On Graphical Methods for Distributional Difference Between Two Populations

Technical Report No. ASU/2022/1

Dated 7 March, 2022

Shyamsundar Sahoo
and
Debasis Sengupta

Indian Statistical Institute
Applied Statistics Unit
Kolkata 700 108

On Graphical Methods for Distributional Difference Between Two Populations

Shyamsundar Sahoo*

Department of Statistics, Haldia Government College, Haldia, WB 721657, India
(*sssahoo.stats@gmail.com*)

and

Debasis Sengupta

Applied Statistics Unit, Indian Statistical Institute, Kolkata, WB 700108, India
(*sdebasis@isical.ac.in*)

March 6, 2022

Abstract

In this paper, we develop graphical tests for the goodness of fit of the two-sample location and location-scale models based on the quantile comparison function (QCF), which is the theoretical equivalent of the graph of a Q-Q plot. In each case, the QCF has a particular parametric form, and thus the estimates of the unspecified parameters lead to 'parametric' estimate of the QCF. The proposed methods are based on comparison of the parametric estimate under a specific model with a nonparametric estimate. We study the difference between the two estimators of the QCF as a stochastic process, and use the maximum difference as a pivot for statistical tests. Monte Carlo simulation study suggests reasonable small sample performance of the proposed methods. Moreover, the proposed graphical test for the location shift model is found to have much higher power than the existing graphical tests, which are based on the confidence bands of certain functions of the QCF. The proposed tests are illustrated through a data set on parallax angles of the Sun measured by two different methods. Proofs of theorems and some illustrations are given in the Appendix.

Keywords: Acceptance band; Confidence band; Graphical test; Quantile comparison function; Quantile-Quantile plot; Two-sample problem.

*Correspondence to: sssahoo.stats@gmail.com

1 Introduction

Graphical tests of hypotheses are diagnostic plots for hypothesized models, augmented with probabilistic limits that are supposed to include, under the null hypothesis, an ideal graph. Such tests have gained popularity because of their ability to combine the intuitive value of diagnostic plots with the formalism of analytical tests. These techniques have been developed mostly in the domain of survival analysis and have evolved sufficiently to have very competitive power vis-a-vis analytical tests; see Sahoo and Sengupta (2016).

Doksum (1974), and several other authors after him, proposed confidence bands for an empirical shift function related to the two-sample Q-Q plot, based on theoretical analysis. These bands can be converted into confidence bands for a Q-Q plot. Atkinson (1985) suggested a confidence envelope of the Q-Q plot by using parametric bootstrap. Under the hypothesis of equal distributions, these bands have specified (large sample) probability of containing the straight line through the origin with unit slope. Hence, these may be regarded as graphical tests of that hypothesis in the formal sense.

Several graphical tests have been proposed in the literature for checking whether two populations have identical distributions up to an unspecified location shift; see Doksum (1974) and Doksum and Sievers (1976). These graphical tests are based on an estimated shift function, and they reject the null hypothesis if no horizontal line can pass through a suitable confidence band of the shift function. There are analogous tests for the goodness of fit of other models that connect the two samples through scale change, location and scale change, and so on. We refer the reader to the works of Doksum (1977), Fisher (1983) and Thas (2009) for graphical methods of comparing two populations and for extension of these methods to several populations. In this paper, we argue that the premise of these graphical tests has an inherent slackness, and show how they can be improved.

In Section 2, we review the existing graphical tests of distributional difference in two populations, and explain the common reason for their inadequacy. In Section 3, we develop

a new graphical test for the location shift hypothesis. In Section 4, we develop another graphical test for the hypothesis of location-scale relation between two populations. For each of the two testing problems, the alternative hypothesis is that the two populations have completely arbitrary distributions. Both the proposed tests are asymptotic. In Section 5, we investigate the small sample performance of the proposed tests through simulations. We illustrate the proposed methods through the analysis of a real data set in Section 6. Section 7 contains some concluding remarks and indications of future work. Proofs of all the theorems, an illustration of efficacy of acceptance band and an additional data analytic illustration are given in Appendices A, B and C, respectively.

