Two-sample density-based empirical likelihood ratio tests based on paired data, with application to a treatment study of Attention-Deficit/Hyperactivity Disorder and Severe Mood Dysregulation

Albert Vexler[a*†], Wan-Min Tsai[a], Gregory Gurevich[b] and Jihnhee Yu[a]

Abstract

It is a common practice to conduct medical trials in order to compare a new therapy with a standard-of-care based on paired data consisted of pre- and post-treatment measurements. In such cases, a great interest often lies in identifying treatment effects within each therapy group as well as detecting a between-group difference. In this article, we propose exact nonparametric tests for composite hypotheses related to treatment effects to provide efficient tools that compare study groups utilizing paired data. When correctly specified, parametric likelihood ratios can be applied, in an optimal manner, to detect a difference in distributions of two samples based on paired data. The recent statistical literature introduces density-based empirical likelihood methods to derive efficient nonparametric tests that approximate most powerful Neyman-Pearson decision rules. We adapt and extend these methods to deal with various testing scenarios involved in the two-sample comparisons based on paired data. We show the proposed procedures outperform classical approaches. An extensive Monte Carlo study confirms that the proposed approach is powerful and can be easily applied to a variety of testing problems in practice. The proposed technique is applied for comparing two therapy strategies to treat children's attention deficit/hyperactivity disorder and severe mood dysregulation.

Keywords: Empirical likelihood; Exact tests; Likelihood ratio; Nonparametric test; Paired data; Paired *t*-test; Two-sample problem; Wilcoxon signed rank test

1 Introduction and Technical Preliminaries

Often, investigators in various fields of medical studies deal with paired data to compare different population groups. In this article, we propose a paired data-based methodology motivated by the

[a] *Department of Biostatisics, The State University of New York, Buffalo, NY 14214, U.S.A.*
[b] *The Department of Industrial Engineering and Management, SCE- Shamoon College of Engineering, Beer-Sheva 84100, Israel*
[*] *Correspondence to: Albert Vexler, Department of Biostatisics, The State University of New York, Buffalo, NY 14214, U.S.A.*
[†] *Email: avexler@buffalo.edu*

following comparative study of Attention-Deficit/Hyperactivity Disorder (ADHD) and Severe Mood Dysregulation (SMD). ADHD is a commonly diagnosed psychiatric disorder in children (e.g., Biederman [1]; Nair et al., [2]). SMD is a diagnostic label recently created by the Leibenluft's laboratory in the National Institute of Mental Health's intramural program to refer to children with an abnormal baseline mood, hyperarousal, and increased reactivity to negative emotional stimuli (e.g., Brotman et al., [3]; Carlson [4]; Leibenluft et al., [5]; Waxmonsky et al., [6]). A novel group therapy study at University at Buffalo enrolled 32 children aged 7-12 with ADHD and SMD. These children were treated for 11 weeks. The study participants were randomized between two therapy groups: experimental group therapy program (case; new therapy group) and community psychosocial treatment (control; old therapy group). An objective of the study was to compare the feasibility and efficacy of these two treatments using the Children's Depression Rating Scale – Revised total score (CDRS-Rts). The Children's Depression Rating Scale, revised version (CDRS-R), is a clinician-rated instrument for the diagnosis of childhood depression and the assessment of the severity of depression in children 6-12 years of age (Poznanski et al., [7, 8]). The CDRS-R consists of 17 clinician rated items, with 14 items based on the child's self-report or reports from the parents or teachers and three items based on the child's nonverbal behavior during the interviews. The CDRS-R provides more reliable depression ratings compared to the other children depression rating scales, since it collects information from more sources through interviewing the child, parents or school teachers, independently, as well as it considers the child's behavior during the interview, and lengthens scales to capture slight differences of symptomatology. On the basis of clinical experience, a CDRS-Rts of below 40, 40-60, and above 60 corresponds to none to mild, moderate, and severe depression, respectively (Poznanski et al., [7, 8]; Ying et al., [9]). Thus, the fact that the CDRS-Rts drops significantly over the course of the study implies an effectiveness of a treatment. To record paired data of this study, two measurements were taken from the same subjects. The paired data were constituted by observed values of CDRS-Rts at week 0 (baseline) and week 11 (endpoint).

In this medical study, main research problems are to test differences between distributions of the two therapy groups as well as to detect treatment effects within each group. Testing the hypothesis of no difference between distributions of two therapy groups using the paired data is only one aspect of the comparisons between treatments. For example, in the context of the treatment study of ADHD and SMD,

For Peer Review

previously, Waxmonsky et al. [6] carried out a study to examine the tolerability and efficacy of methylphenidate (MPH) and behavior modification therapy (BMOD), where multiple comparisons with the Bonferroni technique using the independent sample *t*-tests and the pairwise *t*-tests were conducted. The former tests were implemented to evaluate between-group differences in baseline characteristics and measures of tolerability (the Pittsburgh Side Effect Rating Scale, say PSERS). The latter ones were undertaken in the SMD group to compare pre- and post-differences in CDRS-Rts and PSERS from low-dose MPH to high dose MPH. In this article, we avoid considerations of combined *p*-values, proposing a simple and efficient way to create nonparametric tests that attends to special alternative hypotheses directly, in an analogy to parametric likelihood ratio tests. Note that, in controlling the type I error, the used t-tests are known to be inefficient, when utilized data are skewed, and the applied Bonferroni method tends to be conservative. The nonparametric statistical analyses of two populations described above require to consider more versatile testing methods than those well addressed in the classic literature (e.g., [10, 11]). In this article, we propose and examine distribution-free tests for multiple hypotheses to detect various differences related to treatment effects in study groups based on paired data.

To formalize the testing problems, let (X_{ij}, Y_{ij}) be independent identically distributed (i.i.d.) pairs of observations within a subject *j* from sample *i*, where *i*=1, 2 are referred to as treatments; *j*= $1, \ldots, n_i$ are referred to as subjects. In the nonparametric setting, the classic one-sample tests for paired data, e.g. the paired *t*-test and the Wilcoxon signed rank test, are based on differences $Z_{ij} = Y_{ij} - X_{ij}$, where Z_{ij} denotes a within-pair difference of subject *j* from sample *i*, *i*=1, 2; *j*= $1, \ldots, n_i$. Note that $\{Z_{11}, \ldots, Z_{1n_1}\}$ and $\{Z_{21}, \ldots, Z_{2n_2}\}$ consist of i.i.d. observations from populations Z_1 and Z_2 with distribution functions, say, $F_{Z_1}(.)$ and $F_{Z_2}(.)$, respectively. In contexts of treatment evaluations, Z_{ij} can be defined to be the difference of measurements between pre- and post-treatment. In this article, we consider different hypotheses simultaneously for the symmetry of F_{Z_1} and/or F_{Z_2} (detecting a treatment effect into groups) as well as for the equivalence $F_{Z_1} = F_{Z_2}$. Here we refer to the nonparametric literature to connect the term "treatment effect" with tests for symmetry (e.g., Wilcoxon [10]). Note that the Kolmogorov-Smirnov test is a known procedure to compare distributions of populations, whereas the standard testing procedures such as the paired *t*-test, the sign test, and the Wilcoxon signed rank test can be applied to the symmetric

For Peer Review

problem, i.e., to test for $H_0: F_Z(z) = 1 - F_Z(-z)$. Comparisons between distributions of new therapy and control groups as well as detecting treatment effects may be based on multiple hypotheses tests. To this end, one can create relevant tests combining, for example, the Kolmogorov-Smirnov test and the Wilcoxon signed rank test. The use of the classical procedures commonly requires complex considerations to combine the known nonparametric tests. Alternatively, we will develop a direct distribution-free method for analyzing the two-sample problems. The proposed method can be easily applied to test nonparametrically for different composite hypotheses. The proposed approach approximates nonparametrically most powerful Neyman-Pearson test-rules, providing efficiency of the proposed procedures.

When parametric forms of the relevant distributions are known, corresponding parametric likelihood ratios can be easily applied to test for the problems mentioned above. According to the Neyman-Pearson lemma, the parametric likelihood ratio tests are optimal decision rules (e.g., Lehmann and Romano [11]; Vexler and Wu [12]; Vexler et al., [13]). We propose to approximate corresponding likelihood ratios using an empirical likelihood (EL) concept. The EL methodology has been addressed in the statistical literature as one of powerful nonparametric techniques (e.g., Lazar and Mykand [14]; Owen [15-17]; Qin and Lawless [18]; Vexler et al., [19-21]; Yu et al., [22-23]). The EL methodology allows researchers to use distribution-free procedures with efficient characteristics that are asymptotically close to those of related parametric likelihood approaches (e.g., Lazar and Mykand [14]). The EL approach is developed via terms of cumulative distribution functions (e.g., Owen [17]; Vexler et al., [24]; Vexler and Yu [25]). Vexler and Yu [25] demonstrated that the classical EL method based on distribution functions is well suitable for testing parameters; however, the EL technique based on density functions performs more efficiently to test for distributions. To approximate Neyman-Pearson test statistics, Vexler and Gurevich [26] and Gurevich and Vexler [27] proposed to focus on the density-based EL, $L_f = \prod_{i=1}^{n} f_i$, where $f_i = f\left(T_{(i)}\right)$, $f(\cdot)$ is an unknown density function of the observations $\{T_1, \ldots, T_n\}$ and $T_{(i)}$ denotes the i^{th} ordered statistic based on $\{T_1, \ldots, T_n\}$. In this case, approximate values of f_i are obtained by maximizing L_f subject to an empirical version of the constraint $\int f(u)du = 1$.

For Peer Review

We extend and adapt the density-based EL approach for the two-sample testing issues carrying out multiple testing problems in paired data settings. Despite the fact that many statistical inference procedures have been developed for two-sample problems, to our knowledge, relevant nonparametric likelihood techniques to deal with the presented two-sample issues based on paired data have not been well addressed in the literature. The proposed density-based EL tests are exact that ensures accurate computations of relevant *p*-values based on data with small sample sizes.

The paper is organized as follows. In Section 2, we address the purpose of each testing hypothesis considered in this article, and then we develop corresponding density-based EL ratio test statistics. The theoretical results will be presented to show the asymptotic consistency of the proposed tests. To evaluate the proposed approaches, extensive Monte Carlo studies are carried out in Section 3. An application to analyze the CDRS-Rts data is presented in Section 4. In Section 5, we provide some concluding remarks.

2 Statement of Problems and Methods

2.1 Hypotheses setting

To test for equality of the distribution of the new therapy group and the control therapy group based on paired observations $\{Z_{11}, \dots, Z_{1n_1}\}$ and $\{Z_{21}, \dots, Z_{2n_2}\}$, one may consider the hypotheses,

$$H_{N0}: F_{Z_1} = F_{Z_2} = F_Z \text{ vs. } H_A: F_{Z_1} \neq F_{Z_2}.$$

In order to incorporate evaluation of the treatment effect on each therapy group, we point out three tests related to the null hypothesis, 1) the equality of the distributions of two therapy groups, and 2) no treatment effect in each group. This can be presented by H_0: 1) $F_{Z_1} = F_{Z_2} = F_Z$, and 2) $F_Z(z) = 1 - F_Z(-z)$, for all $z \in (-\infty, \infty)$. Against H_0, we can set up three different alternative hypotheses, namely H_{A1}, H_{A2}, and H_{A3}, where

(i) H_{A1}: not H_0, i.e., $F_{Z_1} \neq F_{Z_2}$ or $F_{Z_1}(z_1) \neq 1 - F_{Z_1}(-z_1)$ or $F_{Z_2}(z_2) = 1 - F_{Z_2}(-z_2)$;

(ii) H_{A2}: There is a treatment effect in one therapy group while there is no treatment effect in the other;

(iii) H_{A3}: One asserts that both therapy groups have the same treatment effect. In this case, since the distributions of two groups are assumed to be identical under H_0 and H_{A3}, a one-sample test for symmetry can be applied.

The cases (i)-(iii) are formally noted in Table 1.

For Peer Review

Table1. Hypotheses of interest to be tested based on paired data.

Null hypothesis	*vs.* Alternative hypothesis
$H_{N0}: F_{Z_1} = F_{Z_2} = F_Z$	$H_A: F_{Z_1} \neq F_{Z_2}$
$H_0: F_{Z_1} = F_{Z_2} = F_Z; F_Z(z) = 1 - F_Z(-z)$, for all $z \in (-\infty, \infty)$	$H_{A1}: F_{Z_1} \neq F_{Z_2}$ or $F_{Z_i}(z_i) \neq 1 - F_{Z_i}(-z_i)$, for i=1 or 2 (i.e. not H_0)
	$H_{A2}: F_{Z_1} \neq F_{Z_2}; F_{Z_1}(z_1) \neq 1 - F_{Z_1}(-z_1)$; $F_{Z_2}(z_2) = 1 - F_{Z_2}(-z_2)$
	$H_{A3}: F_{Z_1} = F_{Z_2} = F_{H_{A3},Z}; F_{H_{A3},Z}(z) \neq 1 - F_{H_{A3},Z}(-z)$

Let Test 1, Test 2, and Test 3, refer to the hypothesis tests for the composite hypotheses H_0 *vs.* H_{A1}, H_0 *vs.* H_{A2}, and H_0 *vs.* H_{A3}, respectively.

2.2 Test statistics

In this section, we develop test statistics for Tests 1-3. The proposed three tests will be shown to be exact.

