



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

THE STATA JOURNAL

Editors

H. JOSEPH NEWTON
Department of Statistics
Texas A&M University
College Station, Texas
editors@stata-journal.com

NICHOLAS J. COX
Department of Geography
Durham University
Durham, UK
editors@stata-journal.com

Associate Editors

CHRISTOPHER F. BAUM, Boston College
NATHANIEL BECK, New York University
RINO BELLOCCO, Karolinska Institutet, Sweden, and
University of Milano-Bicocca, Italy
MAARTEN L. BUIS, WZB, Germany
A. COLIN CAMERON, University of California–Davis
MARIO A. CLEVES, University of Arkansas for
Medical Sciences
WILLIAM D. DUPONT, Vanderbilt University
PHILIP ENDER, University of California–Los Angeles
DAVID EPSTEIN, Columbia University
ALLAN GREGORY, Queen's University
JAMES HARDIN, University of South Carolina
BEN JANN, University of Bern, Switzerland
STEPHEN JENKINS, London School of Economics and
Political Science
ULRICH KOHLER, University of Potsdam, Germany

FRAUKE KREUTER, Univ. of Maryland–College Park
PETER A. LACHENBRUCH, Oregon State University
JENS LAURITSEN, Odense University Hospital
STANLEY LEMESHOW, Ohio State University
J. SCOTT LONG, Indiana University
ROGER NEWSON, Imperial College, London
AUSTIN NICHOLS, Urban Institute, Washington DC
MARCELLO PAGANO, Harvard School of Public Health
SOPHIA RABE-HESKETH, Univ. of California–Berkeley
J. PATRICK ROYSTON, MRC Clinical Trials Unit,
London
PHILIP RYAN, University of Adelaide
MARK E. SCHAFER, Heriot-Watt Univ., Edinburgh
JEROEN WEESIE, Utrecht University
IAN WHITE, MRC Biostatistics Unit, Cambridge
NICHOLAS J. G. WINTER, University of Virginia
JEFFREY WOOLDRIDGE, Michigan State University

Stata Press Editorial Manager

LISA GILMORE

Stata Press Copy Editors

DAVID CULWELL, DEIRDRE SKAGGS, and SHELBI SEINER

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

<http://www.stata-journal.com>

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

<http://www.stata.com/bookstore/sj.html>

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		Printed & electronic	
1-year subscription	\$ 98	1-year subscription	\$138
2-year subscription	\$165	2-year subscription	\$245
3-year subscription	\$225	3-year subscription	\$345
1-year student subscription	\$ 75	1-year student subscription	\$ 99
1-year institutional subscription	\$245	1-year institutional subscription	\$285
2-year institutional subscription	\$445	2-year institutional subscription	\$525
3-year institutional subscription	\$645	3-year institutional subscription	\$765
Electronic only		Electronic only	
1-year subscription	\$ 75	1-year subscription	\$ 75
2-year subscription	\$125	2-year subscription	\$125
3-year subscription	\$165	3-year subscription	\$165
1-year student subscription	\$ 45	1-year student subscription	\$ 45

Back issues of the *Stata Journal* may be ordered online at

<http://www.stata.com/bookstore/sjj.html>

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

<http://www.stata-journal.com/archives.html>

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.



Copyright © 2014 by StataCorp LP

Copyright Statement: The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, **stata**, Stata Press, Mata, **mata**, and NetCourse are registered trademarks of StataCorp LP.

Density-based empirical likelihood procedures for testing symmetry of data distributions and K-sample comparisons

Albert Vexler	Hovig Tanajian
Department of Biostatistics	Department of Biostatistics
New York State University at Buffalo	New York State University at Buffalo
Buffalo, NY	Buffalo, NY
avexler@buffalo.edu	hovigtan@buffalo.edu
Alan D. Hutson	
Department of Biostatistics	
New York State University at Buffalo	
Buffalo, NY	
ahutson@buffalo.edu	

Abstract. In practice, parametric likelihood-ratio techniques are powerful statistical tools. In this article, we propose and examine novel and simple distribution-free test statistics that efficiently approximate parametric likelihood ratios to analyze and compare distributions of K groups of observations. Using the density-based empirical likelihood methodology, we develop a Stata package that applies to a test for symmetry of data distributions and compares K -sample distributions. Recognizing that recent statistical software packages do not sufficiently address K -sample nonparametric comparisons of data distributions, we propose a new Stata command, `vxdbel`, to execute exact density-based empirical likelihood-ratio tests using K samples. To calculate p -values of the proposed tests, we use the following methods: 1) a classical technique based on Monte Carlo p -value evaluations; 2) an interpolation technique based on tabulated critical values; and 3) a new hybrid technique that combines methods 1 and 2. The third, cutting-edge method is shown to be very efficient in the context of exact-test p -value computations. This Bayesian-type method considers tabulated critical values as prior information and Monte Carlo generations of test statistic values as data used to depict the likelihood function. In this case, a nonparametric Bayesian method is proposed to compute critical values of exact tests.

Keywords: `st0338`, `vxdbel`, empirical likelihood, likelihood ratio, nonparametric tests, exact tests, K -sample comparisons, symmetry, p -value computation

1 Introduction

The classical statistical literature proposes parametric likelihood-ratio (LR) decision-making procedures as techniques for testing simple and composite hypotheses. In a variety of scenarios, parametric LR tests are often the most powerful statistical tools (see, for example, Lehmann and Romano [2005] and Vexler, Wu, and Yu [2010]).

The likelihood principle is arguably the most important concept for inference under parametric model assumptions. As an example, consider the goodness-of-fit testing problem, where given a sample of n independent and identically distributed (i.i.d.) observations X_1, \dots, X_n , we are interested in testing the hypothesis

$$H_0: X_1, \dots, X_n \sim F_0 \quad \text{versus} \quad H_1: X_1, \dots, X_n \sim F_1$$

where F_0 and F_1 are hypothesized distributions with the density functions $f_0(x)$ and $f_1(x)$, respectively. By virtue of the Neyman–Pearson lemma, the most powerful test statistic for this hypothesis is the LR $\prod_{i=1}^n f_1(X_i)/f_0(X_i)$, where the density functions f_0 and f_1 are assumed to be completely known.

An attractive property of the parametric LR methodology is that the likelihood principle provides researchers with clear algorithms for constructing efficient statistical tests across various complicated problems as they relate to practical applications. For example, we can easily extend the solution of the goodness-of-fit problem to a two-sample context with i.i.d. observations X_1, \dots, X_{n_1} and i.i.d. observations Y_1, \dots, Y_{n_2} relative to testing

$$H_0: F_X = F_Y \quad \text{versus} \quad H_1: F_X = F_1, F_Y = F_2, F_1 \neq F_2$$

where F_X, F_Y and F_1, F_2 are distribution functions of observed data under the null and alternative hypotheses, respectively. The resulting most powerful test statistic has the form

$$\prod_{i=1}^{n_1} \frac{f_1(X_i)}{f_X(X_i)} \prod_{j=1}^{n_2} \frac{f_2(Y_j)}{f_X(Y_j)}$$

where f_X, f_1 , and f_2 are the corresponding density functions that are assumed to be known.

Our research focuses on modifying the traditional parametric LR testing to the nonparametric setting to gain a degree of robustness while not sacrificing substantial power. In general, the concept of the parametric LR testing method may not be applicable in the nonparametric setting or when the number of unknown parameters involved in hypothesis testing is relatively large (see Fan, Farmen, and Gijbels [1998]). Also, when key assumptions are not met, parametric approaches may be nonrobust to the underlying assumptions, suboptimal, or extremely biased. An important goal in our research is to preserve the efficiency of statistical testing while maintaining robustness via robust distribution-free likelihood-type methods. As part of our approach, we extend and develop new methods based on empirical likelihood (EL) methods (see Owen [2001] and Vexler, Wu, and Yu [2010]).

The EL methodology provides an efficient nonparametric analog to the most powerful parametric likelihood methods. An outline of the EL approach is as follows: For i.i.d. observations X_1, \dots, X_n , the EL function has the form $L_p = \prod_{i=1}^n p_i$, where the components $p_i \in (0, 1)$, $i = 1, \dots, n$, maximize the likelihood L_p . The maximization process is conditional on a set of empirical constraints defined under the null hypothesis.

For example, define the null hypothesis of interest as $H_0 : EX_1 = 0$. It follows that the constraints of interest are $\sum_{i=1}^n p_i = 1$ and $\sum_{i=1}^n p_i X_i = 0$, given by the empirical counterpart of $H_0 : EX_1 = 0$. Computation of the p_i 's is based on a straightforward implementation of the method of Lagrange multipliers. This approach is due to consideration of the distribution function-based likelihood $\prod_{i=1}^n \{F(X_i) - F(X_{i-})\}$ over all distribution functions F (Owen 2001).