2 Weakness of Existing Graphical Tests of Distributional Difference

Let X and Y be two random variables with continuous and strictly increasing distribution functions F and G, respectively. Under the location model, there exists a non-zero parameter θ such that $F(x) = G(x + \theta)$ for all x. On the other hand, one can always write $F(x)$ as $G(x + \Delta(x))$, where $\Delta(\cdot)$ is the shift function defined by

$$\Delta(x) = G^{-1}F(x) - x, \qquad x \in S(F)$$

and $S(F)$ is the support of F. Doksum (1974) studied various properties of the shift function (SF). The quantile comparison function (QCF) is $G^{-1}F(x)$. Its graph is identical to the graph of G^{-1} against F^{-1}, which is the theoretical counterpart of a two-sample Q-Q plot. Under the location shift model, the plot of SF is a horizonal line with intercept θ and the plot of QCF is a straight line with unit slope and intercept θ.

Let F_m and G_n be the empirical distribution functions (EDFs) based on the samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$, respectively. Define the left continuous inverse function

of G_n by $G_n^{-1}(t) = \inf\{u : G_n(u) \geq t\}$. Then, a natural estimator of $\Delta(x)$ is given by

$$\widehat{\Delta}(x) = G_n^{-1}F_m(x) - x, \qquad x \in S(F_m).$$

Note that the graph of $G_n^{-1}F_m$, the empirical QCF, is similar to the Q-Q plot of the second sample against the first.

Doksum (1974) proposed a conservative graphical test based on asymptotic simultaneous confidence band (ASCB) of $\Delta(x)$. This test rejects the location shift hypothesis if no horizontal line can be passed through the band.

Doksum and Sievers (1976) proposed another conservative graphical test by using the distribution-free SCB of $\Delta(x)$ obtained from the two-sample Kolmogorov-Smirnov statistic, and the test rejects the hypothesis if no horizontal line can pass through this band. Doksum and Sievers (1976) referred to this band as the S-band for $\Delta(x)$.

Doksum and Sievers (1976) also proposed a graphical test based on another distribution-free SCB of $\Delta(x)$ obtained from the distribution of a weighted supremum norm Kolmogorov-Smirnov statistic, and the test rejects the hypothesis if no horizontal line fits in the band. Doksum and Sievers (1976) referred to this band as the W-band for $\Delta(x)$.

Construction of the above SCBs entails extensive use of empirical process theory (Shorak and Wellner, 1986). Doksum and Sievers (1976) also proposed variants of the above graphical tests based on several other types of confidence bands of $\Delta(x)$. See the works of Switzer (1976) for other confidence bands that may be turned into graphical tests.

All the above graphical tests are based on confidence bands of $\Delta(x)$ (i.e. of the SF plot). The QCF plot is more informative than the SF plot for studying the distributional relationships between two populations, as it permits interpretation of various forms of departures from the ideal shape, i.e., a straight line (Wilk and Gnanadesikan, 1968). One can easily adapt these plots to confidence bands of $G^{-1}F(x)$ (rather than those of $\Delta(x)$). In this adaptation, a test of the location shift model would consist of checking whether a

suitable confidence band of $G^{-1}F(x)$ contains a straight line with unit slope and arbitrary intercept. Apart from these adapted tests, there is also a graphical test for comparing several populations based on SCBs of multisample QCF plots, proposed by Nair (1982).

