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)\hat{\lambda} + o_p(n^{-\frac{1}{2}}), \end{aligned}$$

which implies

$$\hat{\lambda} = \left[\frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)\right]^{-1}\left[\sum_{j=1}^{2}\Delta_j \frac{\hat{W}_j}{n_j}\sum_{i=1}^{n_j}\frac{I_{X_{ij}\le t} - \hat{F}_0(t)}{w_{ij}}\right] + o_p(n^{-\frac{1}{2}}) \tag{B4}$$

and $\hat{\lambda}\Delta_j \sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t), \hat{W}_j) = \hat{\lambda}^2 \sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j) + o_p(1)$. Substituting the latter into (B3) gives

$$-2\log \mathcal{R}(t) = 2\sum_{j=1}^{2}\left[\frac{\hat{\lambda}^2}{2}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)\right] I_{\hat{\lambda}\le 0} + o_p(1).$$

This and (B4) imply

$$-2\log \mathcal{R}(t) = U_n^2(t) I_{U_n(t)\ge 0} + o_p(1), \tag{B5}$$

where $U_n(t)$ is defined as

$$\left[\frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j))\right]^{-\frac{1}{2}}\left[\sum_{j=1}^{2}(-\Delta_j)\frac{\hat{W}_j}{\sqrt{n_j}\sqrt{\kappa_j}}\sum_{i=1}^{n_j}\frac{I_{X_{ij}\le t} - \hat{F}_0(t)}{w_{ij}}\right]. \tag{B6}$$

Based on (B5), we can obtain the limiting distribution of $-2\log \mathcal{R}(t)$ by studying $U_n(t)$. We begin by finding the weak convergence of the second term in (B6), where we can replace $\hat{W}_j$ and $\hat{F}_0(t)$ by W_j and $F_0(t)$, respectively, because $\hat{F}_0(t) - F_0(t) = O_p(n^{-1/2})$ and $\hat{W}_j - W_j = O_p(n^{-1/2})$. By the Donsker theorem, $(W_j/\sqrt{n_j\kappa_j}) \times \sum_{i=1}^{n_j}[(I_{X_{ij}\le t} - F_0(t))/w_{ij}]$ converges in distribution in $l^\infty([t_1, t_2])$ to a Gaussian process with zero mean and covariance function

$$\frac{W_j^2}{\kappa_j} E\left(\frac{I_{X_{ij}\le s} - F_0(s)}{w_j(X_{ij})}\frac{I_{X_{ij}\le t} - F_0(t)}{w_j(X_{ij})}\right);$$

the relevant class of functions is shown to be G_j-Donsker in Appendix D. Therefore, by independence between the two samples and the continuous mapping theorem, $\sum_{j=1}^{2}(-\Delta_j)(W_j/\sqrt{n_j\kappa_j})\sum_{i=1}^{n_j}[(I_{X_{ij}\le t} - F_0(t))/w_{ij}]$ converges in distribution in $l^\infty([t_1,t_2])$ to a Gaussian process with zero mean and covariance function

$$\sigma(s,t) \equiv \sum_{j=1}^{2}\frac{W_j^2}{\kappa_j}E\left(\frac{I_{X_{ij}\le s} - F_0(s)}{w_j(X_{ij})}\frac{I_{X_{ij}\le t} - F_0(t)}{w_j(X_{ij})}\right).$$

On the other hand,

$$\frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t),\hat{W}_j)) = \sigma(t,t) + o_p(1) \tag{B7}$$

by the Glivenko–Cantelli theorem, $\hat{F}_0(t) - F_0(t) = O_p(n^{-1/2})$ and $\hat{W}_j - W_j = O_p(n^{-1/2})$. Then by Slutsky's lemma and (B6), we have $U_n(t)\xrightarrow{d}U(t)$ in $l^\infty([t_1,t_2])$, where $U(t)$ is a mean-zero Gaussian process with

$$\mathrm{cov}(U(s),U(t)) = \frac{\sigma(s,t)}{\sqrt{\sigma(s,s)}\sqrt{\sigma(t,t)}}.$$

This, (B5) and the continuous mapping theorem imply $-2\log\mathcal{R}(t)\xrightarrow{d}U_+^2(t)$ in $l^\infty([t_1,t_2])$. Then applying the continuous mapping theorem again, we obtain the desired result.

Appendix C. Asymptotic orders of $\hat{F}_0(t)$, $\hat{\lambda}$ and $\hat{W}_j$

First we establish the asymptotic orders of $\hat{\lambda}\hat{W}_j$ and $\hat{F}_0(t)$ uniformly in t. Let $\hat{\lambda} = \theta|\hat{\lambda}|$ such that $|\theta| = 1$ and let $\theta_j = \Delta_j\theta$. Denote $\Delta_j\hat{\lambda}g_{1ij}(\hat{F}_0(t),\hat{W}_j)$ by ζ_{ij}. Substituting $1/(1+\zeta_{ij}) = 1 - \zeta_{ij}/(1+\zeta_{ij})$ into $\theta_j Q_{1j}(\hat{F}_0(t),\hat{W}_j,\hat{\lambda}) = 0$, we get

$$\begin{aligned}
0 &= \frac{1}{n}\theta_j\sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t),\hat{W}_j)\left(1 - \frac{\zeta_{ij}}{1+\zeta_{ij}}\right)\\
&= \theta_j\frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t),\hat{W}_j) - \frac{|\hat{\lambda}|}{n}\sum_{i=1}^{n_j}\frac{\theta_j g_{1ij}(\hat{F}_0(t),\hat{W}_j)g_{1ij}(\hat{F}_0(t),\hat{W}_j)\theta_j}{1+\zeta_{ij}}\\
&= \theta_j\frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t),\hat{W}_j) - |\hat{\lambda}|\frac{1}{n}\sum_{i=1}^{n_j}\frac{g_{1ij}^2(\hat{F}_0(t),\hat{W}_j)}{1+\zeta_{ij}}.
\end{aligned} \tag{C1}$$