Whereas the Neyman–Pearson principle uses a density-based structure of the LR, the classical EL approach employs a distribution-based likelihood. Toward this end, we develop exact tests by modifying the classical EL approach, that is, by basing it directly on density functions (Vexler and Yu 2010). This method provides nonparametric approximations to Neyman–Pearson-type tests.

Vexler and Gurevich (2010) proposed to modify the main idea of the EL technique to develop density-based empirical approximations to the likelihood, having the form

$$L_f = \prod_{i=1}^n f(X_i) = \prod_{i=1}^n f(X_{(i)}) = \prod_{i=1}^n f_i$$

where $f_i = f(X_{(i)})$ and $X_{(i)} \leq \dots \leq X_{(n)}$ are the order statistics derived from the sample X_1, \dots, X_n . In this case, following the maximum EL technique, we can obtain estimated values f_i , $i = 1, \dots, n$, that maximize L_f , satisfying an empirical constraint that corresponds to $\int f(u)du = 1$. Vexler and Gurevich (2010), Vexler and Yu (2010), and Vexler, Gurevich, and Hutson (2013) employed this approach to create powerful density-based empirical likelihood (DBEL) ratio tests for two-sample comparisons and symmetry. Miecznikowski, Vexler, and Shepherd (2013) developed an R package for DBEL ratio goodness-of-fit tests. Tsai, Vexler, and Gurevich (2013) showed that DBEL tests are robust and significantly outperform classical procedures, including the Kolmogorov–Smirnov test, Wilcoxon rank-sum test, and t test.

In this article, we extend these results to K -sample comparisons, including considerations for ordered alternatives (Gurevich 2012). Although investigators need to compare K samples in nonparametric settings in various clinical trials, to our knowledge, software procedures based on the Kruskal–Wallis (KW) test are techniques only to execute nonparametric tests for a three-sample comparison.

The proposed tests are exact; that is, their null distributions are defined independently of data distributions, and their critical values can be evaluated without using asymptotic approximations.

In terms of the inferential procedure, two methods are commonly used for calculating the p -values in statistical software products. The Monte Carlo (MC) method is a well-known technique for accurately approximating the critical values and p -values of exact tests. For some testing situations, the use of the MC technique can be computationally intensive. For example, a relatively large number of MC repetitions—say, t —are needed to evaluate critical values that correspond to the 1% significance level because in this case, the common 95% confidence interval of such an evaluation can be calculated as $[0.01 \pm 1.96\sqrt{(0.01)(1 - 0.01)/t}]$. The use of tables with corresponding critical values is

also a standard method applied in various statistical software routines. In the literature, this method is commonly referred to as interpolation. Providing the tables for use within the testing algorithm improves the execution speed of the test. However, interpolation becomes less reliable when real-data characteristics (for example, sample sizes) differ from those used to tabulate the critical values (Pearson and Hartley 1966).

To advance these methods, we propose a new hybrid method that combines both the interpolation and the MC methods. The result of this hybrid method is an innovation that is related to applications of exact tests in statistical software and that can be applied in the broader setting.

In developing the Stata package, we used the Mata matrix programming language. Using the proposed package, we provide MC comparisons between classical procedures and our newly proposed command. The results are shown in section 5. Section 6 presents applications of the proposed three-sample test to a real-data example. Section 7 provides some concluding remarks.

2 Method

In this section, we develop the DBEL ratio tests for symmetry of data distributions and for comparing K -sample distributions.

2.1 Tests for symmetry

Consider the problem of testing the symmetry of a one-sample distribution about zero. We suppose that the data consist of a sample of i.i.d. observations X_1, \dots, X_n . Our hypothesis of interest is $H_0 : F_x(u) = 1 - F_x(-u)$ for all $-\infty \leq u \leq \infty$ versus $H_1 : F_x(u) \neq 1 - F_x(-u)$ for some $-\infty \leq u \leq \infty$, where the distribution F_x of the observations is assumed to be unknown. In this case, the LR has the form

$$\begin{aligned} \text{LR} &= \frac{\text{likelihood under } H_1}{\text{likelihood under } H_0} = \frac{\prod_{i=1}^n f_{H_1}(X_i)}{\prod_{i=1}^n f_{H_0}(X_i)} = \frac{\prod_{j=1}^n f_{H_1}(X_{(j)})}{\prod_{j=1}^n f_{H_0}(X_{(j)})} \\ &= \frac{\prod_{j=1}^n f_{H_1, j}}{\prod_{j=1}^n f_{H_0, j}} \end{aligned}$$

where $f_{H_k}(X_{(j)}) = f_{H_k, j}$ with $k = 0, 1$ and where $X_{(j)}$ are the order statistics based on the observations X_1, \dots, X_n .

This LR is the most powerful test statistic in the parametric setting. However, in the nonparametric setting, the LR statistic is not computable because the density functions

are unknown. To generate the nonparametric test for the hypothesis above, we consider the likelihood $L_f = \prod_{j=1}^n f_{H_1,j}$. We estimate the values of $f_{H_1,j}$ that maximize L_f given an empirical version of the constraint $\int f_{H_1} = 1$. Following Vexler and Gurevich (2010), this constraint has the form

$$\frac{1}{2m} \sum_{j=1}^n \int_{X_{(j-m)}}^{X_{(j+m)}} \frac{f_{H_1}(u)}{f_{H_0}(u)} f_{H_0}(u) du \leq 1$$

The empirical constraint on the $f_{H_1,j}$'s can then be reformulated as

$$\frac{1}{2m} \sum_{j=1}^n \frac{f_{H_1,j}}{f_{H_0,j}} \Delta_{jm} = 1 - \frac{m+1}{2n}$$

$$\Delta_{jm} = \frac{1}{2n} \sum_{i=1}^n \{I(X_{(j-m)} \leq X_i \leq X_{(j+m)}) + I(X_{(j-m)} \leq -X_i \leq X_{(j+m)})\}$$

where $I(\cdot)$ is an indicator function, $X_{(i)} = X_{(1)}$ if $i \leq 1$, and $X_{(i)} = X_{(n)}$ if $i \geq n$. The method of Lagrange multipliers is then used to find values of $f_{H_1,j}$:

$$f_{H_1,j} = f_{H_0,j} \frac{2m\{1 - (m+1)(2n)^{-1}\}}{n\Delta_{jm}}, \quad j = 1, \dots, n$$

Hence, the EL approximation to L_f and LR can be presented as

$$\prod_{j=1}^n f_{H_0,j} \frac{2m\{1 - (m+1)(2n)^{-1}\}}{n\Delta_{jm}} \quad \text{and} \quad V_{nm} = \prod_{j=1}^n \frac{2m\{1 - (m+1)(2n)^{-1}\}}{n\Delta_{jm}}$$

respectively. Thus it follows that the maximum EL method forms the LR test statistic

$$V_n = \min_{a_n \leq m \leq b_n} V_{nm}$$

where $a_n = n^{0.5+\delta}$, $b_n = \min(n^{1-\delta}, n/2)$, and $\delta \in (0, 0.25)$ (for details, see Vexler, Gurevich, and Hutson [2013]). For practical purposes, we suggest a value of $\delta = 0.1$ for our applications. Tsai, Vexler, and Gurevich (2013) and Vexler, Gurevich, and Hutson (2013) showed that the power of the test statistic does not differ substantially for values of $\delta \in (0, 0.25)$. Similar to the rationale provided in Vexler and Gurevich (2011), we will set $\Delta_{jm} = 1/n$ if $\Delta_{jm} = 0$ in terms of practical applications.

The proposed test is now designed to reject the null hypothesis if $\log(V_n) > C$, where C is the critical value of the test. As Vexler, Gurevich, and Hutson (2013) showed, a test based on the statistic $\log(V_n)$ has an asymptotic power 1; that is, it is a consistent test. It turns out that the null distribution of the test statistic V_n is independent of the distribution of observations X_1, \dots, X_n . Our test statistic is based on indicator functions involved in the definition of Δ_{jm} . Because $I(X > Y) = I\{F_X(X) > F_X(Y)\}$, it follows that

$$P_{H_0}\{\log(V_n) > C\} = P_{X_1, \dots, X_n \sim \text{norm}(0,1)}\{\log(V_n) > C\}$$

By virtue of this result, the proposed test is exact, and the corresponding critical values can be tabulated for fixed sample sizes. In the interpolation method, presented in section 3, the MC approach is used to obtain the related critical values that are stored in Stata beforehand to increase the execution speed of the test.