In order to appreciate the conservative nature of the above tests, let us represent the SF plot and the QCF plot by the graph of a function $h(x)$, for different forms of $h(x)$. For testing the location shift model, the null and the alternative hypotheses can be written as

$$\begin{aligned} H_0 &: h(x) = h_0(x;\theta)\ \forall x \text{ and for some } \theta(\neq 0) \in \Theta_0, \\ \text{vs. } H_1 &: h(x) \neq h_0(x;\theta)\ \forall x \text{ and for any } \theta(\neq 0) \in \Theta_0, \end{aligned}$$

where $h_0(x;\theta)$ is a completely known function of x except for the parameter θ. The forms of $h(x)$ and $h_0(x;\theta)$ for the two plots are given in Table 1.

Table 1. Different plots and, corresponding $h(x)$ and $h_0(x;\theta)$

Plot	Form of $h(x)$	Form of $h_0(x;\theta)$	Interpretation of θ
SF	$G^{-1}F(x) - x$	θ	intercept parameter
QCF	$G^{-1}F(x)$	$x + \theta$	intercept parameter

For testing the composite null hypothesis, the current approach is to obtain a $(1-\alpha)$ level SCB of $h(x)$ in the form of $\widehat{h}_l(x)$ and $\widehat{h}_u(x)$, such that

$$\inf_{\theta\in\Theta_0} P_\theta \left\{ \widehat{h}_l(x) \leq h_0(x;\theta) \leq \widehat{h}_u(x), \forall x \right\} = 1-\alpha.$$

A typical graphical test rejects H_0 at level α if *no* $h_0(x;\phi)$ with $\phi \in \Theta_0$ lies completely within the band (Doksum, 1974; Doksum and Sievers, 1976).

The effective size of the said graphical test is

$$\begin{aligned} &1 - \inf_{\theta\in\Theta_0} P_\theta \left[\bigcup_{\phi\in\Theta_0} \left\{ \widehat{h}_l(x) \leq h_0(x;\phi) \leq \widehat{h}_u(x), \forall x \right\} \right] \\ \leq &1 - \inf_{\theta\in\Theta_0} P_\theta \left\{ \widehat{h}_l(x) \leq h_0(x;\theta) \leq \widehat{h}_u(x), \forall x \right\}. \end{aligned}$$

The right hand side would go to α for sample size going to infinity, as per the construction of the test. The inequality indicates that such a graphical test can have excessive gap between level and size. There appears to be room for improvement of the power.

As an illustration of the excessively conservative nature of the existing graphical tests, let us consider simulated X-samples and Y-samples from $N(0,\ 1)$ and $N(1,\ 2)$ distributions, respectively (where $N(\mu,\ \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2). Thus, the two distributions are *not* connected by location shift. We choose $m = n = 100$ and $\alpha = 0.05$. Figure 1 exhibits the empirical QCF plot together with the 95% ASCB, the 95% S-band and the 95% W-band of the QCF. Figure 1 also shows the graphs of two extreme straight lines with unit slope and different intercepts, which can be drawn within

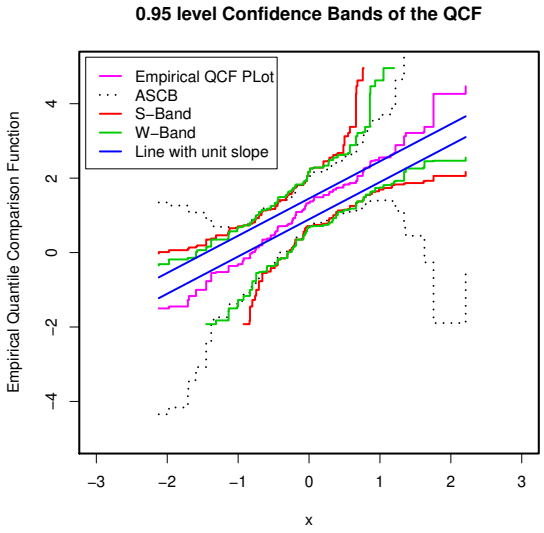


Figure 1: Empirical QCF plot for data simulated from location-scale model, together with the 95% ASCB, the 95% S-band and the 95% W-band of the QCF. The parallel lines represent extreme straight lines with unit slope that can pass through all three bands.

all the three bands. Thus, a conventional graphical test based on either of the three bands would fail to reject the hypothesis of location shift.