Note that $1+\zeta_{ij} > 0$ by $\hat{p}_{ij} > 0$ for all i, j. Thus we can obtain the following (in)equalities:

$$\begin{aligned}
|\hat{\lambda}|\frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}^2(\hat{F}_0(t), \hat{W}_j) &\leq |\hat{\lambda}|\frac{1}{n}\sum_{i=1}^{n_j} \frac{g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)}{1+\zeta_{ij}}\left(1+\max_{1\leq i\leq n_j}\zeta_{ij}\right) \\
&\leq |\hat{\lambda}|\frac{1}{n}\sum_{i=1}^{n_j} \frac{g_{1ij}^2(\hat{F}_0(t), \hat{W}_j)}{1+\zeta_{ij}}\left(1+|\hat{\lambda}|\max_{1\leq i\leq n_j}\left|g_{1ij}(\hat{F}_0(t), \hat{W}_j)\right|\right) \\
&= \theta_j\frac{1}{n}\sum_{i=1}^{n_j} g_{1ij}(\hat{F}_0(t), \hat{W}_j)\left(1+|\hat{\lambda}|\max_{1\leq i\leq n_j}\left|g_{1ij}(\hat{F}_0(t), \hat{W}_j)\right|\right),
\end{aligned}$$

where the last equality follows from (C1). This implies

$$|\hat{\lambda}|\hat{W}_j S_j(\hat{F}_0(t)) \leq \theta_j \bar{g}_{1j}(\hat{F}_0(t))\left(1+|\hat{\lambda}|\hat{W}_j Z_j(\hat{F}_0(t))\right), \tag{C2}$$

where $S_j(\hat{F}_0(t)) = 1/n \times \sum_{i=1}^{n_j}[(I_{X_{ij}\leq t} - \hat{F}_0(t))/(\kappa_j w_{ij})]^2$, $\bar{g}_{1j}(\hat{F}_0(t)) = 1/n \times \sum_{i=1}^{n_j}$ $[(I_{X_{ij}\leq t} - \hat{F}_0(t))/(\kappa_j w_{ij})]$, and $Z_j(\hat{F}_0(t)) = \max_{1\leq i\leq n_j}\left|(I_{X_{ij}\leq t} - \hat{F}_0(t))/(\kappa_j w_{ij})\right|$. Using the assumption $\int_0^\infty w_j(u)^{-1}\,dF_0(u) < \infty$ (i.e., $\int_0^\infty w_j(u)^{-2}\,dG_j(u) < \infty$), we can show the uniform convergence of $S_j(\hat{F}_0(t))$ and $\bar{g}_{1j}(\hat{F}_0(t))$ by the Glivenko–Cantelli theorem and the Donsker theorem (the relevant G_j-Donsker condition is checked in Appendix D), leading to

$$S_j(\hat{F}_0(t)) = \frac{1}{\kappa_j}E\left(\frac{I_{X_{ij}\leq t} - \hat{F}_0(t)}{w_j(X_{ij})}\right)^2 + o(1) \tag{C3}$$

a.s. and

$$\bar{g}_{1j}(\hat{F}_0(t)) = \left(F_0(t) - \hat{F}_0(t)\right)\frac{1}{n_j}\sum_{i=1}^{n_j}\frac{1}{w_{ij}} + O_p(n^{-\frac{1}{2}}). \tag{C4}$$

As for $Z_j(\hat{F}_0(t))$, we can bound it by

$$\frac{2}{\kappa_j}\max_{1\leq i\leq n_j}\left|\frac{1}{w_{ij}}\right| = O(1)o(n^{\frac{1}{2}}) = o(n^{\frac{1}{2}}), \tag{C5}$$

a.s., where the first $o(n^{1/2})$ order is obtained by Lemma 11.2 of Owen (2001). From these uniform convergence results and (C2), we have

$$|\hat{\lambda}|W_1\frac{S_1(\hat{F}_0(t))}{1+|\hat{\lambda}|W_1 Z_1(\hat{F}_0(t))} \leq \theta_1\left(F_0(t) - \hat{F}_0(t)\right)\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{w_{i1}} + O_p(n^{-\frac{1}{2}}), \tag{C6}$$

$$|\hat{\lambda}|W_2\frac{S_2(\hat{F}_0(t))}{1+|\hat{\lambda}|W_2 Z_2(\hat{F}_0(t))} \leq -\theta_1\left(F_0(t) - \hat{F}_0(t)\right)\frac{1}{n_2}\sum_{i=1}^{n_2}\frac{1}{w_{i2}} + O_p(n^{-\frac{1}{2}}). \tag{C7}$$

Multiplying (C6) by $n_1/\sum_{i=1}^{n_1}(1/w_{i1})$ and (C7) by $n_2/\sum_{i=1}^{n_2}(1/w_{i2})$, adding up the two terms, and using the fact that $\sum_{i=1}^{n_j}(1/w_{ij})/n_j = O(1)$ a.s. (by the SLLN), we can bound

$$\sum_{j=1}^{2} \frac{S_j}{Z_j + \frac{1}{|\hat{\lambda}|\hat{W}_j}} \frac{n_j}{\sum_{i=1}^{n_j} \frac{1}{w_{ij}}}$$

above by an $O_p(n^{-1/2})$ term. This and (C3) imply $Z_j + 1/(|\hat{\lambda}|\hat{W}_j)$ must grow faster than $n^{1/2}$ (in probability). Then by (C5) we obtain

$$\hat{\lambda}\hat{W}_j = O_p(n^{-\frac{1}{2}}) \tag{C8}$$

for $j = 1, 2$. This, together with (C3) and (C5), imply that the l.h.s. of (C6) and (C7) are both $O_p(n^{-1/2})$. Then (C6) and (C7) imply that both $\theta_1(\hat{F}_0(t) - F_0(t))$ and $\theta_1(F_0(t) - \hat{F}_0(t))$ are bounded above by $O_p(n^{-1/2})$ terms. And thus

$$\hat{F}_0(t) - F_0(t) = O_p(n^{-\frac{1}{2}}).$$

Next we establish the order of $\hat{W}_j$ and $\hat{\lambda}$. Let $\boldsymbol{g}_{ij}(\hat{F}_0(t), W_j) = [(w_{ij} - W_j)/(\kappa_j w_{ij}), (I_{X_{ij} \le t} - \hat{F}_0(t))/(\kappa_j w_{ij})]^T$ and let $\hat{\boldsymbol{\lambda}} = [0, \hat{\lambda}]^T = \boldsymbol{\theta}\|\hat{\boldsymbol{\lambda}}\|$ such that $\|\boldsymbol{\theta}\| = 1$ and let $\boldsymbol{\theta}_j = \Delta_j \boldsymbol{\theta}$. Then (B2) and $\sum_{i=1}^{n_j} \hat{p}_{ij} = 1$ imply

$$\begin{bmatrix} \hat{W}_j - W_j \\ 0 \end{bmatrix} = \frac{1}{n} \sum_{i=1}^{n_j} \frac{\hat{W}_j \boldsymbol{g}_{ij}(\hat{F}_0(t), W_j)}{1 + \Delta_j \hat{W}_j \hat{\boldsymbol{\lambda}}^T \boldsymbol{g}_{ij}(\hat{F}_0(t), W_j)}. \tag{C9}$$