Next we consider the one-sided version of the two-sided problem (Vexler, Gurevich, and Hutson 2013). We want to verify that the sample is generated from a distribution that is stochastically greater than zero. That is, we want to test for $H_0: F_x(u) = 1 - F_x(-u)$ for all $-\infty \leq u \leq \infty$ versus $H_1: F_x(u) \leq 1 - F_x(-u)$ for some $-\infty \leq u \leq \infty$, where the distribution F_x is assumed to be unknown.

Applying the DBEL concept, we have the EL ratio test statistic

$$V_n^* = \min_{a_n \leq m \leq b_n} \prod_{j=1}^n \frac{2m\{1 - (m+1)(2n)^{-1}\}}{n\Delta_{jm}}, \text{ with}$$

$$\Delta_{jm} = \max \frac{1}{n} \left\{ \sum_{i=1}^n I(X_i \leq X_{(j+m)}), \sum_{i=1}^n I(-X_i \leq X_{(j+m)}) \right\}$$

$$- \max \frac{1}{n} \left\{ \sum_{i=1}^n I(X_i \leq X_{(j-m)}), \sum_{i=1}^n I(-X_i \leq X_{(j-m)}) \right\}$$

The proposed test rejects the null hypothesis if $\log(V_n^*) > C$, where C is the test threshold. As in the two-sided test, the critical values of the one-sided test can be calculated according to the following equation: $P_{H_0}\{\log(V_n^*) > C\} = P_{X_1, \dots, X_n \sim \text{norm}(0,1)}\{\log(V_n^*) > C\}$.

2.2 Two-sample comparison

The two-sample DBEL ratio test has been dealt with extensively in the recent literature (for example, Vexler and Gurevich 2011; Vexler and Yu 2010; Vexler et al. 2012; Miecznikowski, Vexler, and Shepherd 2013). To outline this method, we suppose that data consist of two samples of i.i.d. observations: X_1, \dots, X_n and Y_1, \dots, Y_k . We want to verify that both samples are from the same distribution. That is, we want to test for $H_0: F_x = F_y = F_z$ versus $H_1: F_x \neq F_y$, where the distributions F_x , F_y , and F_z of the observations are unknown. In this case, the most powerful LR statistic has the form of $\text{LR} = \prod_{i=1}^n f_{X,i}/f_{ZX,i} \prod_{j=1}^k f_{Y,j}/f_{ZY,j}$, where $f_X(X_{(i)}) = f_{X,i}$, $f_Z(X_{(i)}) = f_{ZX,i}$, $f_Y(Y_{(j)}) = f_{Y,j}$, and $X_{(i)}$ and $Y_{(j)}$ are the order statistics based on the observations X_1, \dots, X_n and Y_1, \dots, Y_k , respectively.

To construct the corresponding nonparametric test, we consider the likelihood $L_f = \prod_{i=1}^n f_{X,i}$. Toward this end, we find values of $f_{X,i}$ that maximize L_f given an empirical version of the constraint $\int f_X(u)du = 1$. This is quite similar to the one-sample setting described earlier. Define the following H_0 empirical distribution function (EDF):

$$F_{Z(n+k)}(u) = \frac{\sum_{i=1}^n I(X_i \leq u) + \sum_{j=1}^k I(Y_j \leq u)}{(n+k)}$$

After Vexler and Gurevich (2011), one can show that by virtue of the hypotheses setting, $H_0: F_x = F_y = F_z$ versus $H_1: F_x \neq F_y$, the empirical constraint $\tilde{\Delta}_m \leq 1$, with

$$\tilde{\Delta}_m = \frac{\sum_{i=1}^n \{F_{Z(n+k)}(X_{(i+m)}) - F_{Z(n+k)}(X_{(i-m)})\} \frac{f_{X,i}}{f_{ZX,i}}}{2m}$$

approximates the condition

$$\Delta_m = \frac{\prod_{i=1}^n \int_{X(i-m)}^{X(i+m)} \frac{f_Z(u) f_X(u)}{f_Z(u)} du}{2m} \leq 1$$

which reflects the property $\int f_Z(u) f_X(u) / f_Z(u) du = \int f_X(u) du = 1$. Then the Lagrange multiplier method can be used to directly find values of $f_{X,i}$ that maximize L_f , satisfying the empirical constraint $\tilde{\Delta}_m \leq 1$:

$$f_{X,i} = \frac{2m f_{ZX,i}}{n \{F_{Z(n+k)}(X_{(i+m)}) - F_{Z(n+k)}(X_{(i-m)})\}}, \quad i = 1, \dots, n$$

This implies that the EL estimator of $\prod_{i=1}^n f_{X,i} / f_{ZX,i}$ is given as

$$\text{ELR}_{X,m,n} = \prod_{i=1}^n \frac{2m}{n \{F_{Z(n+k)}(X_{(i+m)}) - F_{Z(n+k)}(X_{(i-m)})\}}$$

Similarly, the maximum EL method forms the EL estimator of $\prod_{j=1}^k f_{Y,j} / f_{ZY,j}$ as

$$\text{ELR}_{Y,r,k} = \prod_{i=1}^k \frac{2r}{k \{F_{Z(n+k)}(Y_{(i+r)}) - F_{Z(n+k)}(Y_{(i-r)})\}}$$

where $Y_{(i)} = Y_{(1)}$ if $i \leq 1$ and $Y_{(i)} = Y_{(k)}$ if $i \geq k$.

Thus the proposed test statistic is

$$V_{nk} = \text{ELR}_{X,n} \text{ELR}_{Y,k}$$

where $\text{ELR}_{X,n} = \min_{a_n \leq m \leq b_n} \text{ELR}_{X,m,n}$, $\text{ELR}_{Y,k} = \min_{a_k \leq r \leq b_k} \text{ELR}_{Y,r,k}$, $a_l = l^{0.5+\delta}$, $b_l = \min(l^{1-\delta}, l/2)$, $\delta \in (0, 1/4)$, and $l = n, k$. Here the operation min is used to provide the EL estimators of the parameters m and r in the expressions $\text{ELR}_{X,m,n}$ and $\text{ELR}_{Y,r,k}$, respectively. The bounds a_l and b_l are defined to preserve the asymptotic consistency of the test (see Vexler and Yu [2010] for details). As before, we set a value of $\delta = 0.1$. It is shown in Tsai, Vexler, and Gurevich (2013), Vexler and Yu (2010), Vexler and Gurevich (2010), and Vexler, Gurevich, and Hutson (2013) that the power of the test statistic does not differ substantially for values of $\delta \in (0, 1/4)$. Similar to what was defined in Vexler and Gurevich (2011), $F_{Z(n+k)}(x) - F_{Z(n+k)}(y) = 1/(n+k)$ if $F_{Z(n+k)}(x) = F_{Z(n+k)}(y)$.

The proposed test rejects the null hypothesis if $\log(V_{nk}) > C$, where C is the test threshold. As we saw in section 2.1, it turns out that the null distribution of

the test statistic V_{nk} is independent of the sample distributions. Therefore, we have $P_{H_0}\{\log(V_{nk}) > C\} = P_{X_1, \dots, X_n, Y_1, \dots, Y_k \sim \text{unif}(0,1)}\{\log(V_{nk}) > C\}$. Thus this expression indicates that the probability of a type I error of the test can be calculated exactly.

Now consider the one-sided version of the two-sample problem (Gurevich 2012). We want to test the hypotheses $H_0 : F_x = F_y = F_z$ versus $H_1 : F_x(u) \leq F_y(u)$ for all $-\infty \leq u \leq \infty$, $F_x(u) < F_y(u)$ for some $-\infty \leq u \leq \infty$.

Applying the maximum DBEL concept, one can define the test statistic

$$V_{nk}^* = \text{ELR}_{X,n}^* \text{ELR}_{Y,k}^*$$

where

$$\begin{aligned} \text{ELR}_{X,n}^* &= \min_{a_n \leq m \leq b_n} \prod_{i=1}^n \frac{2m}{n\{F_Z^*(X_{(i+m)}) - F_Z^*(X_{(i-m)})\}} \\ \text{ELR}_{Y,k}^* &= \min_{a_k \leq r \leq b_k} \prod_{i=1}^k \frac{2r}{k\{F_Z^{**}(Y_{(i+r)}) - F_Z^{**}(Y_{(i-r)})\}} \\ F_Z^*(u) &= \max\{F_{X,n}(u), F_{Y,k}(u)\} \\ F_Z^{**}(u) &= \min\{F_{X,n}(u), F_{Y,k}(u)\} \end{aligned}$$

with $F_{X,n}(u) = n^{-1} \sum_{i=1}^n I(X_i \leq u)$ and $F_{Y,k}(u) = k^{-1} \sum_{j=1}^k I(Y_j \leq u)$. Further, if $F_Z^*(X_{(i+m)}) - F_Z^*(X_{(i-m)}) < 0$, then it is set to $(n+k)^{-1}$. Similarly, if $F_Z^{**}(Y_{(i+r)}) - F_Z^{**}(Y_{(i-r)}) \leq 0$ then it is set to $(n+k)^{-1}$. Here, in providing comparisons with the test statistic V_{nk} , instead of using $F_{Z(n+k)}$ —the EDF of identically H_0 -distributed X and Y —we use F_Z^* and F_Z^{**} to depict the fit between data distributions and our alternative hypothesis.