We propose in this paper a new graphical test of the location shift hypothesis based on the difference between parametric and non-parametric estimates of the QCF, which would have higher power than the existing graphical tests.

If the location shift hypothesis does not appear to hold for a particular data, a natural question is: Does a location-scale model hold, i.e. is $G^{-1}F(x) = \delta x + \theta$ for all x and for some $\delta(\neq 1)$ and $\theta(\neq 0)$? To answer this question, we propose another graphical test for checking the location-scale relationship between two populations.

3 Modified Test For Location Shift Model

3.1 Existing Graphical Tests

The level $(1-\alpha)$ ASCB for $G^{-1}F(x)$ proposed by Doksum (1974) is the pair of curves

$$G_n^{-1}F_m(x) \pm \frac{K_\alpha}{\sqrt{M}\ \widehat{g}\left(G_n^{-1}F_m(x)\right)}, \quad x \in S(F_m),$$

where $M = \frac{mn}{m+n}$, $\widehat{g}(\cdot)$ is a kernel density estimate (KDE) of $g(\cdot)$, and K_α is the $(1-\alpha)$th quantile of the supremum norm of the Brownian Bridge obtained from weak convergence of

$$\widehat{g}\left(G_n^{-1}F_m(x)\right)\sqrt{M}(\widehat{\Delta}(x) - \Delta(x)).$$

The level $(1-\alpha)$ S-band (distribution-free SCB) for $G^{-1}F(x)$ proposed by Doksum and Sievers (1976) is given by the pair of curves

$$\left(G_n^{-1}\{F_m(x) - K_{s,\alpha}/\sqrt{M}\},\ G_n^{-I}\{F_m(x) + K_{s,\alpha}/\sqrt{M}\}\right), \quad x \in S(F_m),$$

which is obtained from the distribution of the supremum norm pivot

$$D_{m,n} = \sup_{x \in S(F_m)} \sqrt{M}\,|F_m(x) - G_n(x)|\,.$$

In the above expressions $G_n^{-I}(t) = \sup\{u : G_n(u) \leq t\}$ is the right continuous inverse of G_n, and $K_{s,\alpha}$ is such that $P_{F=G}\{D_{m,n} \leq K_{s,\alpha}\} = 1-\alpha$ and is obtained from the Kolmogorov-

Smirnov tables (Pearson and Hartley, 1972).

The level $(1-\alpha)$ W-band for $G^{-1}F(x)$, also proposed by Doksum and Sievers (1976), is given by the pair of curves

$$\left(G_n^{-1}\left[h^-\{F_m(x)\}\right], G_n^{-I}\left[h^+\{F_m(x)\}\right]\right), \quad x \in S(F_m),$$

which is obtained from the distribution of a weighted supremum norm statistic of the form

$$W_{m,n} = \sup_{x \in S(F_m)} \sqrt{M}\frac{|F_m(x) - G_n(x)|}{\Psi\{H_{m,n}(x)\}},$$

where $H_{m,n}(x) = \widehat{\lambda}F_m(x) + (1-\widehat{\lambda})G_n(x)$, $\widehat{\lambda} = m/(m+n)$ and $\Psi(t) = \{t(1-t)\}^{1/2}$. Explicit forms of the functions h^- and h^+, given in Doksum and Sievers (1976), depend on the quantiles of the distribution of $W_{m,n}$ under $F = G$, obtained in Canner (1975).

If any one of the above three $(1-\alpha)$ level confidence bands of the QCF does not contain a straight line with unit slope and arbitrary intercept, the corresponding graphical test rejects H_0 at level α.