2.2.1 Test 1: H_0 vs. H_{A1}

Consider the scenario where one is interested to test for

$$H_0: F_{Z_1} = F_{Z_2} = F_Z; F_Z(z) = 1 - F_Z(-z), \text{ for all } z \in (-\infty, \infty) \text{ vs. } H_{A1}.$$

The likelihood ratio test statistic based on observations, Z_{ij}, i=1, 2; j= $1, \dots, n_i$, is given by

$$LR_{H_{A1}} = \frac{\prod_{j=1}^{n_1} f_{Z_1}(Z_{1j}) \prod_{j=1}^{n_2} f_{Z_2}(Z_{2j})}{\prod_{j=1}^{n_1} f_Z(Z_{1j}) \prod_{j=1}^{n_2} f_Z(Z_{2j})} = \prod_{j=1}^{n_1} \frac{f_{Z_1,j}}{f_{ZZ_1,j}} \prod_{j=1}^{n_2} \frac{f_{Z_2,j}}{f_{ZZ_2,j}},$$

where f_{Z_i}, i=1, 2, are density functions related to F_{Z_i}, i=1, 2; f_Z is a density function related to a symmetric distribution F_Z; $f_{Z_1,j} = f_{Z_1}(Z_{1(j)})$, $f_{Z_2,j} = f_{Z_2}(Z_{2(j)})$, $f_{ZZ_1,j} = f_Z(Z_{1(j)})$, and $f_{ZZ_2,j} = f_Z(Z_{2(j)})$ as well as $Z_{1(1)} \leq Z_{1(2)} \leq \cdots \leq Z_{1(n_1)}$, $Z_{2(1)} \leq Z_{2(2)} \leq \cdots \leq Z_{2(n_2)}$ are the order statistics based on $\{Z_{11}, \dots, Z_{1n_1}\}$ and $\{Z_{21}, \dots, Z_{2n_2}\}$, respectively. The main novelty of the proposed method for developing the nonparametric test statistic is that we modify the maximum EL concept to obtain directly estimated values of $f_{Z_1,j}, j = 1, \dots, n_1$, maximizing $\prod_{j=1}^{n_1} f_{Z_1,j}$ subject to an empirical constraint. This constraint controls estimated values of $f_{Z_1,j}, j = 1, \dots, n_1$, preserving the main property of the density function f_{Z_1} under the complex structure of the tested hypothesis. To obtain the associated empirical constraint, we utilize the fact that the values of $f_{Z_1,j}$ should be restricted by the equation $\int f_{Z_1}(u)du = 1$. By applying

For Peer Review

the Mean Value Theorem to approximate the constraint $\int f_{Z_1}(u)du = 1$ (for details, see [24-27]), for all positive integer $m \le n/2$, we have

$$(2m)^{-1}\sum_{j=1}^{n_1}\int_{Z_{1(j-m)}}^{Z_{1(j+m)}} f_{Z_1}(u)\,du = (2m)^{-1}\sum_{j=1}^{n_1}\int_{Z_{1(j-m)}}^{Z_{1(j+m)}} f_{Z_1}(u)\frac{f_Z(u)}{f_Z(u)}du$$

$$\cong (2m)^{-1}\sum_{j=1}^{n_1}\frac{f_{Z_1,j}}{f_{ZZ_1,j}}\int_{Z_{1(j-m)}}^{Z_{1(j+m)}} f_Z(u)\,du$$

$$\cong (2m)^{-1}\sum_{j=1}^{n_1}\frac{f_{Z_1,j}}{f_{ZZ_1,j}}\left(F_Z\left(Z_{1(j+m)}\right) - F_Z\left(Z_{1(j-m)}\right)\right). \quad (1)$$

Since under the null hypothesis H_0, the distribution function $F_Z = F_{Z_1} = F_{Z_2}$ is assumed to be symmetric, the idea presented by Schuster [28] can be adapted to estimate $\left(F_Z\left(Z_{1(j+m)}\right) - F_Z\left(Z_{1(j-m)}\right)\right)$ at (1) by using the following estimator, which is denoted as $\eta_{m,j}$,

$$\eta_{m,j} = \left(F_{n_1+n_2}\left(Z_{1(j+m)}\right) - F_{n_1+n_2}\left(Z_{1(j-m)}\right)\right), \quad (2)$$

where $F_{n_1+n_2}(u) = \frac{1}{2(n_1+n_2)}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left[I\left(Z_{ij} \le u\right) + I\left(-Z_{ij} \le u\right)\right]$, $I(.)$ is the indicator function.

By virtue of Lemma 2.1 in [24] and Proposition 2.1 in [26], we have that for all integer $m \le 0.5n_1$,

$$(2m)^{-1}\sum_{j=1}^{n_1}\int_{Z_{1(j-m)}}^{Z_{1(j+m)}} f_{Z_1}(u)\,du$$

$$= \int_{Z_{1(1)}}^{Z_{1(n_1)}} f_{Z_1}(u)\,du - \sum_{r=1}^{m-1}\frac{(m-r)}{2m}\left[\int_{Z_{1(n_1-r)}}^{Z_{1(n_1-r+1)}} f_{Z_1}(u)\,du + \int_{Z_{1(r)}}^{Z_{1(r+1)}} f_{Z_1}(u)\,du\right], \quad (3)$$

where $Z_{1(j-m)} = Z_{1(1)}$, if $j - m \le 1$ and $Z_{1(j+m)} = Z_{1(n_1)}$, if $j + m \ge n_1$.

Since

$$\int_{Z_{1(1)}}^{Z_{1(n_1)}} f_{Z_1}(u)\,du \le \int_{-\infty}^{\infty} f(u)du = 1,$$

For Peer Review

the equation (3) demonstrates that $(2m)^{-1}\sum_{j=1}^{n_1}\int_{Z_{1(j-m)}}^{Z_{1(j+m)}} f_{Z_1}(u)\,du \le 1$, and

$(2m)^{-1}\sum_{j=1}^{n_1}\int_{Z_{1(j-m)}}^{Z_{1(j+m)}} f_{Z_1}(u)\,du \approx 1$ when $m/n_1 \to 0$ as $m, n_1 \to \infty$.

By replacing the distribution functions in (3) by their empirical counterparts, $F_{n_1}(u) = n_1^{-1}\sum_{j=1}^{n_1} I(Z_{1j} \le u)$, the empirical version of the equation (3) then has the form of

$$(2m)^{-1}\sum_{j=1}^{n_1}\int_{Z_{1(j-m)}}^{Z_{1(j+m)}} f_{Z_1}(u)\,du \cong F_{n_1}\left(Z_{1(n_1)}\right) - F_{n_1}\left(Z_{1(1)}\right)$$

$$-\sum_{r=1}^{m-1}\frac{(m-r)}{2m}\left[F_{n_1}\left(Z_{1(n_1-r+1)}\right) - F_{n_1}\left(Z_{1(n_1-r)}\right) + F_{n_1}\left(Z_{1(r+1)}\right) - F_{n_1}\left(Z_{1(r)}\right)\right]. \quad (4)$$

This leads to

$$(2m)^{-1}\sum_{j=1}^{n_1}\int_{Z_{1(j-m)}}^{Z_{1(j+m)}} f_{Z_1}(u)\,du \cong 1 - (m+1)(2n_1)^{-1}. \quad (5)$$

Now, by the equations (1), (2), and (5), the resulting empirical constraint for values of $f_{Z_1,j}$ is

$$(2m)^{-1}\sum_{j=1}^{n_1}\frac{f_{Z_1,j}}{f_{ZZ_1,j}}\eta_{m,j} = 1 - (m+1)(2n_1)^{-1}. \quad (6)$$

To find values of $f_{Z_1,j}$ that maximize the likelihood $\prod_{j=1}^{n_1} f_{Z_1,j}$ provided that the condition (6), we formalize the Lagrange function as

$$\sum_{j=1}^{n_1}\log f_{Z_1,j} + \lambda_1\left[1 - (m+1)(2n_1)^{-1} - (2m)^{-1}\sum_{j=1}^{n_1}\frac{f_{Z_1,j}}{f_{ZZ_1,j}}\eta_{m,j}\right],$$

where λ_1 is a Lagrange multiplier. Maximizing the equation above, the values of $f_{Z_1,1}, \ldots, f_{Z_1,n_1}$ have the form of

$$f_{Z_1,j} = \frac{m(2n_1 - m - 1)}{n_1^2\eta_{m,j}} f_{ZZ_1,j},\ j = 1, \ldots, n_1,$$

For Peer Review

where $Z_{1_{(j-m)}} = Z_{1_{(1)}}$, if $j - m \leq 1$, and $Z_{1_{(j+m)}} = Z_{1_{(n_1)}}$, if $j + m \geq n_1$.

As a consequence, the density-based EL estimator of the ratio $\prod_{j=1}^{n_1} f_{Z_1,j}/f_{ZZ_1,j}$ can be formulated by

$$V_{1,m,1} = \prod_{j=1}^{n_1} \frac{m(2n_1 - m - 1)}{n_1^2 \eta_{m,j}}.$$

One can show that properties of the statistic $V_{1,m,1}$ strongly depend on the selection of values of the integer parameter *m*. A similar problem also arose in the well-known goodness-of-fit tests based on sample entropy, e.g. Vexler et al. [24], Vexler and Gurevich [26], Vasicek [29]. Attending to this issue, we eliminate the dependence on the integer parameter *m*. Towards this end, we utilize the maximum EL concept in a similar manner to arguments proposed in Vexler et al. [24], Vexler and Gurevich [26] and Gurevich and Vexler [27, Appendix A]. Thus, the modified test statistic can be written as

$$V_{1,1} = \min_{{n_1}^{0.5+\delta} \leq m \leq {n_1}^{1-\delta}} \prod_{j=1}^{n_1} \frac{m(2n_1 - m - 1)}{n_1^2 \eta_{m,j}}, \quad \delta \in (0, 1/4), \tag{7}$$

Likewise, the approximation to the likelihood ratio $\prod_{j=1}^{n_2} f_{Z_2,j}/f_{ZZ_2,j}$ is

$$V_{2,2} = \min_{{n_2}^{0.5+\delta} \leq k \leq {n_2}^{1-\delta}} \prod_{j=1}^{n_2} \frac{k(2n_2 - k - 1)}{n_2^2 \varphi_{k,j}}, \quad \delta \in (0, 1/4), \tag{8}$$

where

$$\varphi_{k,j} = \left(F_{n_1+n_2}\left(Z_{2_{(j+k)}}\right) - F_{n_1+n_2}\left(Z_{2_{(j-k)}}\right)\right), \ F_{n_1+n_2} \text{ is defined in (2).}$$

Finally, the proposed test statistic for Test 1 has the form of

$$V_{n_1 n_2}^{H_{A1}} = \prod_{i=1}^{2} V_{i,i} = \min_{{n_1}^{0.5+\delta} \leq m \leq {n_1}^{1-\delta}} \prod_{j=1}^{n_1} \frac{m(2n_1 - m - 1)}{n_1^2 \eta_{m,j}} \min_{{n_2}^{0.5+\delta} \leq k \leq {n_2}^{1-\delta}} \prod_{j=1}^{n_2} \frac{k(2n_2 - k - 1)}{n_2^2 \varphi_{k,j}},$$

that approximates the likelihood ratio $LR_{H_{A1}}$. Consequently, the decision rule is to reject H_0

$$\log\left(V_{n_1 n_2}^{H_{A1}}\right) > C^{H_{A1}}, \tag{9}$$

where $C^{H_{A1}}$ is a test threshold. (Similarly to [30], we will arbitrarily define $\eta_{m,j} = 1/(n_1 + n_2)$ or $\varphi_{k,j} = 1/(n_1 + n_2)$, if $\eta_{m,j} = 0$ or $\varphi_{k,j} = 0$, respectively.) Proposition 1 in Section 2.3 will demonstrate that the proposed test $\log\left(V_{n_1 n_2}^{H_{A1}}\right)$ in (9) is asymptotically consistent. The upper and lower bounds for the integer parameters *m* and *k* in definitions of (7) and (8) were selected to provide the asymptotic consistency. Note that, to test the composite hypotheses H_0 *vs.* H_{A1}, a complex consideration regarding a reasonable combination of the Kolmogorov-Smirnov test and the Wilcoxon signed rank test can be applied (see, for example, Section 3.2). Alternatively, the test (9) uses measurements from the therapy groups, in an approximate Neyman-Person manner, providing a simple procedure to evaluate the treatment effect on each therapy group. Section 3 shows, in various situations, the test (9) is superior to the combinations of the classic tests based on Kolmogorov-Smirnov and Wilcoxon procedures. It is also shown that the proposed nonparametric test has power comparable with that of correct parametric likelihood ratio tests. Thus, in contexts of the study described in Section 1, the direct application of the density-based EL ratio test (9) provides an efficient evaluation of treatment effects with ADHD and SMD in children.

2.2.2 Test 2: H_0 vs. H_{A2}

Our goal is to test for

$$H_0: F_{Z_1} = F_{Z_2} = F_Z, F_Z(z) = 1 - F_Z(-z), \text{for all } z \in (-\infty, \infty),$$

$$\text{vs. } H_{A2}: F_{Z_1} \neq F_{Z_2}, F_{Z_1}(z_1) \neq 1 - F_{Z_1}(-z_1), F_{Z_2}(z_2) = 1 - F_{Z_2}(-z_2).$$

In a similar manner to the development of the density-based EL approximation to the ratio $\prod_{j=1}^{n_1} f_{Z_1,j}/f_{ZZ_1,j}$ mentioned in Section 2.2.1, the EL ratio related to test for H_0 *vs.* H_{A2} can be defined as

$$V_{1,1} = \min_{n_1^{0.5+\delta} \le m \le n_1^{1-\delta}} \prod_{j=1}^{n_1} \frac{m(2n_1 - m - 1)}{n_1^2 \eta_{m,j}}, \delta \in \left(0, \frac{1}{4}\right), \tag{10}$$

where $\eta_{m,j}$ are defined in (2). Consider the density-based EL approximation to the corresponding ratio $\prod_{j=1}^{n_2} f_{Z_2,j}/f_{ZZ_2,j}$. The empirical constraint for values of $f_{Z_2,j}$ can be constructed based on the symmetric property of F_{Z_2}. By analogy with equations (1)-(6), one can show the resulting empirical constraint on values of $f_{Z_2,j}$ in Test 2 has the form of

For Peer Review

$$(2k)^{-1}\sum_{j=1}^{n_2}\frac{f_{Z_2,j}}{f_{ZZ_2,j}}\varphi_{k,j}=\Lambda_{n_2}^{k}, \qquad (11)$$

where $\varphi_{k,j}$ are defined in (8) and

$$\Lambda_{n_2}^{k}=(2n_2)^{-1}\left\{\sum_{j=1}^{n_2}\left[I\left(-Z_{2j}\le Z_{2(n_2)}\right)-I\left(-Z_{2j}\le Z_{2(1)}\right)\right]+n_2-1\right.$$
$$-\sum_{r=1}^{k-1}\frac{(k-r)}{2k}\sum_{j=1}^{n_2}\left[I\left(-Z_{2j}\le Z_{2(n_2-r+1)}\right)-I\left(-Z_{2j}\le Z_{2(n_2-r)}\right)+I\left(-Z_{2j}\le Z_{2(r+1)}\right)\right.$$
$$\left.\left.-I\left(-Z_{2j}\le Z_{2(r)}\right)\right]-\frac{(k-1)}{2}\right\} \qquad (12)$$

The formal derivation of this constraint is given in Appendix A of this article. Then the corresponding Lagrange function can be formulated by

$$\sum_{j=1}^{n_2}\log f_{Z_2,j}+\lambda_2\left[\Lambda_{n_2}^{k}-(2k)^{-1}\sum_{j=1}^{n_2}\frac{f_{Z_2,j}}{f_{ZZ_2,j}}\varphi_{k,j}\right], \qquad (13)$$

where λ_2 is a Lagrange multiplier. Thus approximate values of $f_{Z_2,1},\dots,f_{Z_2,n_2}$ are

$$f_{Z_2,j}=\frac{2k\Lambda_{n_2}^{k}}{n_2\varphi_{k,j}}f_{ZZ_2,j}, j=1,\dots,n_2,$$

where $Z_{2(j-k)}=Z_{2(1)}$, if $j-k\le 1$, and $Z_{2(j+k)}=Z_{2(n_2)}$, if $j+k\ge n_2$.