By (C8) and a similar reasoning as in (C5), we have that $\max_{1 \le i \le n_j} \|\Delta_j \hat{W}_j \hat{\boldsymbol{\lambda}}^T \boldsymbol{g}_{ij}(\hat{F}_0(t), W_j)\| \le \|\Delta_j \hat{W}_j \hat{\boldsymbol{\lambda}}^T\| \max_{1 \le i \le n_j} \|\boldsymbol{g}_{ij}(\hat{F}_0(t), W_j)\| = O_p(n^{-1/2}) o(n^{1/2}) = o_p(1)$. Then we can expand the r.h.s. of (C9) as

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n_j} \left\{ \hat{W}_j \boldsymbol{g}_{ij}(\hat{F}_0(t), W_j) \left[1 - \Delta_j \hat{W}_j \hat{\boldsymbol{\lambda}}^T \boldsymbol{g}_{ij}(\hat{F}_0(t), W_j) + o_p\left(\|\hat{W}_j \hat{\boldsymbol{\lambda}}^T \boldsymbol{g}_{ij}(\hat{F}_0(t), W_j)\| \right) \right] \right\} \\
&\quad = \hat{W}_j \bar{\boldsymbol{g}}_j(\hat{F}_0(t), W_j) - \Delta_j \hat{W}_j^2 \boldsymbol{S}_j(\hat{F}_0(t), W_j) \hat{\boldsymbol{\lambda}} \\
&\qquad + \hat{W}_j \bar{\boldsymbol{g}}_j(\hat{F}_0(t), W_j) o_p\left(\max_{1 \le i \le n_j} \|\hat{W}_j \hat{\boldsymbol{\lambda}}^T \boldsymbol{g}_{ij}(\hat{F}_0(t), W_j)\| \right),
\end{aligned}$$

where $\bar{\boldsymbol{g}}_j(\hat{F}_0(t), W_j) = (1/n) \sum_{i=1}^{n_j} \boldsymbol{g}_{ij}(\hat{F}_0(t), W_j)$ and

$$\boldsymbol{S}_j(\hat{F}_0(t), W_j) = \frac{1}{n} \sum_{i=1}^{n_j} [\boldsymbol{g}_{ij}(\hat{F}_0(t), W_j) \boldsymbol{g}_{ij}^T(\hat{F}_0(t), W_j)].$$

By a similar reasoning as in (C3) and (C4), we can show that $\bar{\boldsymbol{g}}_j(\hat{F}_0(t), W_j) = O_p(n^{-1/2})$ and $\boldsymbol{S}_j(\hat{F}_0(t), W_j) = O(1)$ a.s. Then

$$|\hat{W}_j - W_j| \le \hat{W}_j \|\bar{\boldsymbol{g}}_j(\hat{F}_0(t), W_j)\| + \|\boldsymbol{S}_j(\hat{F}_0(t), W_j)\| |\hat{\lambda}| \hat{W}_j^2 + \hat{W}_j \|\bar{\boldsymbol{g}}_j(\hat{F}_0(t), W_j)\| o_p(1).$$

This and (C8) imply $1 - W_j/\hat{W}_j = O_p(n^{-1/2})$, which gives $\hat{W}_j - W_j = O_p(n^{-\frac{1}{2}})$. Then by (C8) again, we obtain $\hat{\lambda} = O_p(n^{-\frac{1}{2}})$.

Appendix D. Gaussian multiplier bootstrap consistency

To show bootstrap consistency of M_n^*, we start with $U_n^*(t)$. It is easier to first obtain bootstrap consistency of

$$U_n^{**}(t) = \sigma(t,t)^{-\frac{1}{2}} \left[\sum_{j=1}^{2} (-\Delta_j) \frac{W_j}{\sqrt{n_j}\sqrt{\kappa_j}} \sum_{i=1}^{n_j} \xi_{ij} \frac{I_{X_{ij} \le t} - F_0(t)}{w_{ij}} \right].$$

Next we establish asymptotic equivalence of $U_n^*(t)$ and $U_n^{**}(t)$, conditionally on the data almost surely. This implies bootstrap consistency of $U_n^*(t)$. Lastly, by the continuous mapping theorem, we obtain the desired result for M_n^*.

To prove bootstrap consistency of $U_n^{**}(t)$, we make use of the multiplier central limit theorem. Specifically, we first show that for the j-th sample, the class of functions

$$\mathcal{F}_j = \left\{ f_{jt}(x) \equiv \sigma(t,t)^{-\frac{1}{2}} \left[\frac{I_{x \le t} - F_0(t)}{w_j(x)} \right], t \in [t_1, t_2] \right\}$$

is G_j-Donsker. This follows by Donsker preservation (see, e.g., Kosorok 2008, Corollary 9.32) since $\sigma(t,t)$ is bounded away from zero on $[t_1, t_2]$, the Donsker property is preserved under addition of classes of functions, and the class

$$\{x \mapsto I_{x \le t}/w_j(x), t \in [t_1, t_2]\}$$

is G_j-Donsker under the condition $Ew_j(X_j)^{-2} = G_j w_j(x)^{-2} < \infty$. The latter can be seen by adapting the proof of the classical Donsker theorem using bracketing entropy (e.g., van der Vaart (2000), page 271) as follows. Let $\epsilon > 0$ and choose $0 = t_0 < t_1 < \ldots < t_k = \infty$ to have the property that $G_j[w_j(x)^{-2} I(t_{i-1} < x < t_i)] < \epsilon^2$ for each i. Then the brackets $[I(x \le t_{i-1})/w_j(x), I(x < t_i)/w_j(x)]$ have $L_2(G_j)$-size given by the square-root of $G_j[\{I(x < t_i) - I(x \le t_{i-1})\}/w_j(x)]^2 = O(\epsilon^2)$. It follows that the bracketing numbers are of the polynomial order $1/\epsilon$, and we conclude that $\mathcal{F}_j$ is G_j-Donsker.

Secondly, we have $G_j \|f_{jt} - G_j f_{jt}\|_{\mathcal{F}_j}^2 < \infty$ by the assumption $G_j w_j(x)^{-2} < \infty$. These results and the multiplier central limit theorem (see, e.g., van der Vaart and Wellner 1996, Theorem 2.9.7) then imply that, conditionally on the data almost surely, $\sum_{i=1}^{n_j} \xi_{ij} f_{jt}(X_{ij})/\sqrt{n_j}$ converges in distribution in $l^\infty([t_1, t_2])$ to a Gaussian process with mean-zero and covariance function

$$[\sigma(s,s)\sigma(t,t)]^{-\frac{1}{2}} E\left(\frac{I_{X_{ij} \le s} - F_0(s)}{w_j(X_{ij})} \frac{I_{X_{ij} \le t} - F_0(t)}{w_j(X_{ij})} \right).$$

Finally, by independence between the two samples and the continuous mapping theorem, we have that $U_n^{**}(t)$ converges in distribution in $l^\infty([t_1, t_2])$ to $U(t)$, conditionally on the data almost surely.