3.2 Modified Graphical Test

As mentioned in Section 2, the function $h(x)$ has the particular parametric form $h_0(x;\theta)$ under the null hypothesis. Therefore, we can construct a test based on comparing a nonparametric estimator of $h(x)$ with its parametric counterpart under the null hypothesis. If $\widehat{h}$ is a nonparametric estimator of h, and $\widehat{\theta}$ is a consistent estimator of θ under H_0, we would expect the difference $\widehat{h}(x) - h_0(x;\widehat{\theta})$ to be small when H_0 is true. Therefore, the difference function can be used to form a test statistic for H_0.

As for the QCF plot, $h(x) = G^{-1}F(x)$, and $G_n^{-1}F_m(x)$ would be a natural nonparametric estimator of $h(x)$, which is consistent under appropriate conditions. We choose the estimator of $h_0(x;\theta)$ as $x + \widehat{\theta}$, where $\widehat{\theta} = G_n^{-1}F_m(x^0) - x^0$ for some fixed $x^0 \in S(F)$.

Let us denote by $D_L(x)$ the true difference function $h(x) - h_0(x;\theta)$, which is simply zero under H_0. Let us consider the consistent estimator of $D_L(x)$, with $\widehat{\theta}$ chosen as above,

$$\widehat{D}_L(x) = \widehat{h}(x) - h_0(x;\widehat{\theta}) = G_n^{-1}F_m(x) - [x + \widehat{\theta}]. \tag{1}$$

Define $N = m+n$, $M = mn/N$ and $\widehat{\lambda} = m/N$. Let the supports, $S(F)$ and $S(G)$, of F and G be the finite intervals $[a, b]$ and $[c, d]$, respectively. Further, let $W^0(x)$ denote the standard Brownian bridge on $[0, 1]$. Then the following theorem establishes weak convergence of the difference process, which takes place in the space $D[a, b]$ of the functions on $[a, b]$ that are right continuous having finite left limits, equipped with the Skorohod topology.

THEOREM 1. *Let $M \to \infty$ in such a way that $\widehat{\lambda} \to \lambda$ for some constant $\lambda \in (0, 1)$, and G be a strictly increasing function with continuous derivative g on $[c, d]$. Then, under the hypothesis of location shift, the difference process $\sqrt{M}\widehat{D}_L(x)$ converges weakly in the space $D[a, b]$ to a mean zero Gaussian process*

$$\frac{W^0(F(x))}{g(G^{-1}F(x))} - \frac{W^0(F(x^0))}{g(G^{-1}F(x^0))}. \tag{2}$$

The variance function of the process (2) is

$$\begin{aligned}\sigma^2_{D_L}(x) &= \frac{F(x)(1-F(x))}{g^2(G^{-1}F(x))} + \frac{F(x^0)(1-F(x^0))}{g^2(G^{-1}F(x^0))} - \frac{2}{g(G^{-1}F(x))g(G^{-1}F(x^0))} \times \\ &\quad \left\{F(x)(1-F(x^0)) \cdot 1(x \le x^0) + F(x^0)(1-F(x)) \cdot 1(x > x^0)\right\}. \end{aligned} \tag{3}$$

By replacing g with a consistent estimator $\widehat{g}$, one can consistently estimate $\sigma^2_{D_L}(x)$ by

$$\begin{aligned}\widehat{\sigma}^2_{D_L}(x) &= \frac{F_m(x)(1-F_m(x))}{\widehat{g}^2(G_n^{-1}F_m(x))} + \frac{F_m(x^0)(1-F_m(x^0))}{\widehat{g}^2(G_n^{-1}F_m(x^0))} - \frac{2}{\widehat{g}(G_n^{-1}F_m(x))\widehat{g}(G_n^{-1}F_m(x^0))} \times \\ &\quad \left\{F_m(x)(1-F_m(x^0)) \cdot 1(x \le x^0) + F_m(x^0)(1-F_m(x)) \cdot 1(x > x^0)\right\}. \end{aligned}$$