Similarly to (7) and (8), the density-based EL estimator of the ratio $\prod_{j=1}^{n_2} f_{Z_2,j}/f_{ZZ_2,j}$ can be presented as

$$\tilde{V}_{2,2}=\min_{{n_2}^{0.5+\delta}\le k\le {n_2}^{1-\delta}}\prod_{j=1}^{n_2}\frac{2k\Lambda_{n_2}^{k}}{n_2\varphi_{k,j}}, \delta\in(0,1/4). \qquad (14)$$

Finally, taking into account (10) and (14), the proposed test statistic for Test 2 can be constructed as

$$V_{n_1n_2}^{H_{A2}}=\min_{{n_1}^{0.5+\delta}\le m\le {n_1}^{1-\delta}}\prod_{j=1}^{n_1}\frac{m(2n_1-m-1)}{n_1^2\eta_{m,j}}\min_{{n_2}^{0.5+\delta}\le k\le {n_2}^{1-\delta}}\prod_{j=1}^{n_2}\frac{2k\Lambda_{n_2}^{k}}{n_2\varphi_{k,j}},$$

In this case, the decision rule developed for Test 2 is to reject the null hypothesis if

For Peer Review

$$\log\left(V_{n_1 n_2}^{H_{A2}}\right) > C^{H_{A2}}, \tag{15}$$

where $C^{H_{A2}}$ is a test threshold.

2.2.3 Test 3: H_0 vs. H_{A3}

Consider the following hypotheses of interest

$$H_0: F_{Z_1} = F_{Z_2} = F_Z, F_Z(z) = 1 - F_Z(-z), \text{for all } z \in (-\infty, \infty),$$

$$\text{vs. } H_{A3}: F_{Z_1} = F_{Z_2} = F_{H_{A3},Z}, F_{H_{A3},Z}(z) \neq 1 - F_{H_{A3},Z}(-z).$$

The corresponding likelihood ratio test statistic based on observations $Z_{ij}, i = 1,2;\ j = 1,,,n_i$ can be

$$LR_{H_{A3}} = \frac{\prod_{i=1}^{2}\prod_{j=1}^{n_i} f_{H_{A3},Z}(Z_{ij})}{\prod_{i=1}^{2}\prod_{j=1}^{n_i} f_Z(Z_{ij})} = \frac{\prod_{s=1}^{N} f_{H_{A3},Z}(Z_{(s)})}{\prod_{s=1}^{N} f_Z(Z_{(s)})} = \prod_{s=1}^{N}\frac{f_{H_{A3},s}}{f_{H_0,s}},$$

where $N = n_1 + n_2$; $f_{H_0,s} = f_Z(Z_{(s)})$ and $f_{H_{A3},s} = f_{H_{A3},Z}(Z_{(s)})$ denote the density functions of observations Z under H_0 and H_{A3}, respectively, as well as $Z_{(s)}$, $s = 1, \ldots, N$, are the order statistics based on the pooled sample of $\{Z_{11}, \ldots, Z_{1n_1}\}$ and $\{Z_{21}, \ldots, Z_{2n_2}\}$ that are denoted by $Z_{s,}$ $s = 1, \ldots, N$. Using the same technique as in Section 2.2.1, we derive values of $f_{H_{A3},s}$, $s = 1, \ldots, N$, that maximize the log likelihood $\sum_{s=1}^{N}\log(f_{H_{A3},s})$ given a constraint, empirical form of $\int f_{H_{A3}}(u)du = 1$. The proposed test statistic for Test 3 is

$$V_N^{H_{A3}} = \min_{N^{0.5+\delta}\le m\le N^{1-\delta}} \prod_{s=1}^{N}\frac{m(2N-m-1)}{N^2\omega_{m,s}}, \tag{16}$$

where $\omega_{m,s} = (2N)^{-1}\sum_{s=1}^{N}\left[I\left(Z_s \le Z_{(s+m)}\right) + I\left(-Z_s \le Z_{(s+m)}\right) - I\left(Z_s \le Z_{(s-m)}\right) - I\left(-Z_s \le Z_{(s-m)}\right)\right]$ and $\delta \in (0, 1/4)$.

Thus, the decision rule for Test 3 is to reject the null hypothesis if

$$\log\left(V_N^{H_{A3}}\right) > C^{H_{A3}}, \tag{17}$$

where $C^{H_{A3}}$ is a test threshold.

For Peer Review

2.3 Asymptotic consistency of the tests

In this section, we present the following propositions to demonstrate the asymptotic consistency of the proposed tests:

Proposition 1. Let $f_{Z_i}(z)$ be the density function with the expectations $E(\log f_{Z_i}(Z_{i1})) < \infty$ and $E(\log f_{Z_i}(-Z_{i1})) < \infty, i = 1,2$. Let $f_i(u) = \left(f_{Z_i}(u) + f_{Z_i}(-u)\right)/2$. Then under H_0, $(n_1+n_2)^{-1}\log(V_{n_1n_2}^{H_{At}}) \xrightarrow{p} 0$, $t = 1,2$, and under H_{A1} and H_{A2},

$$(n_1+n_2)^{-1}\log\left(V_{n_1n_2}^{H_{At}}\right) \xrightarrow{p} -\frac{\gamma}{1+\gamma}E_{H_{At}}\log\left\{\frac{\gamma}{1+\gamma}+\frac{1}{1+\gamma}\left(\frac{f_2(Z_{11})}{f_1(Z_{11})}\right)\right\}$$
$$-\frac{1}{1+\gamma}E_{H_{At}}\log\left\{\frac{1}{1+\gamma}+\frac{\gamma}{1+\gamma}\left(\frac{f_1(Z_{21})}{f_2(Z_{21})}\right)\right\} \geq 0, t = 1,2,$$

as $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \gamma > 0$, where γ is a constant.

Proof. We outline the proof in Appendix A1 provided in Supplementary material (S1)[§].

Consider the testing problem H_0 *vs*. H_{A3}.

Proposition 2. Let the pooled sample Z_s, $s = 1, \ldots, N$, have the density function, $f(z)$, with the expectations $E(\log f(Z_1)) < \infty$ and $E(\log f(-Z_1)) < \infty$. Then under H_0, $N^{-1}\log(V_N^{H_{A3}}) \xrightarrow{p} 0$, and under H_{A3}, $N^{-1}\log\left(V_N^{H_{A3}}\right) \xrightarrow{p} -E_{H_{A3}}\log\left\{0.5+0.5\left(\frac{f(-Z_1)}{f(Z_1)}\right)\right\} \geq 0$, as $N \to \infty$.

Proof. We omit the proof, since it is similar to the proof of Proposition 1.

2.4 Null distributions of the proposed test statistics

To obtain critical values of the proposed tests, we utilize the fact that the proposed test statistics are based on indicator functions $I(\cdot)$ and $I(Z_1 < Z_2) = I(F_Z(Z_1) < F_Z(Z_2))$ as well as $I(Z_1 < -Z_2) = I(F_Z(Z_1) < F_Z(-Z_2)) = I\left(F_Z(Z_1) < 1 - F_Z(Z_2)\right)$, where the random variables $F_Z(Z_1)$ and $F_Z(Z_2)$ have the uniform distribution, *Unif*[0,1], under H_0. Thus the distributions of the proposed test statistics are independent of the distributions of observations and hence, the critical values of the proposed tests can be exactly computed. For each proposed test, we conducted the following procedures to determine the critical values, C_α, of the null distributions. We first generated data of Z_1 and Z_2 from the standard normal distribution

[§] Supporting information may be found in the online version of this article.

For Peer Review

$N(0, 1)$ and then calculated the test statistics corresponding to each proposed test. At each sample sizes n_1 and n_2, we obtained 50,000 generated values of the test statistics (9), (15), and (17), with $\delta = 0.1$ tabulating the critical values for the null distributions of the test statistics at the significance levels α, α =0.01; 0.05; 0.1 (see, Table 2).

------Table 2------

Remark 1. The definitions (9), (15) and (17) of the proposed test statistic include $\delta \in (0, 1/4)$. We set up $\delta = 0.1$. To investigate the test statistics with different values of δ, we conducted an extensive Monte Carlo study. The Monte Carlo powers of the proposed tests were not found to be significantly dependent on values of $\delta \in (0, 1/4)$. These experimental results are similar to those shown in [24, 25] and [27].

Remark 2. The computer codes related to the outputs of this section are provided in Supplemental Section S2. These codes can be easily modified to obtain results of the next section or to perform the proposed tests based on real data.

3 Simulation Study

In this section, we examine the power properties of the proposed tests in various cases using Monte Carlo simulations. The proposed tests based on (9), (15), and (17), with δ=0.1, are compared with the common test procedures: the maximum likelihood ratio (MLR) tests, assuming parametric conditions on distributions of observations (for details of the constructions and definitions of the MLR tests, see Appendix A2 of the supplementary material, S1); combined classic nonparametric tests with a structure based on the Wilcoxon signed rank test or/and the Kolmogorov-Smirnov test. We fixed the significance level of the tests to be 0.05 in all considered cases.

3.1 Power comparison with the parametric method (the MLR tests)

In order to present the comparative power of the proposed tests versus the corresponding MLR tests, we performed the following Monte Carlo study. Critical values of the MLR test statistics were obtained based on 50,000 simulations under H_0 based on $N(0,1)$-distributed observations Z. To study the powers of the tests, 10,000 samples for each size (n_1, n_2) were generated from a variety of distributions. Tables 3-5 depict the Monte Carlo powers of the proposed tests and those of the corresponding MLR tests.

------Table 3------

For Peer Review

------Table 4------

------Table 5------

When observations are normally distributed, as anticipated, the MLR tests would be more powerful than the proposed nonparametric tests. The tables show the powers of the proposed tests are very close to those of the MLR tests, demonstrating that the density-based EL tests are comparable to the parametric method that utilizes the correct information regarding distributions of observations. Table 6 displays the actual type I errors of the MLR tests under the misspecification of underlying distributions, i.e., when observations were simulated from *t* distributions with different degrees of freedom, a logistic with parameters (0,1), a Laplace with parameters (0,1), and the *Unif*[0,1] under H_0. As can be seen from Table 6, the type I errors of the MLR tests for H_0 *vs*. H_{A1} and H_0 *vs*. H_{A2} are not under control until the degrees of freedom of the *t* distribution < 200. For the cases of the logistic and the Laplace distribution, the type I errors of the MLR tests are not well controlled. When the observations are from *Unif*[0,1], the impact of the misspecification of the model on the type I errors of the MLR tests is more significant. This illustrates the considered MLR tests are strongly dependent on assumptions regarding distributions of observations.

------Table 6------

3.2 Power comparison with classic nonparametric methods

In this section, we compare the power of the proposed tests to the power of procedures based on the classic nonparametric tests. Since Tests 1 and 2 are based on the composite hypotheses regarding between-group differences and treatment effects, the respective Kolmogorov-Smirnov test and Wilcoxon signed rank test cannot be directly applied to test for these hypotheses. In this case, one can perform combined tests based on the Kolmogorov-Smirnov test and the Wilcoxon signed rank test for H_0 *vs*. H_{A1} and H_0 *vs*. H_{A2}. For the comparison, we used combined nonparametric tests with the Bonferroni method. To this end, the R procedure "*p.adjust*" with the method "*bonferroni*" was utilized. Let "W-test" denote the Wilcoxon signed rank test and "K-S test" denote the Kolmogorov-Smirnov test. The combined nonparametric test for H_0 *vs*. H_{A1} consists of two W-tests for symmetry and one K-S test based on Z_1, Z_2 for $F_{Z_1} = F_{Z_2}$. The former tests are employed to assess a treatment effect of each therapy group, whereas

For Peer Review

the latter test is conducted to detect the group difference. Similarly, we performed the combined nonparametric test for H_0 $vs.$ H_{A2} that includes one W-test and one K-S test. The classical procedure for H_0 $vs.$ H_{A3} is the W-test for symmetry.

To test H_0 $vs.$ H_{A1} and H_0 $vs.$ H_{A3}, we assigned different distributions for baseline measurements X and endpoint observations Y in each group under the alternative hypothesis (i.e., (F_{X_1}, F_{Y_1}) $vs.$ (F_{X_2}, F_{Y_2})), whereas when testing for H_0 $vs.$ H_{A2}, we directly generated observations Z_2 from three cases of symmetric distributions under H_{A2}: $N(0, 1)$; $Unif[-1,1]$; t_2 distribution. Tables 7-9 contain the results of the power comparisons of the two different testing procedures: the proposed procedures and the nonparametric testing procedures based on the W-test and/or K-S test using the Bonferroni approach.

------Table 7------

------Table 8------

------Table 9------

The Monte Carlo outputs shown in Tables 7-9 indicate that the new tests have higher powers against the combined nonparametric tests. In particular, for the cases of small sample sizes (e.g., $(n_1, n_2) = (10, 10)$, (25, 25)), the proposed tests are significantly superior to the classic tests. In several cases, the powers of the proposed tests have values that are 3-4 times larger than those of the combined nonparametric tests.

4 Data Analysis

In this section, we apply the proposed method to the study described in Section 1, which evaluates treatment effects of ADHD and SMD in children. Study subjects were randomized to receive either the experimental 11 week group therapy program ($n_1 = 17$) or community psychosocial treatment ($n_2 = 15$). We defined the former as group 1 and the latter as group 2. For each child enrolled in the study, CDRS-Rts was taken at the baseline (week 0) and endpoint (week 11). Specifically, we computed the differences of CDRS-Rts: $Z_{ij} = Y_{ij} - X_{ij}$, for i=1, 2; j= $1, \dots, n_i$, where X_{ij} stands for the CDRS-Rts assessed at baseline before subject j receives treatment i and Y_{ij} represents another CDRS-Rts at the endpoint after subject j receives treatment i.

For Peer Review

The empirical histograms of the CDRS-Rts at baseline and endpoint for each group are shown in Figure 1. As can be seen from Figure 1, it appears that both therapy groups have a decline in the CDRS-Rts after baseline but the decrease in the CDRS-Rts seems to be more significant in the group 1.

------Figure 1------

In the context of the study's interests to test a claim that the distributions of the changes in CDRS-Rts are not equivalent with respect to the therapy groups or at least one therapy group has a treatment effect, we performed the proposed test 1. In this case, the observed value of the test statistic by (9), with δ=0.1, is 22.8217 and the corresponding p-value is 0.00002, indicating the null hypothesis of "no group differences and the lack of treatment effects in both groups" is rejected. The combined nonparametric test (the two "W-tests" and one "K-S test") also rejects the null hypothesis with the p-value 0.000005. Based on these results, there is strong evidence to reject the null hypothesis.

In addition, to demonstrate applicability of the proposed tests, we carried out Test 2 that might be appropriate to test an assertion that there is a treatment effect in one group and no such effect in the other besides a group difference. The observed value of the test statistic by (15), with δ=0.1, is 11.9370 and the corresponding p-value is 4×10^{-5}. The combined nonparametric test (the "W-test" and one "K-S test") with the Bonferroni method also supports the result to reject the null hypothesis with the p-value of 5×10^{-7}. These results show that the proposed procedures are in conjunction with the classic tests, demonstrating that our proposed tests can be utilized in the ADHD and SMD study.