After showing bootstrap consistency of $U_n^{**}(t)$, now we show that $U_n^*(t)$ is asymptotically equivalent to $U_n^{**}(t)$, conditionally on the data almost surely. The task can be broken into three parts. Firstly, we show the (conditional) asymptotic equivalence of $U_n^*(t)\sqrt{\tilde{\sigma}^*(t,t)}$ and $U_n^{**}(t)\sqrt{\sigma(t,t)}$. The second task then involves establishing consistency of the bootstrap estimator $\tilde{\sigma}^*(t,t)$ for $\sigma(t,t)$. The final step is to use the continuous mapping theorem to get (conditional) consistency of $\tilde{\sigma}^{*-1/2}(t,t)$ for $\sigma^{-1/2}(t,t)$, and by Slutsky's lemma we get the desired result. We elaborate the first two tasks as follows.

For the first task, we want to show for all $\varepsilon > 0$,

$$P\left(\sup_{t\in[t_1,t_2]}\left|U_n^*(t)\sqrt{\tilde{\sigma}^*(t,t)} - U_n^{**}(t)\sqrt{\sigma(t,t)}\right| > \varepsilon\,\middle|\,X_{11}, X_{21}, \ldots, X_{12}, X_{22}, \ldots\right) \xrightarrow{a.s.} 0,$$

which follows by Chebyshev's inequality and

$$\begin{aligned}
&E\left(\sup_{t\in[t_1,t_2]}\left|U_n^*(t)\sqrt{\tilde{\sigma}^*(t,t)} - U_n^{**}(t)\sqrt{\sigma(t,t)}\right|^2\,\middle|\,X_{11}, X_{21}, \ldots, X_{12}, X_{22}, \ldots\right)\\
&\le \sum_{j=1}^{2}\frac{1}{n_j\kappa_j}\sum_{i=1}^{n_j}\frac{E\xi_{ij}^2}{w_{ij}^2}\sup_{t\in[t_1,t_2]}\left|\tilde{W}_j\left(I_{X_{ij}\le t} - \tilde{F}_j(t)\right) - W_j\left(I_{X_{ij}\le t} - F_0(t)\right)\right|^2\\
&= \sum_{j=1}^{2}\frac{1}{n_j\kappa_j}\sum_{i=1}^{n_j}\frac{1}{w_{ij}^2}o(1) \longrightarrow 0,
\end{aligned}$$

a.s., where the $o(1)$ is due to the strong consistency of $\tilde{W}_j$ and $\tilde{F}_j(t)$ for W_j and $F_0(t)$ under H_0, respectively (see the Remark in Appendix E), uniformly in $t \in [t_1, t_2]$.

For the second task of showing bootstrap consistency of $\tilde{\sigma}^*(t,t)$ for $\sigma(t,t)$, we use results of bootstrap for Glivenko–Cantelli classes. In detail, note $\tilde{\sigma}^*(t,t)$ can be separated into two terms:

$$\tilde{\sigma}^*(t,t) = \sum_{j=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j}\left(V_{ij}^* - \bar{V}_j^*\right)^2 = \sum_{j=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j}V_{ij}^{*2} - \sum_{j=1}^{2}\bar{V}_j^{*2}, \tag{D1}$$

where $V_{ij}^* = \xi_{ij}\tilde{W}_j[I_{X_{ij}\le t} - \tilde{F}_j(t)]/(\sqrt{\kappa_j}w_{ij})$ and $\bar{V}_j^* = \sum_{i=1}^{n_j}V_{ij}^*/n_j$. It is easier to start with the second term—using results from bootstrap for Glivenko–Cantelli classes (see, e.g., Kosorok 2008, Theorem 10.13), $\bar{V}_j^*$ converges in probability to 0 conditionally on the data almost surely, by strong consistency of $\tilde{W}_j$ and G_j-Glivenko–Cantelli of $\mathcal{F}_j$ (due to its G_j-Donsker property established earlier in this section), for $j = 1, 2$. Then by the continuous mapping theorem and independence between the two samples, we have that $\sum_{j=1}^{2}\bar{V}_j^{*2}$ converges in probability to 0 conditionally on the data almost surely.

As for the first term in (D1), $\sum_{j=1}^{2}\sum_{i=1}^{n_j}V_{ij}^{*2}/n_j$, we now show its (conditional) consistency for $\sigma(t,t)$. It is easier to begin with $\sum_{j=1}^{2}\sum_{i=1}^{n_j}V_{ij}^{**2}/n_j$, where

$$V_{ij}^{**} = \xi_{ij}\frac{W_j}{\sqrt{\kappa_j}}\frac{\left[I_{X_{ij}\le t} - F_0(t)\right]}{w_{ij}}.$$

We can show that conditionally on the data,

$$\sum_{j=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j}V_{ij}^{2**} - \sum_{j=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j}\xi_{ij}^2\frac{W_j^2}{\kappa_j}E\left\{\frac{I_{X_j\le t} - F_0(t)}{w_j(X_j)}\right\}^2 \longrightarrow 0 \tag{D2}$$

uniformly over $t \in [t_1, t_2]$ a.s. The second term in (D2) is strongly consistent for $\sigma(t,t)$ uniformly over $t \in [t_1, t_2]$ a.s., so (D2) implies (conditionally) consistency of $\sum_{j=1}^{2}\sum_{i=1}^{n_j} V_{ij}^{**2}/n_j$ for $\sigma(t,t)$. To obtain (D2), we just need to show that, for the j-th sample, the class of functions

$$\left\{ h_{jt}(x) \equiv \left[\frac{I_{x\le t} - F_0(t)}{w_j(x)}\right]^2, t \in [t_1, t_2] \right\}$$

is G_j-Glivenko–Cantelli (see, e.g., Kosorok 2008, Theorem 10.13). This follows by Glivenko–Cantelli preservation under addition of classes of functions, and the class

$$\{x \mapsto I_{x\le t}^2/w_j^2(x) = I_{x\le t}/w_j^2(x), t \in [t_1, t_2]\}$$

is G_j-Glivenko–Cantelli under the condition $Ew_j(X_j)^{-2} = G_j w_j(x)^{-2} < \infty$. The latter can be seen by a similar reasoning when proving G_j-Donsker of $\mathcal{F}_j$, with the brackets in $L_1(G_j)$ instead.