The function $\sigma^2_{D_L}(x)$ is continuous but not necessarily monotone on $[a, b]$. Therefore, the conventional route of transforming the Brownian motion to the standard Brownian bridge on $[0, 1/2]$ (Shorak and Wellner, 1986; Dabrowska et al., 1989) cannot be used to construct asymptotic confidence bands. Further, the variance function (3) assumes the value 0 at $x = x^0$, and it increases towards the end-points a and b. A weight function which gives less weight towards the end-points and more weight near x^0 would bring stability in the variance of the limiting process. In particular, if the process (2) is scaled by the density function $g(G^{-1}F(x))$ (which coincides with $g(x+\theta)$ under the null hypothesis), the limiting process would have the form

$$W^0(F(x)) - \frac{g(G^{-1}F(x))}{g(G^{-1}F(x^0))} W^0(F(x^0)),$$

which is the same as the process

$$F_L(p) = W^0(p) - \frac{g(G^{-1}(p))}{g(G^{-1}(p^0))} W^0(p^0), \tag{4}$$

defined on $[0, 1]$, where $p = F(x)$ and $p^0 = F(x^0)$.

The following theorem provides convergence of the supremum norm of the scaled difference process, which would be used as a pivot for constructing the graphical test.

THEOREM 2. *Let $M \to \infty$ in such a way that $\widehat{\lambda} \to \lambda$ for some constant $\lambda \in (0, 1)$. Also, let G be a strictly increasing function with continuous derivative g on $[c, d]$, $\widehat{g}$ be a pointwise consistent estimator of g having the property of uniform stochastic equiconinuity, and $\widehat{\theta}$ be a consistent estimator of θ under H_0. Then under the hypothesis of location shift,*

$$\sup_{a \le x \le b} \left| \sqrt{M}\, \widehat{g}\left(x + \widehat{\theta}\right)\, \widehat{D}_L(x) \right| \xrightarrow{\mathcal{D}} \sup_{0 \le p \le 1} |F_L(p)|\,,$$

where $\xrightarrow{\mathcal{D}}$ denotes weak convergence and F_L is the limiting Gaussian process in (4).

REMARK. If $\widehat{g}$ is chosen as the kernel density estimator of g defined by

$$\widehat{g}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - Y_i}{h_n}\right),$$

where $K(\cdot)$ is the kernel function and h_n is the bandwidth parameter, then its pointwise consistency is well established under the assumptions that the kernel $K(\cdot)$ is a non-negative, bounded and symmetric function satisfying $\int K(u)du = 1$, $\int u^2 K(u)du < \infty$ and $|uK(u)| \to 0$ as $|u| \to \infty$, and $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$ (Silverman, 1998). Further, the sufficient condition for uniform stochastic equiconinuity is known to be the stochastic Lipschitz continuity, and it can be shown that the Lipschitz continuity (in probability) of $\widehat{g}(x)$ holds under the bounded derivative assumption of the kernel function $K(\cdot)$.

In practice, the distribution of the supremum norm of F_L, for a specified p^0 in [0, 1], may be simulated by replacing g and G in (4) by the consistent estimators $\widehat{g}$ and G_n, respectively. If d_α is the $(1-\alpha)$th quantile of that distribution, then under the hypothesis of location shift, the line $x + \widehat{\theta}$ would be sandwiched between the graphs of $G_n^{-1}F_m(x) \pm d_\alpha/\{\sqrt{M}\ \widehat{g}(x + \widehat{\theta})\}$ with asymptotic coverage probability $(1 - \alpha)$. We refer to this pair of curves as a level $(1 - \alpha)$ asymptotic *acceptance band* for the parametric estimate of the QCF represented by the straight line $x + \widehat{\theta}$.

Thus, the hypothesis of location shift may be rejected at the level of significance α if $x + \widehat{\theta}$ does not fit into its level $(1 - \alpha)$ asymptotic acceptance band. The width of the band would be a decreasing function of α. The p-value of this test is the smallest level of significance that does not ensure inclusion of $x + \widehat{\theta}$ within its acceptance band.