In addition to the analysis above, we also conducted a bootstrap type study to evaluate the efficiency (power) of the considered tests based on small datasets randomly selected from the original data. To perform this study for Test 1, we executed the following procedure. We randomly selected samples with the sample sizes of (n_1, n_2) = (9, 6), (9, 7), (11, 9), (13, 10), (13, 11), (15, 13) from the original dataset. Then we calculated the corresponding test statistic $\log\left(V_{n_1 n_2}^{H_{A1}}\right)$ by (9), where δ=0.1. We repeated this strategy 10,000 times calculating the proportion of rejections at $\alpha = 0.05$ of the null hypothesis; that is, we computed the percentage of times when $\log\left(V_{n_1 n_2}^{H_{A1}}\right) > C_{\alpha=0.05}$. The bootstrap type study for Test 2 was also carried out following the same procedures as described above. The results regarding the proportion of the rejections of the null hypothesis for each considered test are provided in Table 10.

For Peer Review

------Table 10------

Table 10 demonstrates that the proposed procedures have larger proportion of the rejections in comparison with the combined nonparametric tests. In particular, when the sample sizes are relatively small (e.g., $(n_1, n_2) = (9,\ 6),\ (9,\ 7))$, the differences in the proportions of the rejections between two approaches are strongly recognizable. For example, we selected a sample of size 9 from the group 1 and a sample of size 6 from the group 2. This sub-dataset was tested for the hypotheses H_0 *vs*. H_{A1} (Test 1). In contrast to the result that the nonparametric test based on the two "W-tests" and one "K-S test" for H_0 *vs*. H_{A1} is not statistically significant (the Bonferroni adjusted *p*-values of these classic tests are 0.0617, 0.1050, and 0.9873, respectively), the proposed Test 1, with δ=0.1, is statistically significant (*p*-value=0.0005). Figure 2 shows the empirical histograms of Z_1 and Z_2 from the sub-dataset. All these results indicate that the proposed methods for Tests 1 and 2 are more sensitive to detect the difference between the null hypothesis and the alternative hypotheses involved in Tests 1 and 2 compared to the corresponding combined nonparametric tests.

------Figure 2------

5 Concluding Remarks

In this article, we proposed and examined the two-sample density-based EL ratio tests based on paired observations. While constructing the tests, we used approximations to the most powerful test statistics with respect to the stated problems, providing efficient nonparametric procedures. The proposed tests are shown to be exact and simple in performing. The extensive Monte-Carlo studies confirmed powerful properties of the proposed tests. We showed our tests outperform different tests with a structure based on the Wilcoxon signed rank test and/or the Kolmogorov-Smirnov test, and outperform the parametric likelihood ratio tests when the underlying distributions are misspecified. The data example illustrated that the proposed tests can be easily and efficiently used in practice.

For Peer Review

APPENDIX A. Computing the empirical constraint (11) in the development of Test 2

Similarly to the equations (1)-(4), by virtue of the Mean Value Theorem and Lemma 2.1 in [24], we have

$$(2k)^{-1}\sum_{j=1}^{n_2}\int_{Z_{2(j-k)}}^{Z_{2(j+k)}} f_{Z_2}(u)\,du \cong (2k)^{-1}\sum_{j=1}^{n_2}\frac{f_{Z_2,j}}{f_{ZZ_2,j}}\varphi_{k,j}$$

and

$$(2k)^{-1}\sum_{j=1}^{n_2}\int_{Z_{2(j-k)}}^{Z_{2(j+k)}} f_{Z_2}(u)\,du \;\cong F_{Z_2}\left(Z_{2(n_2)}\right) - F_{Z_2}\left(Z_{2(1)}\right)$$

$$-\sum_{r=1}^{k-1}\frac{(k-r)}{2k}\left[F_{Z_2}\left(Z_{2(n_2-r+1)}\right) - F_{Z_2}\left(Z_{2(n_2-r)}\right) + F_{Z_2}\left(Z_{2(r+1)}\right) - F_{Z_2}\left(Z_{2(r)}\right)\right], \qquad (A1)$$

where $\varphi_{k,j}$ are defined in (8). Here, due to the symmetric property of F_{Z_2} , under the alternative hypothesis H_{A2}, applying the estimation proposed by Schuster [28], $F_{Z_2}(r) - F_{Z_2}(s)$ can be estimated by

$$\tilde{F}_{Z_2}(r) - \tilde{F}_{Z_2}(s) = (2n_2)^{-1}\sum_{j=1}^{n_2}\left[I\left(Z_{2j} \le r\right) + I\left(-Z_{2j} \le r\right) - I\left(Z_{2j} \le s\right) - I\left(-Z_{2j} \le s\right)\right].$$

Now, the right-hand side of equation (A1) can be estimated by $\Lambda_{n_2}^k$ defined in equation (12). This concludes that the resulting empirical constraint on values of $f_{Z_2,j}$ in Test 2 has the form of

$$(2k)^{-1}\sum_{j=1}^{n_2}\frac{f_{Z_2,j}}{f_{ZZ_2,j}}\varphi_{k,j} = \Lambda_{n_2}^k.$$

Acknowledgements

This research is supported by the NIH grant 1R03DE020851 - 01A1 (the National Institute of Dental and Craniofacial Research). The authors are grateful to the Editor, the Associate Editor and the referees for suggestions that led to a substantial improvement in this paper.

References

1. Biederman J. Attention-deficit/hyperactivity disorder: a life-span perspective. *The Journal of Clinical Psychiatry* 1998; **59**:4-16.

2. Nair J, Ehimare U, Beitman BD, Nair SS, Lavin A. Clinical review: evidence-based diagnosis and treatment of ADHD in children. *Missouri Medicine* 2006; **103**:617-621.

3. Brotman MA, Schmajuk M, Rich B, Dickstein DP, Guyer AE, Costello EJ, Egger HL, Angold A, Leibenluft E. Prevalence, clinical correlates and longitudinal course of severe mood dysregulation in children. *Biological Psychiatry* 2006; **60**:991-997.

For Peer Review

4. Carlson GA. Who Are the Children with Severe Mood Dysregulation, a.k.a. "Rages"? *American Journal of Psychiartry* 2007; **164**:1140-1142.

5. Leibenluft E, Charney DS, Towbin KE, Bhangoo RK, Pine DS. Defining Clinical Phenotypes of Juvenile Mania. *American Journal of Psychiartry* 2003; **160**:430-437.

6. Waxmonsky J, Pelham WE, Gnagy E, Cummings MR, O'Connor B, Majumdar A, Verley J, Hoffman MT, Massetti GA, Burrows-MacLean L, Fabiano GA, Waschbusch DA, Chacko A, Arnold FW, Walker KS, Garefino AC, Robb JA. The efficacy and tolerability of methylphenidate and behavior modification in children with attention-deficit/hyperactivity disorder and severe mood dysregulation. *Journal of Child and Adolescent Psychopharmacology* 2008; **18**:573-88.

7. Poznanski EO, Cook SC, Carroll BJ. A depression rating scale for children. *Pediatrics* 1979; **64**:442-450.

8. Poznanski EO, Grossman JA, Buchsbaum Y, Banegas M, Freeman L, Gibbons R. Preliminary studies of the reliability and validity of the children's depression rating scale. *Journal of the American Academy of Child Psychiatry* 1984; **23**:191-197

9. Ying Guo, Mary EN, John H, Michael GW, Graham E. An Exploratory Factor Analysis of the Children's Depression Rating Scale-Revised. *Journal of Child and Adolescent Psychopharmacology* 2006; **16**:482-491.

10. Wilcoxon F. Individual comparisons by ranking methods. *Biometrics* 1945; **1**:80-83.

11. Lehmann EL, Romano JP. *Testing Statistical Hypotheses*. Springer: New York, 2005.

12. Vexler A, Wu C. An Optimal Retrospective Change Point Detection Policy. *Scandinavian Journal of Statistics* 2009; **36**:542-558.

13. Vexler A, Wu C, Yu KF. Optimal hypothesis testing: from semi to fully Bayes factors. *Metrika* 2010; **71**:125-138.

14. Lazar N, Mykland PA. An evaluation of the power and conditionality properties of empirical likelihood. *Biometrika* 1998; **85**:523-534.

15. Owen AB. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 1988; **75**:237–249.

16. Owen AB. Empirical Likelihood for Linear Models. *The Annals of Statistics* 1991; **19**:1725–1747.

17. Owen AB. *Empirical Likelihood*. Chapman and Hall/CRC: New York, 2001.

18. Qin J, Lawless J. Empirical Likelihood and General Estimating Equations. *The Annals of Statistics* 1994; **22**:300-325.

For Peer Review

19. Vexler A, Liu S, Kang L, Hutson AD. Modifications of the Empirical Likelihood Interval Estimation with Improved Coverage Probabilities. *Communications in Statistics (Simulation and Computation)* 2009; **38**:2171-2183.

20. Vexler A, Yu J, Tian L, Liu S. Two-sample nonparametric likelihood inference based on incomplete data with an application to a pneumonia study. *Biometrical Journal* 2010; **52**:348–361.

21. Vexler A, Tsai W-M, Yaakov M. Estimation and testing based on data subject to measurement errors: from parametric to non-parametric likelihood methods. *Statistics in Medicine* 2011. DOI: 10.1002/sim.4304

22. Yu J, Vexler A, Tian L. Analyzing Incomplete Data Subject to a Threshold Using Empirical Likelihood Methods: An Application to a Pneumonia Risk Study in an ICU Setting. *Biometrics* 2010; **66**:123-130.

23. Yu J, Vexler A, Kim S, Hutson AD. Two-sample Empirical likelihood ratio tests for medians application to biomarker evaluations. *The Canadian Journal of Statistics* 2011. In Press. DOI: 10.1002/cjs.10108

24. Vexler A, Shan G, Kim S, Tsai W-M, Tian L, Hutson AD. An empirical likelihood ratio based goodness-of-fit test for Inverse Gaussian distributions. *Journal of Statistical Planning and Inference* 2011; **141**:2128-2140.

25. Vexler A, Yu J. Two-sample density-based empirical likelihood tests for incomplete data in application to a pneumonia study. *Biometrical Journal* 2011; **53**: 628-651.

26. Vexler A, Gurevich G. Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy. *Computational Statistics & Data Analysis* 2010; **54**:531-545.

27. Gurevich G, Vexler A. A two-sample empirical likelihood ratio test based on samples entropy. *Statisitics and Computing* 2011; **21**: 657-670.

28. Schuster EF. Estimating the distribution function of a symmetric distribution. *Biometrika* 1975; **62**: 631-635.

29. Vasicek O. A test for normality based on sample entropy. *Journal of the Royal Statistical Society. Series B (Methodological)* 1976; **38**:54-59.

30. Canner PL. A simulation study of one-and two-sample Kolmogorov-Smirnov statistics with a particular weight function. *Journal of American Statistical Association* 1975; **70,** 209-211.

31. Serfling RJ. *Approximation Theorems of Mathematical Statistics*. Wiley: New York, 1980.

For Peer Review

Table 2. The critical values for Test 1 by (9) (Test 2 by (15)) [Test 3 by (17)] with δ=0.1 for different sample sizes (n_1, n_2) and significance levels α

		n_2					
n_1	α	10	15	20	25	30	35
10	0.01	7.44 (5.75) [4.17]	7.18 (5.85) [4.18]	7.39 (6.02) [4.18]	7.33 (6.17) [4.16]	7.45 (6.33) [4.25]	7.47 (6.20) [4.42]
	0.05	5.22 (3.86) [2.68]	5.06 (3.93) [2.82]	5.25 (4.09) [2.84]	5.35 (4.28) [2.85]	5.38 (4.37) [2.97]	5.42 (4.43) [3.11]
	0.1	4.27 (3.11) [2.14]	4.19 (3.17) [2.27]	4.39 (3.35) [2.30]	4.48 (3.52) [2.36]	4.56 (3.60) [2.46]	4.59 (3.68) [2.60]
15	0.01		6.83 (5.39) [4.15]	6.96 (5.57) [4.16]	6.98 (5.68) [4.24]	6.88 (5.80) [4.33]	6.96 (5.81) [4.39]
	0.05		4.90 (3.73) [2.83]	5.02 (3.89) [2.85]	5.15 (4.04) [2.97]	5.14 (4.11) [3.12]	5.20 (4.18) [3.13]
	0.1		4.09 (3.07) [2.30]	4.25 (3.24) [2.34]	4.38 (3.38) [2.45]	4.36 (3.45) [2.59]	4.44 (3.52) [2.63]
20	0.01			6.86 (5.44) [4.24]	6.96 (5.55) [4.38]	6.99 (5.69) [4.41]	7.01 (5.73) [4.42]
	0.05			5.11 (3.94) [2.96]	5.24 (4.11) [3.11]	5.25 (4.19) [3.16]	5.27 (4.26) [3.19]
	0.1			4.35 (3.33) [2.45]	4.48 (3.47) [2.58]	4.51 (3.55) [2.64]	4.52 (3.62) [2.69]
25	0.01				7.08 (5.62) [4.40]	7.01 (5.76) [4.44]	7.10 (5.85) [4.56]
	0.05				5.33 (4.18) [3.15]	5.36 (4.25) [3.22]	5.37 (4.32) [3.34]
	0.1				4.56 (3.55) [2.64]	4.59 (3.63) [2.70]	4.62 (3.69) [2.82]
30	0.01					6.89 (5.63) [4.55]	6.99 (5.75) [4.68]
	0.05					5.31 (4.23) [3.32]	5.33 (4.27) [3.44]
	0.1					4.59 (3.65) [2.80]	4.61 (3.68) [2.93]
35	0.01						6.91 (5.68) [4.64]
	0.05						5.35 (4.28) [3.47]
	0.1						4.64 (3.69) [2.98]

Table 3. The Monte Carlo powers of Test 1 by (9) vs. the MLR test for H_0 *vs.* H_{A1} with different sample sizes (n_1, n_2) at the significance level $\alpha = 0.05$.

F_{X_1}	F_{Y_1}	F_{X_2}	F_{Y_2}	n_1	n_2	Proposed test at (9)	MLR
$N(0, 1)$	$N(0.2, 0.25^2)$	$N(0.1, 0.5^2)$	$N(0.5, 1)$				
				10	10	0.1541	0.1579
				50	50	0.6232	0.6921
$N(2.5, 0.8^2)$	$N(1.5, 0.5^2)$	$N(1, 1.5^2)$	$N(1.5, 0.6^2)$				
				10	10	0.7267	0.7351
				25	25	0.9956	0.9978
$N(0.3, 0.5^2)$	$N(0.5, 1)$	$N(0.25, 0.25^2)$	$N(0.5, 0.5^2)$				
				10	10	0.2818	0.4497
				50	50	0.9764	0.9993
$N(0.5, 0.5^2)$	$N(1, 1)$	$N(0, 1)$	$N(0, 1)$				
				10	10	0.1723	0.1764
				50	50	0.7171	0.7942
$N(0, 1)$	$N(0, 1)$	$N(1.5, 1.1^2)$	$N(1, 1.3^2)$				
				10	10	0.1125	0.1300
				50	50	0.4321	0.5531
$N(0, 1)$	$N(0.5, 1)$	$N(0.5, 1.2^2)$	$N(1, 0.5^2)$				
				10	10	0.2208	0.2230
				50	50	0.8348	0.8785

Table 4. The Monte Carlo powers of Test 2 by (15) vs. the MLR test for H_0 *vs.* H_{A2} with different sample sizes (n_1, n_2) at the significance level $\alpha = 0.05$.