As in the two-sided test, the critical values can be calculated exactly according to the following equation: $P_{H_0}\{\log(V_{nk}^*) > C\} = P_{X_1, \dots, X_n, Y_1, \dots, Y_k \sim \text{norm}(0,1)}\{\log(V_{nk}^*) > C\}$.

2.3 K-sample comparison

To outline the K -sample procedure, we suppose that the data consist of K independent samples. We want to test whether all K samples are distributed identically in a nonparametric fashion. Let n_1, \dots, n_k denote the respective sample sizes corresponding to the K samples being compared. Assume that the K samples are represented by the vectors of observations given as $\{X_{11}, \dots, X_{1n_1}\}, \dots, \{X_{k1}, \dots, X_{kn_k}\}$ from the corresponding distribution functions F_{X_1}, \dots, F_{X_k} . We now want to test the hypothesis $H_0 : F_{X_1} = \dots = F_{X_k} = F_Z$ versus $H_1 : \text{not all } F_{X_i} = F_{X_j}, i \neq j$. If the corresponding density functions are known, the LR statistic has the form $\text{LR} = \prod_{j=1}^k \prod_{i=1}^{n_j} f_{X_j}(X_{ji}) / f_Z(X_{ji}) = \prod_{j=1}^k \prod_{i=1}^{n_j} f_{X_j,i} / f_{Z,X_j,i}$, where f_{X_j} denotes the density function of the j th sample under H_1 , f_Z is the theoretical density function of observations under H_0 , and $X_{j(i)}$, $j = 1, \dots, k$, are the order statistics per sample based on the observations X_{j1}, \dots, X_{jn_j} . As before, we denote $f_{X_j}(X_{j(i)}) = f_{X_j,i}$ and $f_Z(X_{j(i)}) = f_{Z,X_j,i}$, $j = 1, \dots, k$.

We apply the maximum EL method to obtain the proposed DBEL ratio test statistic $V_{n_1, n_2, \dots, n_k} = \prod_{j=1}^k \text{ELR}_{X_j, n_j}$, where the EL estimator of $\prod_{i=1}^{n_j} f_{X_j, i} / f_{Z(N), i}$, for $j = 1, \dots, k$, is $\text{ELR}_{X_j, n_j} = \min_{a_{nj} \leq m_j \leq b_{nj}} \prod_{i=1}^{n_j} 2m_j / n_j \{F_{Z(N)}(X_{j(i+m_j)}) - F_{Z(N)}(X_{j(i-m_j)})\}$, $a_l = l^{0.5+\delta}$, $b_l = \min(l^{1-\delta}, l/2)$, $\delta \in (0, 1/4)$, and the corresponding EDF under H_0 is given as $F_{Z(N)}(u) = 1/N \sum_{j=1}^k \sum_{i=1}^{n_j} I(X_{ji} \leq u)$, $N = \sum_{j=1}^k n_j$, $X_{j(i)} = X_{j(1)}$ if $i \leq 1$ and $X_{j(i)} = X_{j(n_j)}$ if $i \geq n_j$. As in the two-group setting, we define $F_{Z(N)}(x) - F_{Z(N)}(y) = 1/N$ if $F_{Z(N)}(x) = F_{Z(N)}(y)$.

The type I error of the K -sample test can be monitored through the following probability statement: $P_{H_0} \{\log(V_{n_1, n_2, \dots, n_k}) > C\} = P_{X_{11}, \dots, X_{1n_1}, \dots, X_{k1}, \dots, X_{kn_k} \sim \text{unif}(0,1)} \{\log(V_{n_1, n_2, \dots, n_k}) > C\}$. As above, we set a value of $\delta = 0.1$. The proposed test statistic approximates nonparametrically the most powerful parametric LR test statistic; thus one can assume heuristically that the proposed test also has good relative efficiency.

3 Evaluations of critical values of the exact tests

In this section, we outline three methods to obtain critical values for the exact tests proposed in this article. The MC, interpolation, and hybrid methods will be described in section 3.1, section 3.2, and section 3.3, respectively. These techniques are applied within a newly developed Stata command.

3.1 MC method

The MC method is a well-known approach for obtaining accurate approximations to the critical values (CVs) of exact tests (see Metropolis and Ulam [1949] and Rubinstein and Kroese [2008] for details).

The proposed Stata command provides the option of the MC method using 10,000 MC repetitions to calculate a critical value per application of the command. In the developed procedures, the CVs are calculated using data generated from a standard normal distribution for one-sample tests and a uniform(0,1) distribution for the two-sample and three-sample tests. The generated values of V , the test statistic, are used to determine the critical value C , the test threshold, for the desired significance level $\alpha = P_{H_0}(\log V > C_\alpha)$ via calculating the quantiles of the MC H_0 distribution of $\log V$.

In addition, using the MC procedure, we tabulated tables of CVs. Each critical value was calculated for each proposed test, each set of sample sizes, and each significance level. The MC procedure was based on 50,000 replicate samples of the corresponding test statistic's $\log V$. The resulting tables of CVs are to be used by the command `vxdbel`, a product of this package.

3.2 An approach based on interpolation

Interpolation differs from the MC method in that tables of CVs were calculated beforehand for each proposed test, various sample sizes, and various significance levels α . Following the method in the previous section, we obtained the resulting tables and saved them in Stata data format. Pearson and Hartley (1966) demonstrated the method of interpolation for calculating values within tables. For example, suppose we have data consisting of one sample with size n —say, $n = 37$ —to be tested for symmetry. The needed critical value can be interpolated using tabulated CVs of $n = 30, 35, 40$, and 50 . For our procedure, the tables with CVs are provided for sample sizes: the one-sample tests with sample size n from the set $\{A\}$, where $A = \{1, 2, 3, \dots, 29, 30, 35, 40, 50, 60, 80, 100, 150, 200\}$; the two-sample tests with sample sizes (m, n) from the set $\{A \times A\}$; and the three-sample test with sample sizes (m, n, p) from the set $\{A \times A \times A\}$. The applications of the interpolation and extrapolation procedures decrease the accuracy of the estimates of the CVs when actual samples have sample sizes that differ significantly from those used to tabulate the tables. Providing tables for the test increases the speed to execute the test compared with the MC method, but the interpolation method becomes less reliable.

In the context of the proposed procedures, in the cases of sample sizes not tabulated within the tables, an appropriate subset of the table data is selected. The data to be selected have been defined as related to sample sizes in the tables within a radius of two values around the values of the sample sizes that are needed. For example, if we are interested in a one-sample test based on a sample with size $n = 78$, we use CVs related to $n = 50, 60, 80$, and 100 to estimate the required critical value. On the basis of extensive MC experiments, to interpolate for needed CVs, we conclude that the radius of two values around the CVs best minimizes the bias of the interpolation and extrapolation procedures. To outline this method for one sample, we define $C_{n\alpha}$ as the critical value corresponding to the sample size n and the significance level α . Using the selected table data, we fit $C_{n\alpha}$ via the regression model

$$C_{n\alpha} \cong \beta_0 + \beta_1 n + \beta_2 1/n + \beta_3 n^2 + \beta_4 \alpha + \beta_5 \alpha^{1/2}$$

employing the local maximum-likelihood methodology (Fan, Farman, and Gijbels 1998). In this manner, the coefficient β 's are estimated and yield values of $C_{n\alpha}$ as a function of n and α .

Similarly, for the two- and three-sample tests, we define $C_{nm\alpha}$ and $C_{nmk\alpha}$ as the CVs corresponding to the sample sizes n, m , and k and the significance level α . Using the selected table data, we fit $C_{nm\alpha}$ and $C_{nmk\alpha}$ via the regression models $C_{nm\alpha} \cong \beta_0 + \beta_1 n + \beta_2 m + \beta_3 1/n + \beta_4 1/m + \beta_5 n^2 + \beta_6 m^2 + \beta_7 \alpha + \beta_8 \alpha^{1/2}$ and $C_{nmk\alpha} \cong \beta_0 + \beta_1 n + \beta_2 m + \beta_3 k + \beta_4 1/n + \beta_5 1/m + \beta_6 1/k + \beta_7 n^2 + \beta_8 m^2 + \beta_9 k^2 + \beta_{10} \alpha + \beta_{11} \alpha^{1/2}$. Here we assume $C_{n\alpha} = G_1(n, \alpha)$, $C_{nm\alpha} = G_2(n, m, \alpha)$, and $C_{nmk\alpha} = G_3(n, m, k, \alpha)$, where the functions G_1 , G_2 , and G_3 are unknown but approximated via the equations shown above.