After showing bootstrap consistency of $\sum_{j=1}^{2}\sum_{i=1}^{n_j} V_{ij}^{**2}/n_j$, finally we show that $\sum_{j=1}^{2}\sum_{i=1}^{n_j} V_{ij}^{*2}/n_j$ is asymptotically equivalent to $\sum_{j=1}^{2}\sum_{i=1}^{n_j} V_{ij}^{**2}/n_j$, conditionally on the data almost surely. That is, for all $\varepsilon > 0$,

$$P\left(\sup_{t\in[t_1,t_2]}\left|\sum_{j=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j} V_{ij}^{**2} - \sum_{j=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j} V_{ij}^{*2}\right| > \varepsilon \middle| X_{11}, X_{21}, \ldots, X_{12}, X_{22}, \ldots\right) \xrightarrow{a.s.} 0,$$

which follows by Markov's inequality and

$$\begin{aligned}
&E\left(\sup_{t\in[t_1,t_2]}\left|\sum_{j=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j} V_{ij}^{**2} - \sum_{j=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j} V_{ij}^{*2}\right| \middle| X_{11}, X_{21}, \ldots, X_{12}, X_{22}, \ldots\right)\\
\le &\sum_{j=1}^{2}\frac{1}{n_j\kappa_j}\sum_{i=1}^{n_j}\frac{E\xi_{ij}^2}{w_{ij}^2}\sup_{t\in[t_1,t_2]}\left|\tilde{W}_j^2\left(I_{X_{ij}\le t} - \tilde{F}_j(t)\right)^2 - W_j^2\left(I_{X_{ij}\le t} - F_0(t)\right)^2\right|\\
= &\sum_{j=1}^{2}\frac{1}{n_j\kappa_j}\sum_{i=1}^{n_j}\frac{1}{w_{ij}^2}o(1) \longrightarrow 0,
\end{aligned}$$

a.s., where the $o(1)$ is due to the strong consistency of $\tilde{W}_j$ and $\tilde{F}_j(t)$ for W_j and $F_0(t)$ under H_0, respectively (see the Remark in Appendix E).

Appendix E. Proof of Theorem 2.5

We begin with showing (11) equals (12). It is true because for $j = 1, 2$,

$$
\begin{aligned}
&\sqrt{n}\tilde{\sigma}^{-\frac{1}{2}}(t,t)\left[\tilde{F}_j(t) - F_j(t)\right] \\
=&\tilde{\sigma}^{-\frac{1}{2}}(t,t)\left[\frac{\sqrt{n_j}}{\sqrt{\kappa_j}}\frac{\tilde{W}_j}{n_j}\sum_{i=1}^{n_j}\frac{I_{X_{ij}\le t}}{w_{ij}} - \frac{\sqrt{n_j}}{\sqrt{\kappa_j}}\frac{\tilde{W}_j}{\tilde{W}_j}F_j(t)\right] \\
=&\tilde{\sigma}^{-\frac{1}{2}}(t,t)\frac{\sqrt{n_j}}{\sqrt{\kappa_j}}\tilde{W}_j\left[\frac{1}{n_j}\sum_{i=1}^{n_j}\frac{I_{X_{ij}\le t}}{w_{ij}} - \frac{1}{n_j}\sum_{i=1}^{n_j}\frac{1}{w_{ij}}F_j(t)\right] \\
=&\tilde{\sigma}^{-\frac{1}{2}}(t,t)\left[\frac{\tilde{W}_j}{\sqrt{n_j}\sqrt{\kappa_j}}\sum_{i=1}^{n_j}\frac{I_{X_{ij}\le t} - F_j(t)}{w_{ij}}\right], \qquad \text{(E1)}
\end{aligned}
$$

where the first equality follows by definition of $\tilde{F}_j(t)$ (see beginning of Section 2.1), the second equlity follows by definition of $\tilde{W}_j$, and the third is just re-arranging the terms. By Donsker theorem, $(1/\sqrt{n_j\kappa_j}) \times \sum_{i=1}^{n_j}[(I_{X_{ij}\le t} - F_j(t))/w_{ij}]$ in (E1) converges weakly in $l^\infty([t_1, t_2])$ to a Gaussian process with zero mean and covariance function

$$
\frac{1}{\kappa_j}E\left(\frac{I_{X_{ij}\le s} - F_j(s)}{w_j(X_{ij})}\frac{I_{X_{ij}\le t} - F_j(t)}{w_j(X_{ij})}\right);
$$

the relevant class of functions is shown to be G_j-Donsker in Appendix D (with $F_0(t)$ replaced by $F_j(t)$ in quantities related the j-th sample). By the Central Limit Theorem and the condition $Ew_j(X_j)^{-2} = G_j w_j(x)^{-2} < \infty$, $\tilde{W}_j = n_j/\sum_{i=1}^{n_j}(1/w_{ij})$ can be shown to be $W_j + O_p(n^{-1/2})$. These results give the limiting distribution of the term in bracket in (E1): a Gaussian process with zero mean and covariance function

$$
\frac{W_j^2}{\kappa_j}E\left(\frac{I_{X_{ij}\le s} - F_j(s)}{w_j(X_{ij})}\frac{I_{X_{ij}\le t} - F_j(t)}{w_j(X_{ij})}\right)
$$

in $l^\infty([t_1, t_2])$. This, along with the independence between the two samples and the continuous mapping theorem, implies the term in bracket in (12) converges in distribution in $l^\infty([t_1, t_2])$ to a Gaussian process with zero mean and covariance function

$$
\sigma^B(s,t) \equiv \sum_{j=1}^{2}\frac{W_j^2}{\kappa_j}E\left(\frac{I_{X_{ij}\le s} - F_j(s)}{w_j(X_{ij})}\frac{I_{X_{ij}\le t} - F_j(t)}{w_j(X_{ij})}\right).
$$

On the other hand,

$$
\tilde{\sigma}(t,t) = \sigma^B(t,t) + o_p(1)
$$

by $\tilde{F}_j(t) - F_j(t) = O_p(n^{-1/2})$, $\tilde{W}_j - W_j = O_p(n^{-1/2})$ and the Glivenko–Cantelli theorem; the relevant class of functions is shown to be G_j-Glivenko–Cantelli in Appendix D (with $F_0(t)$ replaced by $F_j(t)$ in quantities related the j-th sample). Then by Slutsky's lemma and (E1), we have

$$
\sqrt{n}\left[\tilde{F}_2(t) - \tilde{F}_1(t) - \{F_2(t) - F_1(t)\}\right] \xrightarrow{d} U^B(t)
$$

in $l^\infty([t_1, t_2])$, where $U^B(t)$ is a mean-zero Gaussian process with covariance function $\mathrm{cov}(U^B(s), U^B(t)) = \sigma^B(s,t)$. Then applying the continuous mapping theorem, we obtain the desired result.