3.3 Asymptotic Relative Efficiency

Efficiency of the nonparametric confidence bands with respect to the proposed acceptance band can be compared on the basis of their asymptotic widths. Doksum and Sievers (1976) derived, under some regularity conditions, the asymptotic widths (as

$M \to \infty$) of the S-band and the W-band which are, respectively, $2K_{s,\alpha}/g(G^{-1}(p))$ and $2K_{w,\alpha}\sqrt{p(1-p)}/g(G^{-1}(p))$ at the pth quantile (x_p) of the distribution F, $0 < p < 1$, and computed the asymptotic relative efficiency of the two bands by the reciprocal ratio of the squares of their asymptotic widths. In case of the acceptance band, it can be shown that the limiting value of $\sqrt{M}$ times the width of the acceptance band at quantile x_p is $2d_\alpha/g(G^{-1}(p))$ as $M \to \infty$. Thus, the asymptotic relative efficiencies of the S-band and the W-band in comparison with the acceptance band become

$$ARE_1(S/A) = \left(\frac{d_\alpha}{K_{s,\alpha}}\right)^2 \quad \text{and} \quad ARE_2(W/A) = \left(\frac{d_\alpha}{K_{w,\alpha}}\right)^2 \frac{1}{p(1-p)},$$

respectively, where the $(1-\alpha)$th quantiles d_α, $K_{s,\alpha}$ and $K_{w,\alpha}$ will be considered as the asymptotic critical values of the corresponding pivots. Further, the limiting width of the ASCB at x_p is $2K_\alpha/g(G^{-1}(p))$, implying that the asymptotic relative efficiency of the ASCB compared to the acceptance band, denoted by ARE_3, is $(d_\alpha/K_\alpha)^2$. Since the two-sample Kolmogorov-Smirnov statistic D_{mn} converges in distribution to the supremum norm of the standard Brownian bridge, $K_{s,\alpha}$ is asymptotically equal to K_α. Therefore, ARE_3 would have approximately the same value as ARE_1.

The computations reported in Section 5 indicate that ARE_1 and ARE_2 are mostly smaller than 1 for some common choices of G, that is, there is considerable advantage of using the proposed acceptance bands.

4 Test For Location-Scale Model

For this testing problem, the null and the alternative hypotheses can be expressed as

$$\begin{aligned} H_0' &: h(x) = h_0(x;\delta,\theta)\ \forall x, \text{ for some } \delta(\neq 1) \in \Theta_{01} \text{ and } \theta(\neq 0) \in \Theta_{02}, \\ \text{vs. } H_1' &: h(x) \neq h_0(x;\delta,\theta)\ \forall x, \text{ for any } \delta(\neq 1) \in \Theta_{01} \text{ and } \theta(\neq 0) \in \Theta_{02}, \end{aligned}$$

where $h(x) = G^{-1}F(x)$ and $h_0(x;\delta,\theta) = \delta x + \theta$.

The null hypothesis, H_0', corresponds to a particular parametric form of h. Therefore, with an argument analogous to the one used in Section 3, we can construct a test based on the supremum norm of the difference between a nonparametric estimator of h and a parametric estimator under H_0'.

For a nonparametric estimator of h, we already have $\widehat{h}(x) = G_n^{-1}F_m(x)$. A parametric estimator of h under H_0' is $h_0(x; \widehat{\delta}, \widehat{\theta}) = \widehat{\delta}x + \widehat{\theta}$, where $\widehat{\delta}$ and $\widehat{\theta}$ are consistent estimators of δ and θ, respectively. One can choose the consistent estimators

$$\widehat{\delta} = \frac{G_n^{-1}F_m(x_3) - G_n^{-1}F_m(x_1)}{x_3 - x_1} \quad \text{and} \quad \widehat{\theta} = G_n^{-1}F_m(x_2) - \widehat{\delta}x_2,$$

for fixed $x_1 < x_2 < x_3$ belonging to the support $S(F)$.