F_{X_1}	F_{Y_1}	F_{Z_2}	n_1	n_2	Proposed	MLR

For Peer Review

					test at (15)	
$N(0, 0.5^2)$	$N(0.5, 1)$	$N(0,1)$				
			10	10	0.2325	0.2351
			50	50	0.7989	0.8560
$N(0, 0.5^2)$	$N(1.5, 1)$	$N(0,1)$				
			10	10	0.9564	0.9593
			15	15	0.9957	0.9967
$N(0, 1)$	$N(1.5, 1)$	$N(0,1)$				
			10	10	0.877	0.8991
			25	25	0.9995	0.9997
$N(1, 0.5^2)$	$N(2, 1.5^2)$	$N(0,1)$				
			10	10	0.5219	0.6194
			50	50	0.9975	0.9997

Table 5. The Monte Carlo powers of Test 3 by (17) vs. the MLR test for H_0 *vs.* H_{A3} with different sample sizes (n_1, n_2) at the significance level $\alpha = 0.05$.

F_{X_1}	F_{Y_1}	F_{X_2}	F_{Y_2}	n_1	n_2	Proposed test at (17)	MLR
$N(1, 1)$	$N(1.5, 1.5^2)$	$N(1, 1)$	$N(1.5, 1.5^2)$				
				10	10	0.2058	0.2191
				50	50	0.6709	0.7919
$N(0, 0.5^2)$	$N(0.6, 1)$	$N(0, 0.5^2)$	$N(0.6, 1)$				
				10	10	0.5922	0.6377
				50	50	0.9974	0.9996
$N(0.5, 0.25^2)$	$N(1, 1)$	$N(0.5, 0.25^2)$	$N(1, 1)$				
				10	10	0.5072	0.5444
				50	50	0.9869	0.9977
$N(2.5, 1.25^2)$	$N(2, 0.5^2)$	$N(2.5, 1.25^2)$	$N(2, 0.5^2)$				
				10	10	0.3383	0.3509
				50	50	0.9004	0.9582

Table 6. The Monte Carlo type I errors of the MLR tests

F_{Z_1}	F_{Z_2}	n_1	n_2	MLR test for H_0 *vs.* H_{A1}	MLR test for H_0 *vs.* H_{A2}	MLR test for H_0 *vs.* H_{A3}
t_3	t_3					
		10	10	0.1599	0.1838	0.0428
		50	50	0.2934	0.3260	0.0465
t_5	t_5					
		10	10	0.0955	0.1066	0.0458
		50	50	0.1447	0.1687	0.0488
t_{200}	t_{200}					
		10	10	0.0507	0.0493	0.0499
		50	50	0.0503	0.0506	0.0502
Logistic(0,1)	*Logistic*(0,1)					
		10	10	0.0717	0.0759	0.0463
		50	50	0.0855	0.0968	0.0508
Laplace(0,1)	*Laplace*(0,1)					
		10	10	0.1108	0.1293	0.0438
		50	50	0.1446	0.1645	0.0488
Unif[0, 1]	*Unif*[0, 1]					
		10	10	1	0.9993	1

Table 7. The Monte Carlo powers of the proposed test (9) vs. the combined nonparametric test (the two Wilcoxon signed rank tests and one Kolmogorov-Smirnov test) at the significance level $\alpha = 0.05$.

For Peer Review

F_{X_1}	F_{Y_1}	F_{X_2}	F_{Y_2}	n_1	n_2	Proposed test at (9)	W and K-S tests
Exp(1)	*Lognorm*(0, 2^2)	*N*(0, 1)	*N*(0.5, 1.5^2)				
				10	10	0.2238	0.0946
				50	50	0.9616	0.6273
Lognorm(1, 1)	*Lognorm*(1, 0.5^2)	*N*(0,1)	*N*(1.5, 2^2)				
				10	10	0.3646	0.2754
				50	50	0.9953	0.9849
Exp(3)	*Lognorm*(0, 2^2)	*Gamma*(5,1)	*Gamma*(1, 5)				
				10	10	0.6321	0.5016
				50	50	1	1
$\chi^2_{(6)}$	*Gamma*(1,10)	*N*(0,1)	*N*(0.5, 2^2)				
				10	10	0.3815	0.1179
				50	50	1	0.8576
Exp(1)	*Cauchy*(1,1)	*N*(0.5, 1)	*N*(1.5, 2^2)				
				10	10	0.1819	0.1255
				50	50	0.7928	0.7426
Exp(1)	*Lognorm*(0, 2^2)	*Unif*[-1, 1]	*Unif*[-1, 1]				
				10	10	0.2325	0.0836
				50	50	0.9981	0.6939

Table 8. The Monte Carlo powers of the proposed test (15) vs. the combined nonparametric test (the one Wilcoxon signed rank test and one Kolmogorov-Smirnov test) at the significance level $\alpha = 0.05$.

F_{X_1}	F_{Y_1}	F_{Z_2}	n_1	n_2	Proposed test at (15)	W and K-S tests
Exp(3)	*N*(1.5, 2^2)	*N*(0, 1)				
			10	10	0.5638	0.2876
			50	50	0.9995	0.9816
Exp(1)	*Beta*(1,1)	*N*(0, 1)				
			10	10	0.2256	0.1193
			50	50	0.9983	0.8046
Exp(1)	*Cauchy*(1,1)	*N*(0, 1)				
			10	10	0.1613	0.0401
			50	50	0.9736	0.2323
Exp(1.5)	*N*(0.5,1)	*Unif*[-1, 1]				
			10	10	0.1677	0.0384
			50	50	0.9984	0.3042
Exp(1.5)	*Beta*(3,1)	*Unif*[-1, 1]				
			10	10	0.1328	0.0934
			50	50	0.8774	0.4128
Exp(1)	*Cauchy*(1,1)	*Unif*[-1, 1]				
			10	10	0.4052	0.0608
			25	25	0.9952	0.9042
Lognorm(1, 1)	*Lognorm*(1.2, 1)	t_2				
			10	10	0.2892	0.0657
			50	50	0.8807	0.6635
$\chi^2_{(6)}$	*Gamma*(10,1)	t_2				
			10	10	0.9044	0.6699
			25	25	0.9999	0.9914
Exp(1)	*Cauchy*(1,1)	t_2				
			10	10	0.0792	0.0316
			50	50	0.2891	0.0877

For Peer Review

Table 9. The Monte Carlo powers of the proposed test (17) vs. the Wilcoxon signed rank test at the significance level $\alpha = 0.05$.

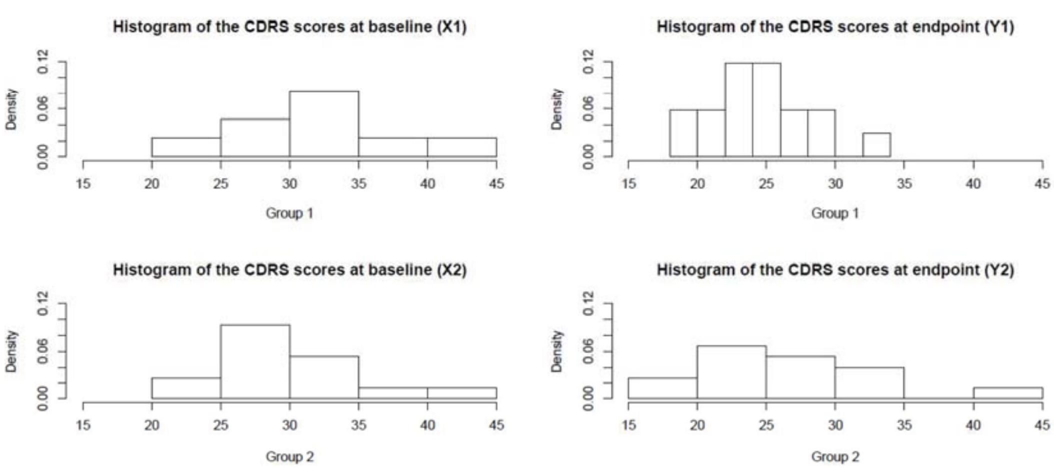
F_{X_1}	F_{Y_1}	F_{X_2}	F_{Y_2}	n_1	n_2	Proposed test at (17)	W test
Exp(1)	*Lognorm*(0, 2^2)	*Exp*(1)	*Lognorm*(0, 2^2)				
				10	10	0.4136	0.2933
				50	50	0.9992	0.9223
Lognorm(1, 1)	*Lognorm*(1, 0.5^2)	*Lognorm*(1, 1)	*Lognorm*(1, 0.5^2)				
				10	10	0.1218	0.0731
				50	50	0.7906	0.2208
Gamma(5,1)	*Gamma*(1, 5)	*Gamma*(5,1)	*Gamma*(1, 5)				
				10	10	0.1218	0.0886
				50	50	0.8074	0.2294
$\chi^2_{(6)}$	*Gamma*(1,10)	$\chi^2_{(6)}$	*Gamma*(1,10)				
				10	10	0.3125	0.2344
				50	50	0.9629	0.8517
Beta(0, 0.8)	*Exp*(1.5)	*Beta*(0, 0.8)	*Exp*(1.5)				
				10	10	0.2094	0.1518
				50	50	0.9928	0.6003

Table 10. The proportions of rejections[a] based on the bootstrap method for each considered test.

Bootstrapped sample sizes	Test 1		Test 2	
(n_1, n_2)	Proposed test (9)	Classic test[b]	Proposed test (15)	Classic test[c]
(9, 6)	0.9858	0.7135	0.9755	0.9134
(9, 7)	0.9870	0.7172	0.9800	0.9165
(11, 9)	0.9989	0.9795	0.9955	0.9844
(13, 10)	0.9998	0.9962	0.9993	0.9972
(13, 11)	0.9999	0.9965	0.9997	0.9975
(15, 13)	1	0.9997	0.9999	0.9994

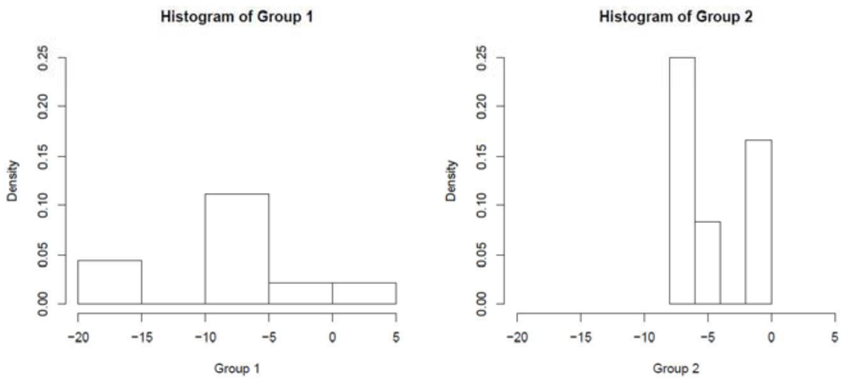
a. The proportion of rejections of each test from the bootstrap method was computed based on sample sizes (n_1, n_2) and 10,000 replications; b. The combined classic test for H_0 *vs*. H_{A1} is based on two W-tests and one K-S test; c. The combined classic test for H_0 *vs*. H_{A2} is based on one W-test and one K-S test.

Figure 1. Histograms of CDRS-Rts related to the baseline and endpoint in group 1: (X_1, Y_1), and in group 2: (X_2, Y_2).



For Peer Review

Figure 2. Histograms of the paired observations Z_1 and Z_2 based on the CDRS-Rts data, with sample sizes (n_1, n_2)=(9, 6), that were sampled from the original data set.



For Peer Review

For Peer Review

Supplementary material (Proofs and R Codes) to

Two-sample density-based empirical likelihood ratio tests based on paired data, with application to a treatment study of Attention-Deficit/Hyperactivity Disorder and Severe Mood Dysregulation

Albert Vexler[a*†], Wan-Min Tsai[a], Gregory Gurevich[b] and Jihnhee Yu[a]

S1:

Appendix A1.

Proof of proposition 1

A1.1 Proposed test 1

Consider the case of testing H_0 *vs.* H_{A1} $(t = 1)$ in Proposition 1. Towards this end, we define

$$V^{*}_{n_1 m} = \sum_{j=1}^{n_1}\left[\log\left(\frac{2m}{n_1 \eta_{m,j}}\right) + \log\left(1 - \frac{m+1}{2n_1}\right)\right],\ \ V^{**}_{n_2 k} = \sum_{j=1}^{n_2}\left[\log\left(\frac{2k}{n_2 \varphi_{k,j}}\right) + \log\left(1 - \frac{k+1}{2n_2}\right)\right]. \qquad \text{(A1.1.1)}$$

Then the proposed test statistic at (9), $(n_1 + n_2)^{-1}\log\left(V^{H_{A1}}_{n_1 n_2}\right)$, can be expressed as

$$(n_1 + n_2)^{-1}\log(V^{H_{A1}}_{n_1 n_2}) =$$

$$\min_{{n_1}^{0.5+\delta} \le m \le {n_1}^{1-\delta}} (n_1 + n_2)^{-1} V^{*}_{n_1 m} + \min_{{n_2}^{0.5+\delta} \le k \le {n_2}^{1-\delta}} (n_1 + n_2)^{-1} V^{**}_{n_2 k}. \qquad \text{(A1.1.2)}$$

We first investigate the first term of the right-hand side of the equation (A1.1.2). To this end, we define the distribution function $Q(x)$ to be

$$Q(x) = \left(n_1 F_1(x) + n_2 F_2(x)\right)/(n_1 + n_2),$$

where

$$F_1(x) = \left(F_{Z_1}(x) + 1 - F_{Z_1}(-x)\right)/2 \text{ and } F_2(x) = \left(F_{Z_2}(x) + 1 - F_{Z_2}(-x)\right)/2.$$