Remark. Proofs in this appendix can be applied to show strong consistency results used in Appendix D. We can show $\tilde{W}_j = W_j + o(1)$ a.s. by the SLLN. Then $\tilde{F}_j(t) = F_j(t) + o(1)$ a.s. follows because a G_j-Donsker class is automatically strong G_j-Glivenko–Cantelli (see, e.g., Kosorok 2008, Lemma 8.17).

References

Chang, H., and McKeague, I.W., "Empirical likelihood based tests for stochastic ordering under right censorship." (2014).

Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H., and Nielsen, R. (2005), "Ascertainment bias in studies of human genome-wide polymorphism," *Genome Research*, 15, 1496–1502.

Davidov, O., Fokianos, K., and Iliopoulos, G. (2010), "Order-restricted semiparametric inference for the power bias model," *Biometrics*, 66, 549–557.

Davidov, O., and Herman, A. (2009), "New tests for stochastic order with application to case control studies," *Journal of Statistical Planning and Inference*, 139, 2614–2623.

Duffy, S.W., Nagtegaal, I.D., Wallis, M., Cafferty, F.H., Houssami, N., Warwick, J., Allgood, P.C., Kearins, O., Tappenden, N., O'Sullivan, E., and Lawrence, G. (2008), "Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival," *American Journal of Epidemiology*, 168, 98–104.

Einmahl, J.H.J., and McKeague, I.W. (2003), "Empirical likelihood based hypothesis testing," *Bernoulli*, 9, 267–290.

El Barmi, H., and McKeague, I.W. (2013), "Empirical likelihood based tests for stochastic ordering," *Bernoulli*, 19, 295–307.

El Barmi, H., and Rothmann, M. (1998), "Nonparametric estimation in selection biased models in the presence of estimating equations," *Journal of Nonparametric Statistics*, 9, 381–399.

El Barmi, H., and Mukerjee, H. (2005), "Inferences under a stochastic ordering constraint: the k-sample case," *Journal of the American Statistical Association*, 100, 252–261.

Gilbert, P.B., Lele, S.R., and Vardi, Y. (1999), "Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials," *Biometrika*, 86, 27–43.

Gill, R.D., Vardi, Y., and Wellner, J.A. (1988), "Large sample theory of empirical distributions in biased sampling models," *The Annals of Statistics*, 16, 1069–1112.

Hall, P., *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag (1992).

Hall, P., and Wilson, S.R. (1991), "Two guidelines for bootstrap hypothesis testing," *Biometrics*, 47, 757–762.

Harford, T. (2014), "Big data: a big mistake?," *Significance*, 11, 14–19.

Kitamura, Y., Santos, A., and Shaikh, A.M. (2012), "On the asymptotic optimality of empirical likelihood for testing moment restrictions," *Econometrica*, 80, 413–423.

Kosorok, M.R., *Introduction to Empirical Processes and Semiparametric Inference*, New York: Springer (2008).

Kvam, P.H., Singh, H., and Tiwari, R.C. (1999), "Nonparametric estimation of the survival function based on censored data with additional observations from the residual life distribution," *Statistica Sinica*, 9, 229–246.

McKeague, I.W., and Zhao, Y. (2002), "Simultaneous confidence bands for ratios of survival functions via empirical likelihood," *Statistics & Probability Letters*, 60, 405–415.

Nair, V.N. (1984), "Confidence bands for survival functions with censored data: a comparative study," *Technometrics*, 26, 265–275.

Ning, J., Qin, J., Asgharian, M., and Shen, Y. (2013), "Empirical likelihood-based confidence intervals for length-biased data," *Statistics in Medicine*, 32, 2278–2291.

Nowell, C., and Stanley, L.R. (1991), "Length-biased sampling in mall intercept surveys," *Journal of Marketing Research*, 28, 475–479.

Oluyede, B.O., and George, E.O. (2002), "On stochastic inequalities and comparisons of reliability measures for weighted distributions," *Mathematical Problems in Engineering*, 8, 1–13.

Owen, A.B. (1988), "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, 75, 237–249.

Owen, A.B., *Empirical Likelihood*, Boca Raton: Chapman & Hall/CRC (2001).

Patil, G.P., and Rao, C.R. (1978), "Weighted distributions and size-biased sampling with applications to wildlife populations and human families," *Biometrics*, 34, 179–189.

Qin, J. (1993), "Empirical likelihood in biased sample problems," *The Annals of Statistics*, 21, 1182–1196.

Ramírez, P., and Vidakovic, B. (2010), "Wavelet density estimation for stratified size-biased sample," *Journal of Statistical Planning and Inference*, 140, 419–432.

Rémillard, B., and Scaillet, O. (2009), "Testing for equality between two copulas," *Journal of Multivariate Analysis*, 100, 377–386.

van der Vaart, A.W., *Asymptotic Statistics*, Cambridge Series on Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press (2000).

van der Vaart, A.W., and Wellner, J.A., *Weak Convergence and Empirical Processes*, New York: Springer-Verlag (1996).

Vardi, Y. (1982), "Nonparametric estimation in the presence of length bias," *The Annals of Statistics*, 10, 616–620.

Walker, G.V., Grant, S.R., Guadagnolo, B.A., Hoffman, K.E., Smith, B.D., Koshy, M., Allen, P.K., and Mahmood, U. (2014), "Disparities in stage at diagnosis, treatment, and survival in nonelderly adult patients with cancer according to insurance status," *Journal of Clinical Oncology*, 32, 3118–3125.

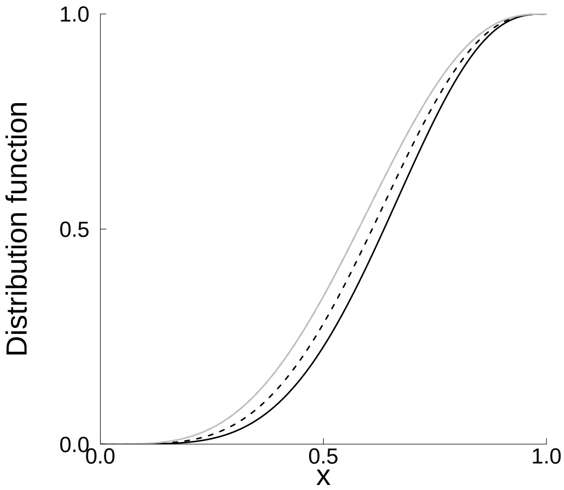
Zelen, M., and Feinleib, M. (1969), "On the theory of screening for chronic diseases," *Biometrika*, 56, 601–614.