The regression equation with the parameter estimates is solved backward for an estimate of the type I error, α , using the equation solvers `mm_root` and `optimize` in

Stata. This is accomplished by plugging the value of the test statistic, $\log V$, which is based on the observed data for the critical value and the sample sizes of the observed data, into the regression equation and solving for α ; that is, $\log V = \hat{\beta}_0 + \hat{\beta}_1 n + \hat{\beta}_2 1/n + \hat{\beta}_3 n^2 + \hat{\beta}_4 \alpha + \hat{\beta}_5 \alpha^{1/2}$ should be solved with respect to α for the one-sample test. Thus the value obtained for α is an estimate of the p -value for the test.

3.3 A novel hybrid technique that combines MC and interpolation

This method combines the MC and interpolation methods depicted in section 3.1 and section 3.2, respectively. Toward this end, we propose a nonparametric Bayesian-type approach for constructing the posterior expectations of the needed CVs. Thereby, we incorporate the efficiency of the interpolation method and the accuracy of the MC method.

Lazar (2003) and Vexler, Deng, and Wilding (2013) showed that the ELs can be used in Bayesian statistical inferences instead of the corresponding parametric likelihoods. This provides nonparametric Bayes procedures. This concept is applied in the proposed command to calculate the CVs. Because distributions of test-statistic values are unknown, the likelihoods are presented in the EL form.

The proposed algorithm is conducted in two stages and repeated until a stopping condition is met. The following notations are used in the description of the procedures:

- t_k denotes the number of MC simulations related to stage $k = 1, 2$;
- α denotes the level of significance (we use $\alpha = 0.05$ as the default level in the command);
- $T_{(1)}^k < T_{(2)}^k < \dots < T_{(t_k)}^k$ denotes the order statistics based on the test statistic values $T_1^k, T_2^k, \dots, T_{t_k}^k$, generated on stage k ($T = \log V$);
- J_k denotes the interval $[\max\{1, (1 - \alpha)t_k - t_k^{1/2}\log(t_k)\}, \min\{(1 - \alpha)t_k + t_k^{1/2}\log(t_k), t_k\}]$;
- $L_k(q) = \exp(t_k F_{k,t_k}(q))[\log(1 - \alpha) - \log\{F_{k,t_k}(q)\} + t_k\{1 - F_{k,t_k}(q)\}][\log\alpha - \log\{1 - F_{k,t_k}(q)\}]$ denotes the EL function $= \max\{\prod_{i=1}^{t_k} p_i : \sum_{i=1}^{t_k} p_i = 1, \sum_{i=1}^{t_k} p_i I(T_i^k < q) = \alpha\}$, where $F_{k,t_k}(q) = \sum_{i=1}^{t_k} I(T_i^k < q)/t_k$; and
- the quantile q_α , such that $P_{H_0}(\log V > q_\alpha) = \alpha$, defines the needed critical value of the test statistic $\log V$.

We begin by fitting this information with a functional form. The tabulated values are assumed to provide prior information regarding the target CVs. We obtain the prior distribution function $\pi(q)$, with the parameters (μ_0, σ_0) , by using the local maximum-likelihood method (Fan, Farman, and Gijbels 1998) based on tabulated CVs. Here

$$\pi(q) = (2\pi\sigma_0)^{-0.5} \exp\{-(q - \mu_0)^2/(2\sigma_0^2)\}^{-1}$$

The normal function form of $\pi(q)$ was used because quantile estimators are commonly normally distributed when sample sizes are relatively large. For example, in the context of the two-sample test, following section 3.2, we can present $\mu_0 = \hat{\beta}_0 n_0 + \hat{\beta}_1 m_0 + \hat{\beta}_2 1/n_0 = \hat{\beta}_3 1/m_0 + \hat{\beta}_4 n_0^2 + \hat{\beta}_5 m_0^2 + \hat{\beta}_6 \alpha$, where we assume that the observed data consist of two samples of sizes n_0 and m_0 . In this case, the σ_0 can be estimated using standard regression analysis.

In the first MC simulation step, 200 generations ($t_1 = 200$) of test statistic values are conducted; then test statistics $T^1 = (T_1^1, \dots, T_{t_1}^1)$, based on the generated data, are calculated. Next the posterior expectation $\hat{q}_{1,\alpha}$, of the quantile q_α , is calculated. We compute the posterior expectation of quantiles as follows:

$$\hat{q}_{k,\alpha} = \frac{\int_{T_{(1)}^1}^{T_{(t_1)}^k} q L_k(q) \pi(q) dq}{\int_{T_{(1)}^k}^{T_{(t_1)}^k} L_k(q) \pi(q) dq}, \quad k = 1$$

Now one can show that

$$\hat{q}_{1,\alpha} = \frac{\sum_{j \in J_1} \exp[-t_1(1 - \alpha - \frac{j}{t_1})^2 \{2\alpha(1 - a)\}^{-1}] \int_{T_{(j-1)}^c}^{T_{(j)}^1} q \pi(q) dq}{\sum_{j \in J_1} \exp[-t_1(1 - \alpha - \frac{j}{t_1})^2 \{2\alpha(1 - a)\}^{-1}] \int_{T_{(j-1)}^1}^{T_{(j)}^1} \pi(q) dq}$$

where, in general, $\int_a^b q \pi(q) dq = \sqrt{\sigma_0^2/2\pi} [\exp\{-(a-\mu_0)^2/(2\sigma_0^2)\} - \exp\{-(b-\mu_0)^2/(2\sigma_0^2)\}] + \mu_0 \int_a^b \pi(q) dq$, and $\int_a^b \pi(q) dq = \text{normal}((b - \mu_0)/\sigma_0) - \text{normal}((a - \mu_0)/\sigma_0)$. In this equation, the Stata function `normal()` returns the cumulative standard normal distribution.

During the second stage, 200 additional generations ($t_2 = 200$) of the test statistic values are conducted; then test statistics $T^2 = (T_1^2, \dots, T_{t_2}^2)$, based on simulated data, are calculated. We then estimate $f_{2,t_2}(\hat{q}_{1,\alpha})$, an estimate of the density function of the test statistic, by using the following kernel estimator (see Gibbons and Chakraborti [2005] for details):

$$f_{2,t_2}(\hat{q}_{1,\alpha}) = (t_2)^{-1} \sum_{j=1}^{t_2} (2\pi h^2)^{-1/2} \exp\{-(\hat{q}_{1,\alpha} - T_j^2)^2 (2h^2)^{-1}\}$$

where $h = 1.06 \hat{\sigma}_{t_2} t^{-1/5}$ and $\hat{\sigma}_{t_2}$ is the standard deviation of the test statistics T_j^2 . Then we calculate the estimated variance of $\hat{q}_{1,\alpha}$ as $V_k \equiv F_{k+1,t_{k+1}}(\hat{q}_{k,\alpha}) \{1 - F_{k+1,t_{k+1}}(\hat{q}_{k,\alpha})\} f_{k+1,t_{k+1}}(\hat{q}_{k,\alpha})^{-2} / t_{k+1}$, $k = 1$. (For details of this approximation to the variance, see Serfling [1980].)

To define the stopping rule of the procedure, we evaluate whether $V_1 \leq \sigma_0^2$; if this is so, we stop the procedure and calculate the posterior expectation of the quantiles with the combined values of the test statistics $T^c = (T_1^1, \dots, T_{t_1}^1, T_1^2, \dots, T_{t_2}^2)$ and $t_c = t_1 + t_2$:

$$\hat{q}_{c,\alpha} = \frac{\int_{T_{(1)}^c}^{T_{(t_c)}^c} q L_c(q) \pi(q) dq}{\int_{T_{(1)}^c}^{T_{(t_c)}^c} L_c(q) \pi(q) dq} \cong \frac{\sum_{j \in J_1} \exp[-t_c(1 - \alpha - \frac{j}{t_c})^2 \{2\alpha(1 - a)\}] \int_{T_{(j-1)}^c}^{T_{(j)}^c} q \pi(q) dq}{\sum_{j \in J_1} \exp[-t_c(1 - \alpha - \frac{j}{t_c})^2 \{2\alpha(1 - a)\}] \int_{T_{(j-1)}^c}^{T_{(j)}^c} \pi(q) dq}$$

In this case, we reach the estimated value of the variance of the CVs comparable with the variance of the CVs found in the table. If the value of the test statistic based on the data, t_0 , is greater than $\hat{q}_{c,\alpha}$, then we reject the null.