Also, an empirical distribution function is defined by

[a] *Department of Biostatisics, The State University of New York, Buffalo, NY 14214, U.S.A.*
[b] *The Department of Industrial Engineering and Management, SCE- Shamoon College of Engineering, Beer-Sheva 84100, Israel*
[*] *Correspondence to: Albert Vexler, Department of Biostatisics, The State University of New York, Buffalo, NY 14214, U.S.A.*
[†] *Email: avexler@buffalo.edu*

$$G_{n_1+n_2}(x) = \left(n_1 G_{n_1}(x) + n_2 G_{n_2}(x)\right)/(n_1+n_2),$$

where

$$G_{n_1}(x) = \left({n_1}^{-1}\sum_{j=1}^{n_1} I\left(Z_{1j} \le x\right) + {n_1}^{-1}\sum_{j=1}^{n_1} I\left(-Z_{1j} \le x\right)\right)\Big/2$$

and

$$G_{n_2}(x) = \left({n_2}^{-1}\sum_{j=1}^{n_2} I\left(Z_{2j} \le x\right) + {n_2}^{-1}\sum_{j=1}^{n_2} I\left(-Z_{2j} \le x\right)\right)\Big/2$$

are the empirical distribution functions based on observations $Z_{11},\ldots,Z_{1n_1}$ and $Z_{21},\ldots,Z_{2n_2}$, respectively. Consequently, the first term of $V^*_{n_1m}$ in (A1.1.1) can be reformulated as

$$\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{2m}{n_1\eta_{m,j}}\right) = -\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{n_1\left[G_{n_1+n_2}\left(Z_{1(j+m)}\right) - G_{n_1+n_2}\left(Z_{1(j-m)}\right)\right]}{2m}\right)$$

$$= -\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{Q\left(Z_{1(j+m)}\right) - Q\left(Z_{1(j-m)}\right)}{F_1\left(Z_{1(j+m)}\right) - F_1\left(Z_{1(j-m)}\right)}\right)$$

$$+\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{Q\left(Z_{1(j+m)}\right) - Q\left(Z_{1(j-m)}\right)}{G_{n_1+n_2}\left(Z_{1(j+m)}\right) - G_{n_1+n_2}\left(Z_{1(j-m)}\right)}\right)$$

$$-\frac{1}{n_1+n_2}\sum_{j=1}^{n_1} log\left(\frac{n_1\left[F_1\left(Z_{1(j+m)}\right) - F_1\left(Z_{1(j-m)}\right)\right]}{2m}\right). \quad \text{(A1.1.3)}$$

The first term in the right-hand of the equation (A1.1.3) can be expressed as

$$\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{Q\left(Z_{1(j+m)}\right) - Q\left(Z_{1(j-m)}\right)}{F_1\left(Z_{1(j+m)}\right) - F_1\left(Z_{1(j-m)}\right)}\right) = \frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{Q\left(Z_{1(j+m)}\right) - Q\left(Z_{1(j-m)}\right)}{Z_{1(j+m)} - Z_{1(j-m)}}\right)$$

$$-\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{F_1\left(Z_{1(j+m)}\right) - F_1\left(Z_{1(j-m)}\right)}{Z_{1(j+m)} - Z_{1(j-m)}}\right). \quad \text{(A1.1.4)}$$

The result shown in Theorem 1 of Vasicek [29] leads to

$$\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{Q\left(Z_{1(j+m)}\right) - Q\left(Z_{1(j-m)}\right)}{Z_{1(j+m)} - Z_{1(j-m)}}\right) = \frac{n_1}{n_1+n_2}\frac{1}{2m}\sum_{d=1}^{2m} S_d, \quad \text{(A1.1.5)}$$

where

For Peer Review

$$S_d = \sum_{j=1}^{n_1} \log\left(\frac{Q\left(Z_{1(j+m)}\right)-Q\left(Z_{1(j-m)}\right)}{Z_{1(j+m)}-Z_{1(j-m)}}\right)\left(F_1\left(Z_{1(j+m)}\right) - F_1\left(Z_{1(j-m)}\right)\right), j \equiv d \text{ (mod 2m)}.$$

Let $f_i(x) = dF_i(x)/dx = \left(f_{Z_i}(x) + f_{Z_i}(-x)\right)/2, i = 1, 2.$

Suppose $Z_{1(j-m)}$ and $Z_{1(j+m)}$ is within an interval in which

$$q(x) = dQ(x)/dx = n_1 f_1(x)/(n_1 + n_2) + n_2 f_2(x)/(n_1 + n_2)$$

is positive and continuous, then

$$q(Z'_d) = \frac{Q\left(Z_{1(j+m)}\right) - Q\left(Z_{1(j-m)}\right)}{Z_{1(j+m)} - Z_{1(j-m)}},$$

for some existing value $Z'_d \in (Z_{1(j-m)}, Z_{1(j+m)})$. (The assumption that $q(x)$ is positive and continuous, when $x \in \left(Z_{1(j-m)}, Z_{1(j+m)}\right)$, is used to simplify the proof and this condition can be excluded, for example, see the proof scheme applied in [29].) It follows that S_d can be written as

$$S_d = \sum_{j=1}^{n_1} \log\left(q(Z'_d)\right)\left(F_1\left(Z_{1(j+m)}\right) - F_1\left(Z_{1(j-m)}\right)\right), j \equiv d \text{ (mod 2m)}.$$

Let us define a density function $\bar{q}(x)$ that approximates $q(x)$ as follows:

$$\bar{q}(x) = \gamma f_1(x)/(1+\gamma) + f_2(x)/(1+\gamma).$$

For each $\varepsilon > 0$ and sufficiently large n_1 and n_2,

$$n_1/n_2 \to \gamma \text{ so that we have } (1-\varepsilon)\bar{q}(x) \le q(x) \le (1+\varepsilon)\bar{q}(x).$$

It follows that for sufficiently large n_1 and n_2,

$$S_{d,(-\varepsilon)} \le S_d \le S_{d,\varepsilon},$$

where $S_{d,(-\varepsilon)} = \sum_{j=1}^{n_1} \log\left((1-\varepsilon)\bar{q}(Z'_d)\right)\left(F_1\left(Z_{1(j+m)}\right) - F_1\left(Z_{1(j-m)}\right)\right)$,

and $S_{d,\varepsilon} = \sum_{j=1}^{n_1} \log\left((1+\varepsilon)\bar{q}(Z'_d)\right)\left(F_1\left(Z_{1(j+m)}\right) - F_1\left(Z_{1(j-m)}\right)\right)$.

i.e. $S_{d,(-\varepsilon)}$ and $S_{d,\varepsilon}$ are Stieltjes sums of the function $\log((1-\varepsilon)\bar{q}(Z_{11}))$ and $\log((1+\varepsilon)\bar{q}(Z_{11}))$, respectively, with respect to the measure F_1 over the sum of intervals of continuity of $f_1(x)$ and $f_2(x)$ in which $\bar{q}(x) > 0$. Since in any interval in which $\bar{q}(x)$ is positive, $Z_{1(j+m)} - Z_{1(j-m)} \to 0$ as $n_1 \to \infty$ uniformly over $m \in \left[{n_1}^{0.5+\delta}, {n_1}^{1-\delta}\right]$, $\delta \in (0,1/4)$, and uniformly over Z_1, $S_{d,(-\varepsilon)}$ converges in probability to $\int_{-\infty}^{\infty} \log\left((1-\varepsilon)\bar{q}(Z_{11})\right) dQ(Z_{11}) = E\{\log((1-\varepsilon)\bar{q}(Z_{11}))\}$ as $n_1 \to \infty$. Furthermore, this convergence is uniformly over j

and $n_1{}^{0.5+\delta} \leq m \leq n_1{}^{1-\delta}$, $\delta \in (0,1/4)$. Similarly, $S_{d,\varepsilon}$ converges in probability to $E\{\log((1+\varepsilon)\bar{q}(Z_{11}))\}$, uniformly over j and $n_1{}^{0.5+\delta} \leq m \leq n_1{}^{1-\delta}$, $\delta \in (0,1/4)$.

Therefore,

$$E\{\log((1-\varepsilon)\bar{q}(Z_{11}))\} \leq \frac{1}{2m}\sum_{t=1}^{2m} S_t \leq E\{\log((1+\varepsilon)\bar{q}(Z_{11}))\},$$

as $n_1 \to \infty$, uniformly over $n_1{}^{0.5+\delta} \leq m \leq n_1{}^{1-\delta}$, $\delta \in (0,1/4)$.

Recalling from (A1.1.5), we find

$$(n_1+n_2)^{-1}\sum_{j=1}^{n_1}\log\left(\frac{Q\left(Z_{1(j+m)}\right)-Q\left(Z_{1(j-m)}\right)}{Z_{1(j+m)}-Z_{1(j-m)}}\right)\xrightarrow{P}\frac{\gamma}{1+\gamma}E\{\log(\bar{q}(Z_{11}))\}, \qquad \text{(A1.1.6)}$$

as $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \gamma > 0$, uniformly over $n_1{}^{0.5+\delta} \leq m \leq n_1{}^{1-\delta}$, $\delta \in (0,1/4)$.

Similarly, we have

$$(n_1+n_2)^{-1}\sum_{j=1}^{n_1}\log\left(\frac{F_1\left(Z_{1(j+m)}\right)-F_1\left(Z_{1(j-m)}\right)}{Z_{1(j+m)}-Z_{1(j-m)}}\right)\xrightarrow{P}\frac{\gamma}{1+\gamma}E\left\{\log\left(f_1(Z_{11})\right)\right\}, \qquad \text{(A1.1.7)}$$

as $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \gamma > 0$, uniformly over $n_1{}^{0.5+\delta} \leq m \leq n_1{}^{1-\delta}$, $\delta \in (0,1/4)$.

Combining the results of (A1.1.4), (A1.1.6), and (A1.1.7) yields

$$(n_1+n_2)^{-1}\sum_{j=1}^{n_1}\log\left(\frac{Q\left(Z_{1(j+m)}\right)-Q\left(Z_{1(j-m)}\right)}{F_1\left(Z_{1(j+m)}\right)-F_1\left(Z_{1(j-m)}\right)}\right)$$

$$\xrightarrow{P}\frac{\gamma}{1+\gamma}E\left(\log\left(\frac{\bar{q}(Z_{11})}{f_1(Z_{11})}\right)\right)=\frac{\gamma}{1+\gamma}E\log\left\{\frac{\gamma}{1+\gamma}+\frac{1}{1+\gamma}\left(\frac{f_2(Z_{11})}{f_1(Z_{11})}\right)\right\}, \qquad \text{(A1.1.8)}$$

as $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \gamma > 0$, uniformly over $n_1{}^{0.5+\delta} \leq m \leq n_1{}^{1-\delta}$, $\delta \in (0,1/4)$.

Now, we consider the second term in the right-hand side of the equation (A1.1.3). By Theorem A in Serfling [31], we know that for $0 \leq \epsilon \leq \delta/2$,

$$\Pr\left(\sup_{-\infty<x<\infty}\left|F_1(x)-F_{n_1}(x)\right| > n_1{}^{-0.5+\epsilon}\right)\xrightarrow{n_1\to\infty}0,$$

and

$$\Pr\left(\sup_{-\infty<x<\infty}\left|F_2(x)-F_{n_2}(x)\right| > n_2{}^{-0.5+\epsilon}\right)\xrightarrow{n_2\to\infty}0,$$

implying that $\Pr\left(\sup_{-\infty<x<\infty}\left|F_2(x)-F_{n_2}(x)\right|>(2n_1/\gamma)^{-0.5+\epsilon}\right)\xrightarrow{n_1\to\infty,n_2\to\infty,n_1/n_2\to\gamma}0$.

Hence, $\Pr\left(\sup_{-\infty<x<\infty}\left|Q(x)-G_{n_1+n_2}(x)\right|>{n_1}^{-0.5+2\epsilon}\right)\xrightarrow{n_1\to\infty,n_2\to\infty}0$, for $0\le\epsilon\le\delta/2$.

Now we consider the case when $\sup_{-\infty<x<\infty}\left|Q(x)-G_{n_1+n_2}(x)\right|\le{n_1}^{-0.5+2\epsilon}$. According to the definition of $G_{n_1+n_2}(x)$, we have the inequality

$$G_{n_1+n_2}\left(Z_{1(j+m)}\right)-G_{n_1+n_2}\left(Z_{1(j-m)}\right)\ge n_1(n_1+n_2)^{-1}2m(n_1)^{-1}=2m/(n_1+n_2).$$

Thus, for the case of $\sup_{-\infty<x<\infty}\left|Q(x)-G_{n_1+n_2}(x)\right|\le{n_1}^{-0.5+2\epsilon}$, we have

$$\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{Q\left(Z_{1(j+m)}\right)-Q\left(Z_{1(j-m)}\right)}{G_{n_1+n_2}\left(Z_{1(j+m)}\right)-G_{n_1+n_2}\left(Z_{1(j-m)}\right)}\right)$$

$$\le\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{G_{n_1+n_2}\left(Z_{1(j+m)}\right)-G_{n_1+n_2}\left(Z_{1(j-m)}\right)+{n_1}^{-0.5+\delta/2}}{G_{n_1+n_2}\left(Z_{1(j+m)}\right)-G_{n_1+n_2}\left(Z_{1(j-m)}\right)}\right)$$

$$\le\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(1+\frac{{n_1}^{-0.5+\frac{\delta}{2}}}{2m/(n_1+n_2)}\right)\le\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\left(\frac{{n_1}^{-0.5+\delta/2}}{2{n_1}^{0.5+\delta}/(n_1+n_2)}\right)=\frac{1}{2{n_1}^{\delta/2}}\to 0,$$

for a sufficiently large n_1 and $m\in\left[{n_1}^{0.5+\delta},{n_1}^{1-\delta}\right]$.

Also, note that

$$\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{Q\left(Z_{1(j+m)}\right)-Q\left(Z_{1(j-m)}\right)}{G_{n_1+n_2}\left(Z_{1(j+m)}\right)-G_{n_1+n_2}\left(Z_{1(j-m)}\right)}\right)$$

$$\ge\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(\frac{G_{n_1+n_2}\left(Z_{1(j+m)}\right)-G_{n_1+n_2}\left(Z_{1(j-m)}\right)-{n_1}^{-0.5+\delta/2}}{G_{n_1+n_2}\left(Z_{1(j+m)}\right)-G_{n_1+n_2}\left(Z_{1(j-m)}\right)}\right)$$

$$\ge\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\log\left(1-\frac{{n_1}^{-0.5+\frac{\delta}{2}}}{2m/(n_1+n_2)}\right)\ge-\frac{1}{n_1+n_2}\sum_{j=1}^{n_1}\left(\frac{2{n_1}^{-0.5+\delta/2}}{2{n_1}^{0.5+\delta}/(n_1+n_2)}\right)=-\frac{1}{{n_1}^{\delta/2}}\to 0,$$

for a sufficiently large n_1 and $m \in \left[{n_1}^{0.5+\delta}, {n_1}^{1-\delta}\right]$. Hence, we prove that the second term in the right-hand side of the equality (A1.1.3) converges to zero in probability. That is,

$$\frac{1}{n_1+n_2}\sum_{j=1}^{n_1} \log\left(\frac{Q\left(Z_{1(j+m)}\right)-Q\left(Z_{1(j-m)}\right)}{G_{n_1+n_2}\left(Z_{1(j+m)}\right)-G_{n_1+n_2}\left(Z_{1(j-m)}\right)}\right) \overset{P}{\to} 0, \quad \text{(A1.1.9)}$$

as $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \gamma > 0$, uniformly over ${n_1}^{0.5+\delta} \leq m \leq {n_1}^{1-\delta}$.