Table 1. Empirical significance levels based on 10,000 replications, each with 1000 bootstrap samples. **Scenario A**: distribution functions displayed in Figure 1, upper panel. **Scenario B**: distribution functions displayed in Figure 1, lower panel.

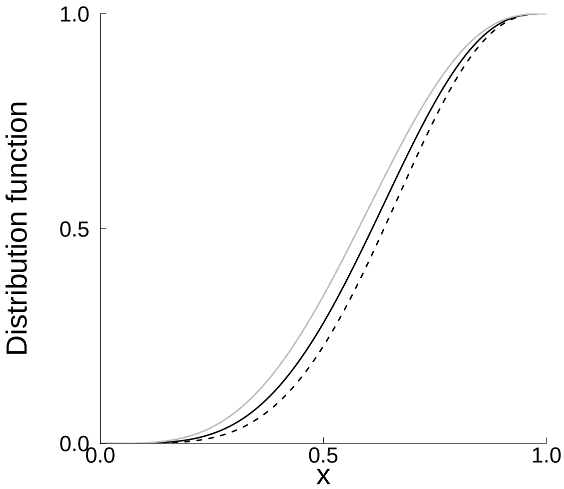
Scenario	group	$\alpha = 0.05$			$\alpha = 0.01$		
	size	M_n	M_n^{ign}	Wald	M_n	M_n^{ign}	Wald
A	50	0.055	0.161	0.056	0.012	0.044	0.011
	80	0.060	0.217	0.063	0.013	0.070	0.012
B	50	0.058	0.014	0.035	0.011	0.001	0.005
	80	0.062	0.012	0.037	0.013	0.001	0.006

Table 2. Power simulation results based on 10,000 replications, each with 1000 bootstrap samples. **Scenario C**: distribution functions displayed in Figure 2, first column. **Scenario D**: distribution functions displayed in Figure 2, second column.

Scenario	group	$\alpha = 0.05$			$\alpha = 0.01$		
	size	M_n	M_n^{ign}	Wald	M_n	M_n^{ign}	Wald
C	50	0.608	0.354	0.529	0.331	0.137	0.237
	80	0.808	0.513	0.756	0.561	0.254	0.467
D	50	0.763	0.416	0.676	0.496	0.177	0.361
	80	0.915	0.590	0.873	0.745	0.322	0.644

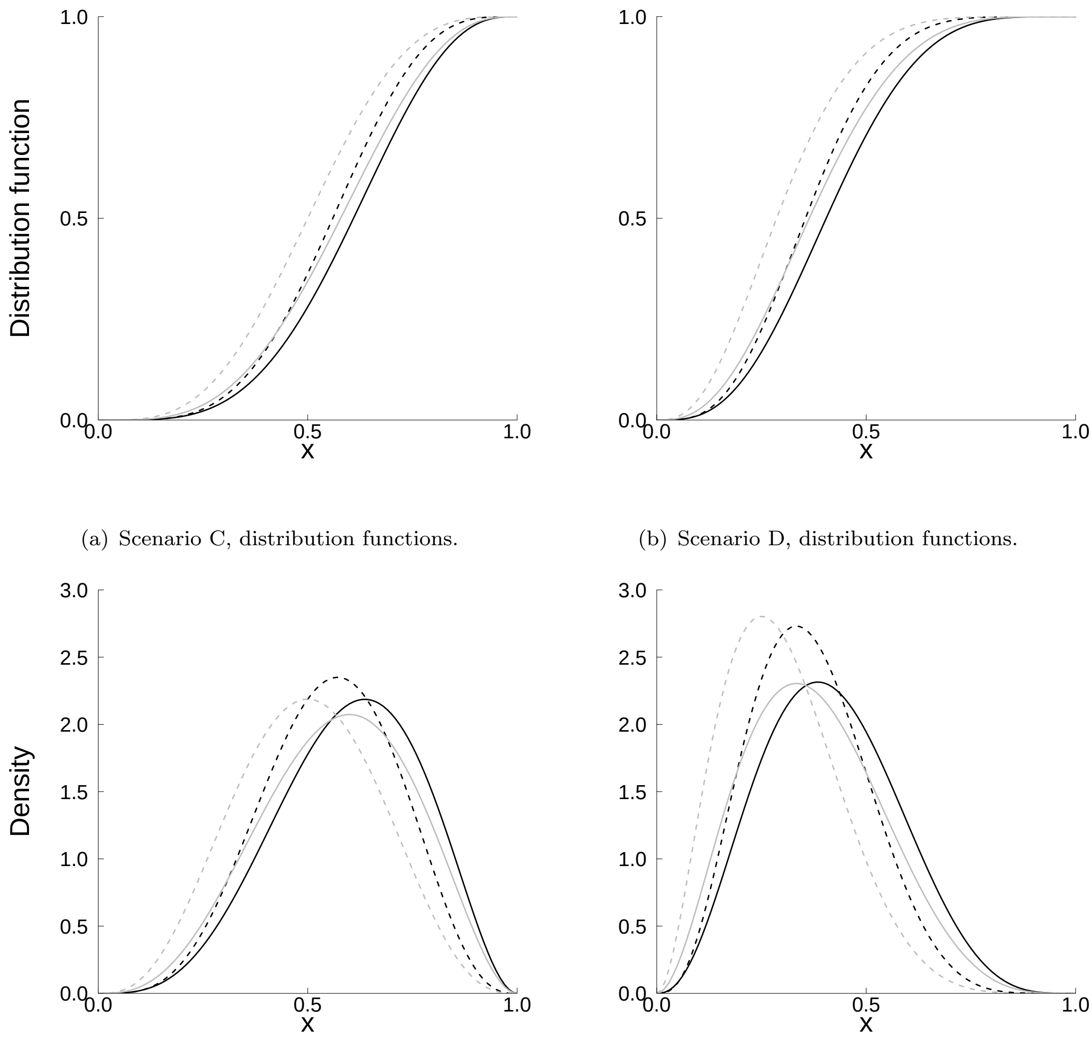


(a) Scenario A.



(b) Scenario B.

Figure 1. For computing empirical levels, the underlying (gray) and weighted (black) distribution functions in Scenario A (top) and Scenario B (bottom): F_1 and G_1 (solid) versus F_2 and G_2 (dashed).