If $V_1 > \sigma_0^2$, combine T^1 and T^2 into a new T^1 so that the new t_1 is equal to $t_1 + t_2$. Repeat stages one and two until the stop condition ($V_1 \leq \sigma_0^2$) is reached or the number of the new combined values of test statistics, T^1 , is greater than 35,000.

In summary, the algorithm is performed iteratively, and the decision to perform another iteration of the scheme is based on a comparison of the variance estimator V_k and the parameter σ_0^2 .

4 The vxdbel command

4.1 Description

In this article, we present a Stata implementation of the DBEL ratio test in the command `vxdbel`. This command conducts a one-sample test of symmetry, two-sample comparison, or three-sample comparison as described above. The command output presents an applicable test statistic and p -value.

4.2 Syntax

The syntax of the command to execute the test of symmetry and K -sample distribution-free DBEL ratio test is

```
vxdbel varlist [ , sided(description) method(type) level(#) reps(#) ]
```

where *varlist* specifies up to three variable names.

4.3 Options

`sided(description)` specifies the test as one sided or two sided. `sided(greater)` and `sided(less)` perform a one-sided test. `sided(two.sided)` (the default) conducts a two-sided test.

`method(type)` specifies the test method. `method(interpolation)` (the default) defines it as interpolation, `method(mc)` defines it as MC, and `method(hybrid)` defines it as hybrid.

`level(#)` specifies the significance level, as a percentage, for the hybrid method. The default is `level(95)`.

`reps(#)` specifies the number of repetitions to use in calculating the critical values for the MC method. The default is `reps(10000)`.

4.4 Remarks

Three Stata tables are provided with the command. For the MC method, the critical values are obtained by MC simulations for each test conducted as described in section 3.1. For the interpolation method, the provided tables of CVs are used to obtain the p -value as described in section 3.2. For the hybrid method, which is based on the Bayesian approach, the provided tables of CVs are used to obtain initial parameter estimates (μ_0, σ_0) as described in section 3.3.

Further, assuming the `method()` option is set to interpolation or hybrid, the provided tables of CVs are complete for sample sizes less than or equal to 30 but are incomplete for sample sizes greater than 30. If all the sample sizes are 30 or less, then the CVs are directly available in the tables. If any of the sample sizes is greater than 30, then the CVs are not directly available in the tables. In this case, the hybrid and interpolation methods are used to obtain the desired CVs.

5 Simulations

In this section, we examine the power of the proposed tests. In particular, we compare the power of our tests with that of the classical tests, and we compare the new hybrid method with the MC and interpolation methods. Simulations were conducted for 10,000 repetitions for all tests with varying sample sizes, at a significance level of 0.05. In the interest of economy, we present just a part of the obtained MC results to show the general picture of the evaluations. The power of the proposed tests was simulated under various alternatives. To describe the MC simulations, we use the following abbreviations: $N()$ is the normal distribution with mean μ and standard deviation σ ; $u(a, b)$ is the uniform distribution between a and b ; $e(\beta)$ is the exponential distribution with parameter β ; and $\log N(\mu, \sigma)$ is the lognormal distribution with mean μ and standard deviation σ .

To complete table 1, we compare the powers of proposed one-sample tests with those of Wilcoxon (W) tests. In accordance with the statistical literature, with regard to normally distributed data and the change of location, the W test is considered to be close to an optimal test. We can see that our test provided powers that are not significantly different from those of the W test. However, in cases different from the one above, the proposed test is superior to the W test. For example, when the two-sided test is based on $u(-1, 0.8)$ distributed data with $n = 40$, the proposed test provides a power of 0.3965 versus the W test power of 0.1997, a 99% increase relative to the W test.

In tables 2 and 3, the powers of proposed two-sample tests are compared with those of W and Kolmogorov–Smirnov (KS) tests. In accordance with the statistical literature, with regard to normally distributed data and change of location, the W test is very powerful. In such cases, we observed that our test provided powers that are not significantly different from those of the W test. However, in cases different from the one above, the proposed test is superior to the W and KS tests. For example, when the two-sided test is based on $N(1, 1)$ and $N(0, 2)$ distributed data with sample sizes of 10 and 25, respectively, the power of the proposed test was significantly higher than that of the W and KS tests. The proposed, W, and KS test powers were 0.6603, 0.3257, and 0.3727, respectively, with the proposed test showing a 103% and a 77% increase relative to the W and the KS, respectively. Similar results were seen for the two-sided test with sample distributions $u(1, 2)$ and $\log N(0, 1)$. The proposed, W, and KS test powers at sample sizes of 10 and 20 were 0.9934, 0.2351, and 0.5446, respectively, with the proposed test showing a 323% and an 82% increase relative to the W and the KS, respectively. The same situations are observed, but not shown, when the test is based on two exponentially distributed data or two Cauchy-distributed data.

Table 2. The MC power comparisons between the proposed, W, and KS two-sample two-sided tests

Design		$N(1,1)$ and $N(0,0.5)$		$N(1,1)$ and $N(0,2)$		$u(-1,1)+e(1)$ and $u(-1,1)$	
n	m	Proposed	W	KS	Proposed	W	KS
10	5	0.5872	0.4967	0.3542	0.1980	0.1878	0.1384
	10	0.7906	0.7045	0.5221	0.3792	0.2477	0.1332
	15	0.8223	0.7935	0.8050	0.5445	0.2931	0.3716
	20	0.8663	0.8255	0.7809	0.5864	0.3174	0.3095
	25	0.8502	0.8577	0.8632	0.6603	0.3257	0.3727
	30	0.8613	0.8623	0.8605	0.7065	0.3485	0.3943
15	35	0.8657	0.8771	0.8929	0.7275	0.3589	0.4476
	5	0.6410	0.5672	0.4629	0.2039	0.2205	0.1874
	15	0.9417	0.8772	0.8527	0.6125	0.3604	0.3507
	20	0.9680	0.9083	0.9334	0.6852	0.4113	0.5209
	25	0.9671	0.9345	0.9532	0.7707	0.4505	0.5606
	30	0.9699	0.9426	0.9529	0.8181	0.4657	0.5659
20	35	0.9721	0.9489	0.9616	0.8492	0.4892	0.6014
	5	0.6805	0.6162	0.6556	0.2152	0.2542	0.2477
	20	0.9815	0.9596	0.9611	0.7655	0.4746	0.5153
	25	0.9862	0.9675	0.9770	0.8629	0.5060	0.6271
	30	0.9908	0.9729	0.9848	0.9051	0.5498	0.6837
	35	0.9900	0.9809	0.9887	0.9361	0.5734	0.7234
30	5	0.7258	0.7084	0.6853	0.2084	0.2907	0.2537
	25	0.9987	0.9928	0.9972	0.8930	0.5939	0.7415
	30	0.9991	0.9949	0.9963	0.9344	0.6505	0.7492
	35	0.9996	0.9965	0.9983	0.9628	0.6846	0.8295