Finally, using the result of Lemma 1 of Vasicek [29], the last term in the right-hand side of the equality of (A1.1.3) also converges to zero in probability. That is,

$$-(n_1+n_2)^{-1}\sum_{j=1}^{n_1} \log\left(\frac{n_1\left[F_1\left(Z_{1(j+m)}\right)-F_1\left(Z_{1(j-m)}\right)\right]}{2m}\right) \overset{P}{\to} 0, \quad \text{(A1.1.10)}$$

as $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \gamma > 0$, uniformly over ${n_1}^{0.5+\delta} \leq m \leq {n_1}^{1-\delta}$, $\delta \in (0,1/4)$.

By (A1.1.8), (A1.1.9), and (A1.1.10), we show that

$$(n_1+n_2)^{-1}V_{n_1 m}^{*} \overset{P}{\to} -\frac{\gamma}{1+\gamma} E\log\left\{\frac{\gamma}{1+\gamma}+\frac{1}{1+\gamma}\left(\frac{f_2(Z_{11})}{f_1(Z_{11})}\right)\right\}, \quad \text{(A1.1.11)}$$

as $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \gamma > 0$, uniformly over ${n_1}^{0.5+\delta} \leq m \leq {n_1}^{1-\delta}$, $\delta \in (0, 1/4)$.

Likewise, following the same procedure as shown in the proof of $V_{n_1 m}^{*}$, we have

$$(n_1+n_2)^{-1}V_{n_2 k}^{**} = (n_1+n_2)^{-1}\sum_{j=1}^{n_2}\left[\log\left(\frac{2k}{n_2\varphi_{k,j}}\right)+\log\left(1-\frac{k+1}{2n_2}\right)\right] \overset{P}{\to}$$

$$-\frac{1}{1+\gamma}E\log\left\{\frac{\gamma}{1+\gamma}\left(\frac{f_1(Z_{21})}{f_2(Z_{21})}\right)+\frac{1}{1+\gamma}\right\},$$

as $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \gamma > 0$, uniformly over ${n_2}^{0.5+\delta} \leq k \leq {n_2}^{1-\delta}$, $\delta \in (0,1/4)$.

This and (A1.1.11) conclude that

$$(n_1+n_2)^{-1}\log(V_{n_1 n_2}^{H_{A1}}) \overset{p}{\to} -\frac{\gamma}{1+\gamma}E\log\left\{\frac{\gamma}{1+\gamma}+\frac{1}{1+\gamma}\left(\frac{f_2(Z_{11})}{f_1(Z_{11})}\right)\right\}$$

$$-\frac{1}{1+\gamma}E\log\left\{\frac{\gamma}{1+\gamma}\left(\frac{f_1(Z_{21})}{f_2(Z_{21})}\right)+\frac{1}{1+\gamma}\right\},$$

as $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \gamma > 0$.

Hence, under H_0,

For Peer Review

$$\frac{f_2(Z_{11})}{f_1(Z_{11})} = \frac{\left(f_{Z_2}(Z_{11}) + f_{Z_2}(-Z_{11})\right)/2}{\left(f_{Z_1}(Z_{11}) + f_{Z_1}(-Z_{11})\right)/2} = \frac{f_{Z_2}(Z_{11})}{f_{Z_1}(Z_{11})},$$

$$\frac{f_1(Z_{21})}{f_2(Z_{21})} = \frac{\left(f_{Z_1}(Z_{21}) + f_{Z_1}(-Z_{21})\right)/2}{\left(f_{Z_2}(Z_{21}) + f_{Z_2}(-Z_{21})\right)/2} = \frac{f_{Z_1}(Z_{21})}{f_{Z_2}(Z_{21})},$$

$$(n_1+n_2)^{-1}\log(V_{n_1n_2}^{H_{A1}}) \xrightarrow{p} -\frac{\gamma}{1+\gamma}E_{H_0}\log\left\{\frac{\gamma}{1+\gamma}+\frac{1}{1+\gamma}\left(\frac{f_{Z_2}(Z_{11})}{f_{Z_1}(Z_{11})}\right)\right\}$$

$$-\frac{1}{1+\gamma}E_{H_0}\log\left\{\frac{\gamma}{1+\gamma}\left(\frac{f_{Z_1}(Z_{21})}{f_{Z_2}(Z_{21})}\right)+\frac{1}{1+\gamma}\right\} = 0,$$

and under H_{A1},

$$(n_1+n_2)^{-1}\log(V_{n_1n_2}^{H_{A1}}) \xrightarrow{p} -\frac{\gamma}{1+\gamma}E_{H_{A1}}\log\left\{\frac{\gamma}{1+\gamma}+\frac{1}{1+\gamma}\left(\frac{f_2(Z_{11})}{f_1(Z_{11})}\right)\right\}$$

$$-\frac{1}{1+\gamma}E_{H_{A1}}\log\left\{\frac{\gamma}{1+\gamma}\left(\frac{f_1(Z_{21})}{f_2(Z_{21})}\right)+\frac{1}{1+\gamma}\right\}$$

$$\geq -\frac{\gamma}{1+\gamma}\log\left\{\frac{\gamma}{1+\gamma}+\frac{1}{1+\gamma}E_{H_{A1}}\left(\frac{f_2(Z_{11})}{f_1(Z_{11})}\right)\right\} - \frac{1}{1+\gamma}\log\left\{\frac{\gamma}{1+\gamma}E_{H_{A1}}\left(\frac{f_1(Z_{21})}{f_2(Z_{21})}\right)+\frac{1}{1+\gamma}\right\}$$

≥ 0, as $n_1\to\infty$, $n_2\to\infty$, $n_1/n_2\to\gamma>0$.

We complete the proof of Proposition 1 for the case of $t=1$, i.e. the consistency related to the proposed Test 1.

A1.2 Proposed test 2

Here, we will consider the case of $t=2$. That is, we will show that

$$(n_1+n_2)^{-1}\log(V_{n_1n_2}^{H_{A2}}) \xrightarrow{p} -\frac{\gamma}{1+\gamma}E\log\left\{\frac{\gamma}{1+\gamma}+\frac{1}{1+\gamma}\left(\frac{f_2(Z_{11})}{f_1(Z_{11})}\right)\right\}$$

$$-\frac{1}{1+\gamma}E\log\left\{\frac{\gamma}{1+\gamma}\left(\frac{f_1(Z_{21})}{f_2(Z_{21})}\right)+\frac{1}{1+\gamma}\right\},$$

as $n_1\to\infty$, $n_2\to\infty$, $n_1/n_2\to\gamma>0$.

It is clear that if one can show that $\log\left(\Lambda_{n_2}^k\right)\xrightarrow{p}0$ as $n_2\to\infty$, where $\Lambda_{n_2}^k$ is defined by (12), the rest of the proof is similar to the proof shown in Section A1.1 regarding the test statistic of the proposed Test 1, $\log(V_{n_1n_2}^{H_{A1}})$. To consider $\log\left(\Lambda_{n_2}^k\right)$, as $n_2\to\infty$, we begin with a proof that

For Peer Review

$$\tilde{F}_{Z_2}\left(Z_{2(n_2)}\right) - \tilde{F}_{Z_2}\left(Z_{2(1)}\right) = \frac{1}{2n_2}\sum_{j=1}^{n_2}\left[I\left(Z_{2j} \le Z_{2(n_2)}\right) + I\left(-Z_{2j} \le Z_{2(n_2)}\right)\right]$$

$$-\frac{1}{2n_2}\sum_{j=1}^{n_2}\left[I\left(Z_{2j} \le Z_{2(1)}\right) + I\left(-Z_{2j} \le Z_{2(1)}\right)\right] \xrightarrow{p} 1, \text{as } n_2 \to \infty.$$

To this end, we apply Theorem A of Serfling [31], having that for $\epsilon \in (0, 1/2)$,

$$\sup_{-\infty<u<\infty}\left|\tilde{F}_{Z_2}(u) - F_{Z_2}(u)\right| = o({n_2}^{-0.5+\epsilon}) \text{ as } n_2 \to \infty.$$

Thus, $\tilde{F}_{Z_2}\left(Z_{2(n_2)}\right) - \tilde{F}_{Z_2}\left(Z_{2(1)}\right) = F_{Z_2}\left(Z_{2(n_2)}\right) - F_{Z_2}\left(Z_{2(1)}\right) + o({n_2}^{-0.5+\epsilon})$.

It is obvious that $F_{Z_2}\left(Z_{2(n_2)}\right) \to 1$ and $F_{Z_2}\left(Z_{2(1)}\right) \to 0$ as $n_2 \to \infty$.

Hence,

$$\tilde{F}_{Z_2}\left(Z_{2(n_2)}\right) - \tilde{F}_{Z_2}\left(Z_{2(1)}\right) \xrightarrow{p} 1 \text{ as } n_2 \to \infty. \quad \text{(A1.2.1)}$$

Next, we will show that the part of $\Lambda_{n_2}^k$, $\sum_{r=1}^{k-1}(2k)^{-1}(k-r)\sum_{j=1}^{n_2}\left[\tilde{F}_{Z_2}\left(Z_{2(r+1)}\right) - \tilde{F}_{Z_2}\left(Z_{2(r)}\right)\right] \xrightarrow{p} 0$ as $n_2 \to \infty$.

Let $\tilde{F}_{-Z_2}(u)$ denote the empirical distribution function of

$-Z_2$ distributed with $\left(1 - F_{Z_2}(-u)\right)$ (Here the symmetry of Z_2 distribution under H_0 and H_{A_2} is used). Then

$$\sum_{r=1}^{k-1}\frac{(k-r)}{2k}\left[\tilde{F}_{Z_2}\left(Z_{2(r+1)}\right) - \tilde{F}_{Z_2}\left(Z_{2(r)}\right)\right]$$

$$= \frac{1}{2n_2}\sum_{r=1}^{k-1}\frac{(k-r)}{2k}\sum_{j=1}^{n_2}\left[I\left(Z_{2j} \le Z_{2(r+1)}\right) + I\left(-Z_{2j} \le Z_{2(r+1)}\right) - I\left(Z_{2j} \le Z_{2(r)}\right) - I\left(-Z_{2j} \le Z_{2(r)}\right)\right]$$

$$= \frac{(k-r)}{4n_2} + \sum_{r=1}^{k-1}\frac{(k-r)}{2k}\left[\tilde{F}_{-Z_2}\left(Z_{2(r+1)}\right) - \tilde{F}_{-Z_2}\left(Z_{2(r)}\right)\right]. \quad \text{(A1.2.2)}$$

Since the first term of (A1.2.2), $(4n_2)^{-1}(k-r)$, vanishes to zero as $n_2 \to \infty$, we focus on the remaining terms of the equation (A1.2.2), which can be reorganized as follows:

$$\sum_{r=1}^{k-1}\frac{(k-r)}{2k}\left[\tilde{F}_{-Z_2}\left(Z_{2(r+1)}\right) - \tilde{F}_{-Z_2}\left(Z_{2(r)}\right)\right] = \sum_{r=1}^{k-1}\frac{(k-(r-1))}{2k}\tilde{F}_{-Z_2}\left(Z_{2(r)}\right) - \frac{k}{2k}\tilde{F}_{-Z_2}\left(Z_{2(1)}\right)$$

$$+\frac{(k-(k-1))}{2k}\tilde{F}_{-Z_2}\left(Z_{2(k)}\right) - \sum_{r=1}^{k-1}\frac{(k-r)}{2k}\tilde{F}_{-Z_2}\left(Z_{2(r)}\right)$$

$$= \frac{1}{2k}\sum_{r=1}^{k-1}\tilde{F}_{-Z_2}\left(Z_{2(r)}\right) - \frac{1}{2}\tilde{F}_{-Z_2}\left(Z_{2(1)}\right) + \frac{1}{2k}\tilde{F}_{-Z_2}\left(Z_{2(k)}\right). \quad \text{(A1.2.3)}$$

For Peer Review

In respect to the empirical distribution function, $\tilde{F}_{-Z_2}(u)$, appeared in (A1.2.3), again by virtue of Theorem A of Serfling [31], we have that for $\epsilon \in (0, 1/2)$,

$$\sum_{r=1}^{k-1} \frac{(k-r)}{2k}\left[\tilde{F}_{-Z_2}\left(Z_{2(r+1)}\right) - \tilde{F}_{-Z_2}\left(Z_{2(r)}\right)\right]$$

$$= \frac{1}{2k}\sum_{r=1}^{k-1} F_{-Z_2}\left(Z_{2(r)}\right) - \frac{1}{2}F_{-Z_2}\left(Z_{2(1)}\right) + \frac{1}{2k}F_{-Z_2}\left(Z_{2(k)}\right) + o\left({n_2}^{-0.5+\epsilon}\right). \quad \text{(A1.2.4)}$$

Clearly, $F_{-Z_2}\left(Z_{2(1)}\right)/2 \to 0$ and $F_{-Z_2}\left(Z_{2(k)}\right)/2k \to 0$ as $n_2 \to \infty$. Now, we prove that the first item of (A1.2.4) converges to zero in probability as $n_2 \to \infty$.

Since the distribution of $Z_{21}, \dots, Z_{2n_2}$ is symmetric under H_0 and the statistic, $\Lambda_{n_2}^k$, is based on $I\left(-Z_{2j} \le Z_{2(r)}\right)$ under H_0 and H_{A_2}, the distribution of $Z_{21}, \dots, Z_{2n_2}$ can be taken as the uniform distribution on the interval [-1, 1]. Thus, we obtain $F_{-Z_2}\left(Z_{2(r)}\right) = \left(1 + Z_{2(r)}\right)/2$, where $U_{(r)} = \left(1 + Z_{2(r)}\right)/2$ is the r^{th} order statistic based on a standard uniformly distributed, *Unif*[0,1], random variable. Since

$$E\left\{(2k)^{-1}\sum_{r=1}^{k-1} U_{(r)}\right\} = \frac{k(k-1)}{4k(n_2+1)} \to 0, \text{as } n_2 \to \infty,$$

applying the Chebyshev's inequality yields $(2k)^{-1}\sum_{r=1}^{k-1} U_{(r)} \xrightarrow{p} 0$ as $n_2 \to \infty$.

Combining (A1.2.2)-(A1.2.4), we conclude that

$$\sum_{r=1}^{k-1} \frac{(k-r)}{2k}\left[\tilde{F}_{Z_2}\left(Z_{2(r+1)}\right) - \tilde{F}_{Z_2}\left(Z_{2(r)}\right)\right] \xrightarrow{p} 0 \text{ as } n_2 \to \infty. \quad \text{(A1.2.5)}$$

Similarly, one can show that

$$\sum_{r=1}^{k-1} \frac{(k-r)}{2k}\left[\tilde{F}_{Z_2}\left(Z_{2(n_2-r+1)}\right) - \tilde{F}_{Z_2}\left(Z_{2(n_2-r)}\right)\right] \xrightarrow{p} 0 \text{ as } n_2 \to \infty. \quad \text{(A1.2.6)}$$

The results of (A1.2.1), (A1.2.5), and (A1.2.6) complete the proof of $\log\left(\Lambda_{n_2}^k\right) \xrightarrow{p} 0$ as $n_2 \to \infty$.