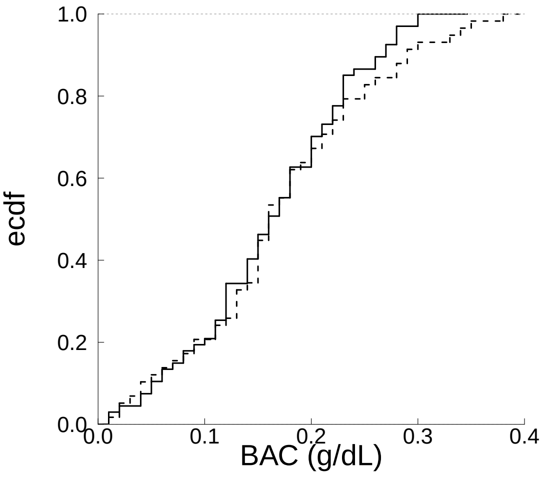
(a) Scenario C, distribution functions.

(b) Scenario D, distribution functions.

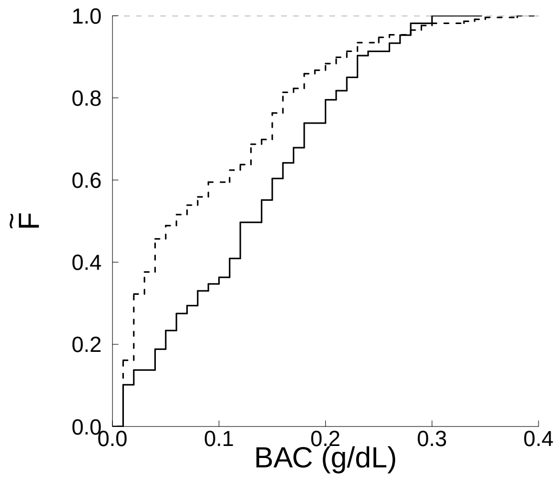
(c) Scenario C, density functions.

(d) Scenario D, density functions.

Figure 2. For power comparisons, the underlying (gray) and weighted (black) distribution (top row) and density (bottom row) functions in Scenario C (first column) and Scenario D (second column): F_1 and G_1 (solid) versus F_2 and G_2 (dashed).



(a) The empirical cdfs.



(b) The NPMLEs for the underlying distribution functions.

Figure 3. The empirical cdf (top) and the NPMLE for the underlying distribution function (bottom) of BAC values for drivers of age less than 30 (solid) and at least 30 (dashed); the weight functions for the NPMLEs are taken to be $w_y(x) = \sqrt{x}$ and $w_o(x) = x$, respectively.

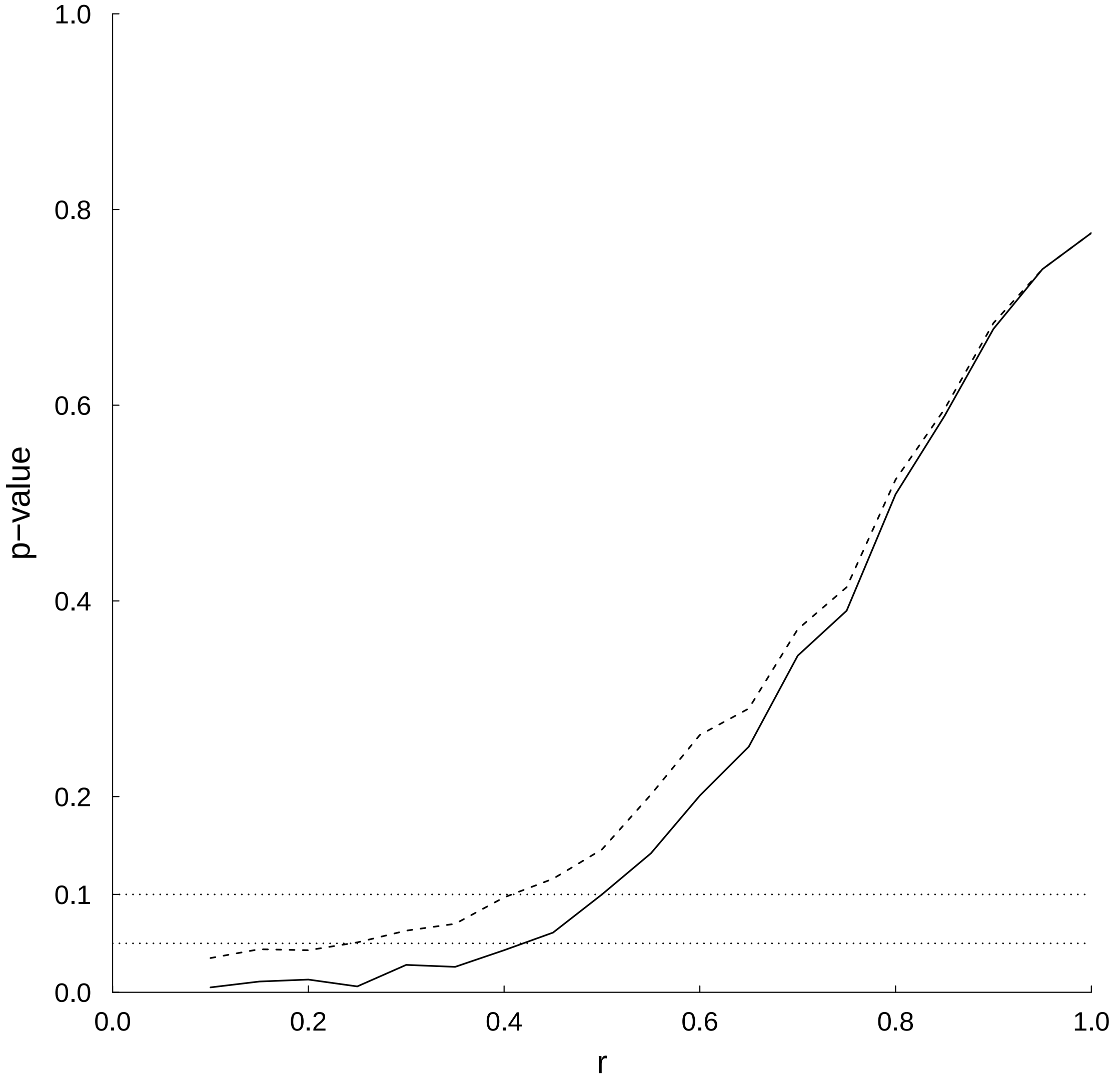


Figure 4. The p-value of M_n (solid) and Wald test (dashed) for comparing BAC in the two age groups when the exponent r in $w_y(x) = x^r$ changes. The horizontal dotted lines indicate the 0.05 and 0.1 significance levels.