Continued on next page

Design		$e(1)+u(0,0.5)$ and $e(1)$				$u(1,2)$ and $\log N(0,1)$				
n	m	Proposed	W	KS	Proposed	W	KS	Proposed	W	KS
10	5	0.0944	0.0936	0.0466	0.2835	0.1614	0.2617	0.2835	0.1614	0.2617
	10	0.1360	0.1056	0.0366	0.8104	0.2149	0.2450	0.8104	0.2149	0.2450
	15	0.1810	0.1314	0.1251	0.9571	0.2175	0.6398	0.9571	0.2175	0.6398
	20	0.1859	0.1424	0.0869	0.9934	0.2351	0.5446	0.9934	0.2351	0.5446
	25	0.1961	0.1437	0.1042	0.9971	0.2380	0.6575	0.9971	0.2380	0.6575
	30	0.2028	0.1459	0.1043	0.9996	0.2392	0.7225	0.9996	0.2392	0.7225
15	35	0.2207	0.1522	0.1184	0.9998	0.2249	0.7686	0.9998	0.2249	0.7686
	5	0.0898	0.1116	0.0662	0.2930	0.1783	0.2806	0.2930	0.1783	0.2806
	15	0.1885	0.1618	0.0931	0.9859	0.2877	0.6142	0.9859	0.2877	0.6142
	20	0.2189	0.1728	0.1546	0.9994	0.3096	0.8510	0.9994	0.3096	0.8510
	25	0.2450	0.1894	0.1671	0.9997	0.3340	0.8792	0.9997	0.3340	0.8792
	30	0.2793	0.1935	0.1537	>0.9999	0.3484	0.8976	>0.9999	0.3484	0.8976
20	35	0.2826	0.2041	0.1693	0.9999	0.3490	0.9218	0.9999	0.3490	0.9218
	5	0.0889	0.1208	0.1063	0.3168	0.2857	0.3235	0.3168	0.2857	0.3235
	20	0.2671	0.2054	0.1414	0.9998	0.3801	0.8458	0.9998	0.3801	0.8458
	25	0.3284	0.2225	0.1811	>0.9999	0.3980	0.9511	>0.9999	0.3980	0.9511
	30	0.3649	0.2283	0.2111	>0.9999	0.4134	0.9639	>0.9999	0.4134	0.9639
	35	0.3842	0.2452	0.2147	>0.9999	0.4377	0.9854	>0.9999	0.4377	0.9854
30	5	0.0841	0.1454	0.1044	0.3059	0.3244	0.3256	0.3059	0.3244	0.3256
	25	0.3479	0.2715	0.2471	>0.9999	0.4773	0.9851	>0.9999	0.4773	0.9851
	30	0.3762	0.2866	0.2255	>0.9999	0.5066	0.9859	>0.9999	0.5066	0.9859
	35	0.4133	0.3180	0.2906	>0.9999	0.5456	0.9950	>0.9999	0.5456	0.9950

Table 3. The MC power comparisons between the proposed, W, and KS two-sample one-sided tests (greater)

Design		$N(1,1)$ and $N(0,0.5)$			$N(1,1)$ and $N(0,2)$			$u(-1,1)+e(1)$ and $u(-1,1)$			$e(1)+u(0,0.5)$ and $e(1)$		
n	m	Proposed	W	KS	Proposed	W	KS	Proposed	W	KS	Proposed	W	KS
10	5	0.7073	0.6792	0.5926	0.3756	0.3085	0.2330	0.5236	0.5866	0.4055	0.1695	0.1749	0.0991
	10	0.9014	0.8169	0.7506	0.4183	0.3572	0.3111	0.7563	0.7295	0.5594	0.1843	0.1880	0.1147
	15	0.9477	0.8618	0.8787	0.4529	0.3984	0.4295	0.8650	0.8053	0.7326	0.1892	0.2080	0.1705
	20	0.9665	0.9015	0.8780	0.4724	0.4603	0.4471	0.8993	0.8431	0.7240	0.1939	0.2327	0.1475
	30	0.9798	0.9201	0.9096	0.5106	0.4850	0.5062	0.9554	0.8727	0.7641	0.1956	0.2267	0.1581
15	5	0.9825	0.9195	0.9216	0.5252	0.5114	0.5470	0.9530	0.8827	0.7809	0.2185	0.2468	0.1746
	10	0.7930	0.7453	0.6346	0.4179	0.3436	0.2377	0.6235	0.6438	0.4103	0.1896	0.1907	0.1023
	15	0.9440	0.8921	0.8869	0.5336	0.4175	0.4685	0.8371	0.8197	0.7054	0.2410	0.2138	0.1779
	20	0.9809	0.9396	0.9394	0.5812	0.5043	0.5384	0.9277	0.8855	0.8048	0.2569	0.2534	0.1901
	30	0.9913	0.9551	0.9461	0.6344	0.5368	0.5609	0.9635	0.9227	0.8117	0.2620	0.2795	0.1846
20	5	0.9950	0.9705	0.9762	0.6822	0.6100	0.6831	0.9859	0.9465	0.8915	0.2758	0.3009	0.2399
	10	0.8151	0.7891	0.6510	0.4648	0.3599	0.2418	0.6393	0.6741	0.4181	0.2132	0.1935	0.1016
	15	0.9643	0.9396	0.9084	0.6110	0.4687	0.4584	0.8848	0.8715	0.7349	0.2695	0.2486	0.1733
	20	0.9907	0.9703	0.9596	0.6739	0.5438	0.5417	0.9576	0.9289	0.8374	0.3174	0.2942	0.1877
	30	0.9964	0.9793	0.9827	0.7428	0.5967	0.6869	0.9809	0.9582	0.9206	0.3125	0.3027	0.2574
30	5	0.9990	0.9889	0.9899	0.7941	0.6715	0.7514	0.9945	0.9749	0.9317	0.3450	0.3424	0.2614
	10	0.8424	0.8464	0.7714	0.5489	0.4689	0.3356	0.6866	0.7097	0.5197	0.2570	0.2054	0.1410
	15	0.9975	0.9907	0.9901	0.7902	0.6010	0.6571	0.9837	0.9627	0.9289	0.3809	0.3297	0.2516
	20	0.9992	0.9948	0.9960	0.8511	0.6576	0.7265	0.9958	0.9818	0.9611	0.4251	0.3530	0.2712
	30	0.9999	0.9987	0.9988	0.9104	0.7589	0.8426	0.9994	0.9923	0.9851	0.4734	0.4074	0.3358
35	5	0.9999	0.9988	0.9992	0.9248	0.7869	0.8830	0.9995	0.9950	0.9877	0.5068	0.4242	0.3717

To complete table 4, we compared the powers of proposed three-sample tests with those of ANOVA F and KW tests. In accordance with the statistical literature, with regard to normally distributed data and the change of location, the F test is considered to be close to an optimal test. We observed that our test provided powers that are not significantly different from those of the F test. However, in cases different from the one above, the proposed test is superior to the F test. For example, when the test is based on $N(0, 1)$, $N(0, 1)$, and $u(-1, 1)$ distributed data with sample sizes of 37, 45, and 50, respectively, the proposed test provides a power of 0.9809 versus the F test power of 0.0610 and the KW test power of 0.0577, with the proposed test having a 1508% and a 1600% increase relative to the F and the KW test, respectively. The proposed three-sample test outperformed the F and KW tests under almost all the tested alternatives.

Table 4. The MC power comparisons between the proposed, F , and KW three-sample tests

Distributions	n	m	k	Proposed	F	KW
$N(0,1) \times 2: u(-1,1)$	25	15	20	0.3745	0.0573	0.0526
	25	25	25	0.5578	0.0529	0.0521
	37	45	50	0.9809	0.0610	0.0577
$e(1) \times 2: \log N(0,1)$	25	15	20	0.2313	0.2306	0.2386
	37	45	50	0.6023	0.5481	0.5241
$N(0,1) \times 2: N(0,2.25)$	25	15	20	0.5908	0.0685	0.0599
	25	25	25	0.7719	0.0624	0.0576
	37	45	50	0.9941	0.0497	0.0544
$N(0,1) \times 2: N(0,0.25)$	25	15	20	0.9950	0.0610	0.0664
	25	25	25	0.9998	0.0584	0.0651
$e(1): \log N(0,1): u(0,1)$	25	15	20	0.9645	0.8004	0.6745
	25	25	25	0.9980	0.9121	0.8311

To analyze the methods for calculating critical values of the exact tests mentioned in sections 3.1, 3.2, and 3.3, we conducted the following MC comparisons. In table 5, the type I errors for the proposed two-sample test were compared for the MC, interpolation, and hybrid methods. The two-sample test was based on $N(0, 1)$ and $N(0, 1)$ distributed data samples. The type I errors for the test were appropriately controlled at 5% for all three methods. For example, when the test is based on sample sizes of 35 and 37, the actual type I errors are 0.0483, 0.0526, and 0.0526 for the MC, interpolation, and hybrid methods, respectively.

Table 5. The MC type I error comparisons between proposed methods for a two-sample test

Design	n	m	MC	Interpolation	Hybrid
$N(0,1) \times 2$	32	35	0.0538	0.0520	0.0472
		40	0.0533	0.0583	0.0513
		50	0.0519	0.0446	0.0498
	35	37	0.0483	0.0526	0.0526
		50	0.0476	0.0327	0.0488
	37	50	0.0513	0.0467	0.0526
	40	50	0.0469	0.0351	0.0498

The type I errors for the proposed three-sample test were compared for the MC, interpolation, and hybrid methods in table 6. The three-sample test was based on sample distributions of $N(0, 1)$, $N(0, 1)$, and $N(0, 1)$. The type I errors for the test were appropriately controlled at 5% for all three methods. The hybrid and MC methods are comparable, but the hybrid method on average uses five times fewer MC repetitions.