A2. Mathematical derivation of maximum likelihood ratio tests

A2.1 Maximum likelihood ratio test statistic for Test 1

Assume $Z_{ij} \sim i.i.d.\, N(\mu_{Z_i}, \sigma_{Z_i}^2)$, where μ_{Z_i} and $\sigma_{Z_i}^2$, i =1,2, are unknown. The following null hypothesis, H_0^{MLR}, is equivalent to the null hypothesis, H_0, that presented in the article

$$H_0^{MLR}: \mu_{Z_1} = \mu_{Z_2} = 0;\ \sigma_{Z_1}^2 = \sigma_{Z_2}^2 = \sigma^2.$$

Hence, the corresponding hypothesis of interest for Test 1 using the maximum likelihood ratio test is H_0^{MLR} *vs.* H_1^{MLR}: not H_0^{MLR}. Under normal assumptions, the MLR test statistic is given by

$$MLR_{A_1} = \frac{\max_{\mu_{Z_1},\sigma^2_{Z_1}} \prod_{j=1}^{n_1}\left(2\pi\sigma^2_{Z_1}\right)^{-1/2} e^{-\frac{(Z_{1j}-\mu_{Z_1})^2}{2\sigma^2_{Z_1}}} \max_{\mu_{Z_2},\sigma^2_{Z_2}} \prod_{j=1}^{n_2}\left(2\pi\sigma^2_{Z_2}\right)^{-1/2} e^{-\frac{(Z_{2j}-\mu_{Z_2})^2}{2\sigma^2_{Z_2}}}}{\max_{\sigma^2} \prod_{j=1}^{n_1}(2\pi\sigma^2)^{-1/2} e^{-\frac{Z_{1j}^2}{2\sigma^2}} \prod_{j=1}^{n_2}(2\pi\sigma^2)^{-1/2} e^{-\frac{Z_{2j}^2}{2\sigma^2}}}$$

$$= \frac{(2\pi\hat{\sigma}^2_{Z_1})^{-n_1/2}(2\pi\hat{\sigma}^2_{Z_2})^{-n_2/2} e^{-(n_1+n_2)/2}}{(2\pi\hat{\sigma}^2)^{-(n_1+n_2)/2} e^{-(n_1+n_2)/2}} = \frac{(\hat{\sigma}^2_{Z_1})^{-n_1/2}(\hat{\sigma}^2_{Z_2})^{-n_2/2}}{(\hat{\sigma}^2)^{-(n_1+n_2)/2}},$$

where the associated maximum likelihood estimators (MLEs) of μ_{Z_1}, μ_{Z_2}, $\sigma^2_{Z_1}$, $\sigma^2_{Z_2}$, and $\sigma^2$are $\hat{\mu}_{Z_1} = \sum_{j=1}^{n_1} Z_{1j}/n_1 = \bar{Z}_1$, $\hat{\mu}_{Z_2} = \sum_{j=1}^{n_2} Z_{2j}/n_2 = \bar{Z}_2$, $\hat{\sigma}^2_{Z_1} = \sum_{j=1}^{n_1}(Z_{1j} - \bar{Z}_1)^2/n_1$, $\hat{\sigma}^2_{Z_2} = \sum_{j=1}^{n_2}(Z_{2j} - \bar{Z}_2)^2/n_2$, and $\hat{\sigma}^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i} {Z_{ij}}^2/(n_1 + n_2)$, respectively.

A2.2 Maximum likelihood ratio test statistic for Test 2

Under the assumption that $Z_{ij} \sim i.i.d.\, N(\mu_{Z_i}, \sigma^2_{Z_i})$, where μ_{Z_i} and $\sigma^2_{Z_i}$, i =1,2, are unknown, the hypotheses H_0 *vs.* H_{A2} are equivalent to the following hypotheses:

$$H_0^{MLR}\ vs.\ H_{A_2}^{MLR}: \mu_{Z_1} \neq 0,\ \mu_{Z_2} = 0;\ \sigma^2_{Z_1} \neq \sigma^2_{Z_2}.$$

Thus, the maximum likelihood ratio for Test 2 can be formulated by

$$MLR_{A_2} = \frac{\max_{\mu_{Z_1},\sigma^2_{Z_1}} \prod_{j=1}^{n_1}\left(2\pi\sigma^2_{Z_1}\right)^{-1/2} e^{-\frac{(Z_{1j}-\mu_{Z_1})^2}{2\sigma^2_{Z_1}}} \max_{\sigma^2_{Z_2}} \prod_{j=1}^{n_2}\left(2\pi\sigma^2_{Z_2}\right)^{-1/2} e^{-\frac{Z_{2j}^2}{2\sigma^2_{Z_2}}}}{\max_{\sigma^2} \prod_{j=1}^{n_1}(2\pi\sigma^2)^{-1/2} e^{-\frac{Z_{1j}^2}{2\sigma^2}} \prod_{j=1}^{n_2}(2\pi\sigma^2)^{-1/2} e^{-\frac{Z_{2j}^2}{2\sigma^2}}}.$$

Substituting the associated MLEs of μ_{Z_1}, $\sigma^2_{Z_1}$, and σ^2 into the above likelihood ratio, MLR_{A_2}, yields the following maximum likelihood ratio test statistic for Test 2:

$$MLR_{A_2} = (\hat{\sigma}^2_{Z_1})^{-n_1/2}(\hat{\sigma}^2_{Z_2})^{-n_2/2}(\hat{\sigma}^2)^{(n_1+n_2)/2}$$

where $\hat{\mu}_{Z_1} = \sum_{j=1}^{n_1} Z_{1j}/n_1 = \bar{Z}_1$, $\hat{\sigma}^2_{Z_1} = \sum_{j=1}^{n_1}(Z_{1j} - \bar{Z}_1)^2/n_1$, $\hat{\sigma}^2_{Z_2} = \sum_{j=1}^{n_2} Z_{2j}^2/n_2$ and $\hat{\sigma}^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_i} {Z_{ij}}^2/(n_1 + n_2)$.

A2.3 Maximum likelihood ratio test statistic for the Test 3

Assume $Z_{ij} \sim i.i.d.\, N(\mu_{Z_i}, \sigma^2_{Z_i})$, where μ_{Z_i} and $\sigma^2_{Z_i}$, i =1,2, are unknown. The hypotheses: H_0 *vs.* H_{A3} are equivalent to the following hypotheses:

$$H_0^{MLR}\ vs.\ H_{A_3}^{MLR}: \mu_{Z_1} = \mu_{Z_2} = \mu_1 \neq 0;\ \sigma^2_{Z_1} = \sigma^2_{Z_2} = \sigma^2_1.$$

Accordingly, the corresponding MLR test statistic is

For Peer Review

$$MLR_{A_3} = \frac{\max_{\mu_1,\sigma_1^2} \prod_{j=1}^{n_1} (2\pi\sigma_1^2)^{-1/2} e^{-\frac{(Z_{1j}-\mu_1)^2}{2\sigma_1^2}} \prod_{j=1}^{n_2} (2\pi\sigma_1^2)^{-1/2} e^{-\frac{(Z_{2j}-\mu_1)^2}{2\sigma_1^2}}}{\max_{\sigma^2} \prod_{j=1}^{n_1} (2\pi\sigma^2)^{-1/2} e^{-\frac{Z_{1j}^2}{2\sigma^2}} \prod_{j=1}^{n_2} (2\pi\sigma^2)^{-1/2} e^{-\frac{Z_{2j}^2}{2\sigma^2}}}.$$

Replacing the parameters of μ_1, σ_1^2, and $\sigma^2$by their MLEs, the following maximum likelihood ratio test statistic for the Test 3 can be formulated by

$$MLR_{A_3} = (\hat{\sigma}_1^2/\hat{\sigma}^2)^{-(n_1+n_2)/2},$$

where $\hat{\mu}_1 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} Z_{ij}/(n_1+n_2) = \bar{Z}$, $\hat{\sigma}_1^2 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (Z_{ij}-\bar{Z})^2/(n_1+n_2)$, and $\hat{\sigma}^2 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} {Z_{ij}}^2/(n_1+n_2)$.

For Peer Review

S2: R Codes applied to the Monte Carlo Simulations.

```
###################################
############## Test 1 #############
###################################

#The next codes present computations of the test-1-statistic (9)
# number of the Monte Carlo iterations
k<-50000

# sample sizes
n1<-10
n2<-10

# delta value used in the definitions (7) and (8)
delta<-0.1

# Storage for the values of the proposed test statistic
EntrF<-array()


for(i in 1:k)
{
#generation of Sample 1 (paired data z1)
z1<-rnorm(n1,0,1)
#generation of Sample 2 (paired data z2)
z2<-rnorm(n2,0,1)
z<-c(z1,z2)      # combination of the samples, called z

 sz1<-sort(z1)    # sorting of the generated paired data z1
 sz2<-sort(z2)    # sorting of the generated paired data z2
 sz<-sort(z)      # sorting of the combined sample z

 # density-based test statistic (log of eq. (7)) based on Sample 1 (z1)
```

```
 LogM<-array()

# loop for each value of m
for(m in round(n1^(delta+0.5)):min(c(round((n1)^(1-delta)),round(n1/2))))
{
 Log<-0
   for(j in 1:n1)
   {
    D<-1*(j-m<1)+(j-m)*(j-m>=1)
    U<-n1*(j+m>n1)+(j+m)*(j+m<=n1)
    Uz<-length(sz[sz<=sz1[U]])+length(sz[-sz<=sz1[U]])
    Dz<-length(sz[sz<=sz1[D]])+length(sz[-sz<=sz1[D]])
    Delt<-(Uz-Dz)/(2*(n1+n2))
#Similarly to Canner (J.Am.Stat.Assoc. 70, 209-211(1975)), we will
#arbitrarily define Delt=1/(n1+n2), if Uz=Dz
    if (Delt==0) Delt<-1/(n1+n2)
    Log<-Log-2*log(n1)-log(Delt)+log(m)+log(2*n1-m-1)
   }
 LogM[m-round(n1^(delta+0.5))+1]<-Log
 }

 EntrF[i]<-min(LogM)

# density-based test statistic (log of eq. (8)) based on Sample 2 (z2)

 LogM<-array()
for(m in round(n2^(delta+0.5)):min(c(round((n2)^(1-delta)),round(n2/2))))
# loop for each value of m
 {
  Log<-0

  for(j in 1:n2)
  {
   D<-1*(j-m<1)+(j-m)*(j-m>=1)
   U<-n2*(j+m>n2)+(j+m)*(j+m<=n2)
   Uz<-length(sz[sz<=sz2[U]])+length(sz[-sz<=sz2[U]])
   Dz<-length(sz[sz<=sz2[D]])+length(sz[-sz<=sz2[D]])
   Delt<-(Uz-Dz)/(2*(n1+n2))
   if (Delt==0) Delt<-1/(n1+n2)
   Log<-Log-2*log(n2)-log(Delt)+log(m)+log(2*n2-m-1)
  }
 LogM[m-round(n2^(delta+0.5))+1]<-Log
 }
 EntrF[i]<-min(LogM)+EntrF[i]   # the proposed test statistic
}#end of the MC repetitions

### Calculate critical values of the density-based EL test statistic (9)
### for Test 1 #####

quantile(EntrF, 0.99) # alpha=0.01
quantile(EntrF, 0.95) # alpha=0.05
quantile(EntrF, 0.9)  # alpha=0.1
```

For Peer Review

```
####################################
############## Test 2 #############
####################################


#The next codes present computations of the test-2-statistic
#based on arrays sz, sz1, sz2 mentioned above.
#Density-based test statistic based on Sample 1 (log of eq. (10))

LogM<-array()
for(m in round(n1^(delta+0.5)):min(c(round((n1)^(1-delta)),round(n1/2))))
#loop for each value of m
 {
  Log<-0
  for(j in 1:n1)
  {
   D<-1*(j-m<1)+(j-m)*(j-m>=1)
   U<-n1*(j+m>n1)+(j+m)*(j+m<=n1)
   Uz<-length(sz[sz<=sz1[U]])+length(sz[-sz<=sz1[U]])
   Dz<-length(sz[sz<=sz1[D]])+length(sz[-sz<=sz1[D]])
   Delt<-(Uz-Dz)/(2*(n1+n2))
   if (Delt==0) Delt<-1/(n1+n2)
   Log<-Log-2*log(n1)-log(Delt)+log(m)+log(2*n1-m-1)
  }
  LogM[m-round(n1^(delta+0.5))+1]<-Log
 }
 EntrF[i]<-min(LogM) #here "i" is Monte Carlo index

#density-based test statistic based on Sample 2
 LogM<-array()
for(m in round(n2^(delta+0.5)):min(c(round((n2)^(1-delta)),round(n2/2))))
# loop for each value of m
 {
  Log<-0
  for(j in 1:n2)
  {
    D<-1*(j-m<1)+(j-m)*(j-m>=1)
    U<-n2*(j+m>n2)+(j+m)*(j+m<=n2)
    Uz<-length(sz[sz<=sz2[U]])+length(sz[-sz<=sz2[U]])
    Dz<-length(sz[sz<=sz2[D]])+length(sz[-sz<=sz2[D]])
    Delt<-(Uz-Dz)/(2*(n1+n2))
    if (Delt==0) Delt<-1/(n1+n2)
    d<-c()
    dd<-c()
    tuz<-length(sz2[-sz2<=sz2[n2]])-length(sz2[-sz2<=sz2[1]])
    for (r in 1:m-1)
    {
d[r]<-length(sz2[-sz2<=sz2[n2-r+1]])-length(sz2[-sz2<=sz2[n2-
r]])+length(sz2[-sz2<=sz2[r+1]])-length(sz2[-sz2<=sz2[r]])
     dd[r]<-((m-r)/(2*m))*d[r]
    }
    tud<-sum(dd)
    a<-(tuz-tud+n2-1-(m-1)/2)/(2*n2)
    Log<-Log-log(n2)-log(Delt)+log(2)+log(m)+log(a)
```

For Peer Review

```
    }
    LogM[m-round(n2^(delta+0.5))+1]<-Log
  }
#Test statistic at (15)
EntrF[i]<-min(LogM)+EntrF[i] #here "i" is Monte Carlo index
```

```
###################################
############## Test 3 #############
###################################
```

```
#The next codes present computations of the test-3-statistic(log of (16))

 N<-n1+n2         # total sample size
LogM<-array()
for(m in round(N^(delta+0.5)):min(c(round((N)^(1-delta)),round(N/2))))
{
  Log<-0
  for(j in 1:N)
  {
   D<-1*(j-m<1)+(j-m)*(j-m>=1)
   U<-N*(j+m>N)+(j+m)*(j+m<=N)
   Uz<-length(sz[sz<=sz[U]])+length(sz[-sz<=sz[U]])
   Dz<-length(sz[sz<=sz[D]])+length(sz[-sz<=sz[D]])
   Delt<-(Uz-Dz)/(2*(n1+n2))
   if (Delt==0) Delt<-1/(n1+n2)
   Log<-Log-2*log(N)-log(Delt)+log(m)+log(2*N-m-1)
  }
 LogM[m-round(N^(delta+0.5))+1]<-Log
 }
EntrF[i]<-min(LogM) #here "i" is Monte Carlo index
}
```

For Peer Review