Table 6. The MC type I error comparisons between proposed methods for a three-sample test

Design	n	m	k	MC	Interpolation	Hybrid
$N(0,1) \times 3$	32	40	50	0.0493	0.0454	0.0519
	35	35	35	0.0546	0.0415	0.0501
	50	37	40	0.0477	0.0441	0.0466

As mentioned in sections 3.1, 3.2, and 3.3, the interpolation and hybrid methods use tables of CVs.

6 Application

In this section, the proposed three-sample comparison test is illustrated via a dataset of blood test results for patients with anemia (Wians et al. 2001). A total of 134 patients with anemia underwent a series of blood tests. To eliminate the bias that might be caused by gender, researchers limited the analysis to the 55 female study patients. Ferritin concentration provides a useful screening test for iron deficiency anemia (IDA). Nonpregnant women with anemia and a ferritin concentration less than $20 \mu\text{g/l}$ were assigned to the IDA group, while those with anemia and a ferritin concentration greater than $240 \mu\text{g/l}$ were assigned to the anemia of chronic disease (ACD) group. The interme-

diate group consists of the women with ferritin concentration between 20 and 240 $\mu\text{g/l}$. There were 12, 14, and 29 females in the ACD, intermediate, and IDA groups, respectively. We want to compare the sample distributions of the two rapid blood tests—that is, total iron-binding capacity (TIBC) and percent transferrin saturation (%TS)—to discriminate between the ACD, intermediate, and IDA groups (Tian et al. 2011).

Tian et al. (2011) focus on the confidence interval estimation of the differences in paired volumes under surfaces (VUS) and paired partial volumes under surfaces (PVUS). The 95% confidence intervals were [0.1103, 0.5139] for ΔVUS and [0.038, 0.1515] for ΔPVUS . Both confidence intervals showed that TIBC had better diagnostic ability than %TS. The assumption of multivariate normality for each group was tested and not rejected.

We use the `vxdBEL` command to conduct three-sample comparison tests, comparing the three groups within the two rapid blood tests. In addition, we use the `vxdBEL` command to conduct two-sample comparison tests, comparing each group between the two rapid blood tests. First, we conduct the three-sample comparison of the three groups for the TIBC.

```
. use tbc
. vxdBEL ICD Intermediate ACD
```

	T0	P-value
DBEL (3-sample)	44.2553	<0.0001

The test results indicate that the distributions of the three groups are not equal.

Similarly, we conduct the three-sample comparison of the three groups for the %TS.

```
. use ts
. vxdBEL ICD Intermediate ACD
```

	T0	P-value
DBEL (3-sample)	25.1539	0.0003

The test results indicate that the distributions of the three groups are not equal.

Thus the proposed three-sample test can discriminate between the three groups for both the TIBC and the %TS blood tests.

In addition, we conducted a two-sample comparison of TIBC versus %TS for each group. We note that the %TS and TIBC result values have completely different scales. The %TS results range within single digits, and the TIBC results range within hundreds, so a standardization transformation was applied to make the results of the blood tests comparable. For brevity, the `vxdBEL` results are not presented. The p -values for the three two-sample tests were 0.4556, 0.9647, and 0.0939 for the IDA, intermediate, and ACD groups, respectively. The test results suggest that the distributions between TIBC and %TS for each group are equal.

In conducting clinical experiments, researchers often desire to compare populations receiving different treatments to establish equivalence or desire to test identical dis-

tribution assumptions of various statistical tests. Thus the proposed Stata command can be used in such experiments without assumptions on data distributions (Tian et al. [2011] assumed normal distributions of observations). In general, there is excellent applicability for discriminating between samples in any context.

7 Concluding remarks

In this article, we proposed and examined the one-sample symmetry and K -sample comparison DBEL ratio tests. The tests are shown to be exact and robust nonparametric tests that approximate the optimal LR test statistic. The powers of the tests were comparable and in many cases outperformed the classical tests.

To date, simple DBEL tests have not been presented in Stata but are considered very efficient in practice. We developed and presented a command for Stata to perform the discussed approaches. The data example was used to demonstrate that the command is simple to use. The command performs both one-sided and two-sided alternatives for one-sample symmetry and two-sample comparison tests; it also performs a three-sample comparison test. The command can perform three methods: the MC method for best accuracy and the interpolation and hybrid methods for improved speed. The command is freely available for download at

<http://sphhp.buffalo.edu/biostatistics/research-and-facilities/software/stata.html>

The MC study performed in this article confirmed powerful properties of the proposed tests. We demonstrated that our one- and two-sample tests outperform the classical nonparametric tests of W and KS and that the three-sample test outperforms the F and KW tests under various alternatives. Further, this is accomplished while appropriately controlling the type I error for all the tests. The data example shows that the proposed test can be easily and efficiently used by practitioners.

8 Acknowledgments

This research is supported by the National Institutes of Health grant 1R03DE020851-01A1 (the National Institute of Dental and Craniofacial Research). The authors are grateful to the editor and the referee for suggestions that led to a substantial improvement in this article and the proposed command.

9 References

- Fan, J., M. Farman, and I. Gijbels. 1998. Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society, Series B* 60: 591–608.
- Gibbons, J. D., and S. Chakraborti. 2005. *Nonparametric Statistical Inference*. 4th ed. New York: CRC Press.

- Gurevich, G. 2012. Two-sample density-based empirical likelihood tests for stochastically ordered alternatives. In *Proceedings of the 6th International Days of Statistics and Economics*, 425–435. Prague: Melandrium.
- Lazar, N. A. 2003. Bayesian empirical likelihood. *Biometrika* 90: 319–326.
- Lehmann, E. L., and J. P. Romano. 2005. *Testing Statistical Hypotheses*. New York: Springer.
- Metropolis, N., and S. Ulam. 1949. The Monte Carlo method. *Journal of the American Statistical Association* 44: 335–341.
- Miecznikowski, J. C., A. Vexler, and L. Shepherd. 2013. dbEmplikeGOF: An R package for nonparametric likelihood ratio tests for goodness-of-fit and two-sample comparisons based on sample entropy. *Journal of Statistical Software* 54: 1–19.
- Owen, A. B. 2001. *Empirical Likelihood*. New York: Chapman & Hall/CRC.
- Pearson, E., and H. Hartley. 1966. *Biometrika Tables for Statisticians Volume I*. New York: Cambridge.
- Rubinstein, R., and D. Kroese. 2008. *Simulation and the Monte Carlo Method*. 2nd ed. Hoboken, NJ: Wiley.
- Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
- Tian, L., C. Xiong, C. Lai, and A. Vexler. 2011. Exact confidence interval estimation for the difference in diagnostic accuracy with three ordinal diagnostic groups. *Journal of Statistical Planning and Inference* 141: 549–558.
- Tsai, W., A. Vexler, and G. Gurevich. 2013. An extensive power evaluation of a novel two-sample density-based empirical likelihood ratio test for paired data with an application to a treatment study of attention-deficit/hyperactivity disorder and severe mood dysregulation. *Journal of Applied Statistics* 40: 1189–1208.
- Vexler, A., W. Deng, and G. E. Wilding. 2013. Nonparametric Bayes factors based on empirical likelihood ratios. *Journal of Statistical Planning and Inference* 143: 611–620.
- Vexler, A., and G. Gurevich. 2010. Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy. *Computational Statistics and Data Analysis* 54: 531–545.
- . 2011. A note on optimality of hypothesis testing. *Journal Mesa* 2: 243–250.
- Vexler, A., G. Gurevich, and A. D. Hutson. 2013. An exact density-based empirical likelihood ratio test for paired data. *Journal of Statistical Planning and Inference* 143: 334–345.

- Vexler, A., W. Tsai, G. Gurevich, and J. Yu. 2012. 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. *Statistics in Medicine* 31: 1821–1837.
- Vexler, A., C. Wu, and K. F. Yu. 2010. Optimal hypothesis testing: From semi to fully Bayes factors. *Metrika* 71: 125–138.
- Vexler, A., and J. Yu. 2010. Two-sample density-based empirical likelihood ratio tests for incomplete data in application to a pneumonia study. *Biometrical Journal* 52: 348–361.
- Wians, F. H., J. E. Urban, J. H. Keffer, and S. H. Kroft. 2001. Discriminating between iron deficiency anemia and anemia of chronic disease using traditional indices of iron status vs transferrin receptor concentration. *American Journal of Clinical Hematopathology* 115: 112–118.

About the authors

Albert Vexler is an associate professor in the Department of Biostatistics at New York State University at Buffalo. The density-based empirical likelihood methodology belongs to his main research interests.

Hovig Tanajian was employed as a research assistant and was a PhD student in the Department of Biostatistics at New York State University at Buffalo.

Alan D. Hutson is the chair of and a professor in the Department of Biostatistics at New York State University at Buffalo.