Let $D_{LS}(x) = h(x) - h_0(x; \delta, \theta)$, which is zero under H_0'. Let us denote the difference process, for $\widehat{\delta}$ and $\widehat{\theta}$ chosen as above, as

$$\widehat{D}_{LS}(x) = \widehat{h}(x) - h_0(x; \widehat{\delta}, \widehat{\theta}) = G_n^{-1}F_m(x) - [\widehat{\delta}x + \widehat{\theta}]. \tag{5}$$

The following theorem shows weak convergence of this difference process in $D[a, b]$.

THEOREM 3. *Let $M \to \infty$ in such a way that $\widehat{\lambda} \to \lambda$ for some constant $\lambda \in (0, 1)$, and G be a strictly increasing function with continuous derivative g on $[c, d]$. Then, under the hypothesis of location-scale model, the difference process $\sqrt{M}\widehat{D}_{LS}(x)$ converges weakly in the space $D[a, b]$ to the mean zero Gaussian process*

$$\frac{W^0(F(x))}{g(G^{-1}F(x))} - \frac{W^0(F(x_2))}{g(G^{-1}F(x_2))} - \left(\frac{x - x_2}{x_3 - x_1}\right)\left[\frac{W^0(F(x_3))}{g(G^{-1}F(x_3))} - \frac{W^0(F(x_1))}{g(G^{-1}F(x_1))}\right]. \tag{6}$$

The shape of the variance function of the limiting Gaussian process (6) will play an important role in developing the new test. One can easily obtain the variance function of this process, which is found to be continuous on $[a, b]$ (taking the minimum value 0 at

$x = x_2$). However, this function is generally nonmonotone. Therefore, the conventional argument that banks on monotonicity cannot be used to construct asymptotic confidence bands. We may simulate the limiting Gaussian process in order to develop a test procedure.

The limiting process (6) defined over $[a, b]$, scaled by the factor $g(G^{-1}F(x))$, can be written as the process

$$W^0(p) - \frac{g(G^{-1}(p))}{g(G^{-1}(p_2))}W^0(p_2)$$

$$- \left(\frac{F^{-1}(p) - F^{-1}(p_2)}{F^{-1}(p_3) - F^{-1}(p_1)}\right)\left[\frac{g(G^{-1}(p))}{g(G^{-1}(p_3))}W^0(p_3) - \frac{g(G^{-1}(p))}{g(G^{-1}(p_1))}W^0(p_1)\right] \tag{7}$$

defined over $[0, 1]$, where $p = F(x)$, $p_i = F(x_i)$, $i = 1, 2, 3$. The following convergence result would be used to construct the graphical test.

THEOREM 4. *Let $M \to \infty$ in such a way that $\widehat{\lambda} \to \lambda$ for some constant $\lambda \in (0, 1)$. Also, let G be a strictly increasing function with continuous derivative g on $[c, d]$, $\widehat{g}$ be a pointwise consistent estimator of g having the property of uniform stochastic equiconinuity, and $\widehat{\delta}$ and $\widehat{\theta}$ be consistent estimators of δ and θ, respectively, under H_0'. Then, if we denote by F_{LS} the limiting Gaussian process in (7), we have under H_0',*

$$\sup_{a \le x \le b} \left|\sqrt{M}\,\widehat{g}\left(\widehat{\delta}x + \widehat{\theta}\right)\,\widehat{D}_{LS}(x)\right| \xrightarrow{\mathcal{D}} \sup_{0 \le p \le 1} |F_{LS}(p)|\,.$$

Let e_α be the $(1-\alpha)$th quantile of the distribution of the supremum norm of F_{LS}, for a specified p_i, $i = 1, 2, 3$ in $[0, 1]$, with g, G and F replaced by their consistent estimators $\widehat{g}$, G_n and F_m, respectively. Then, under the hypothesis of the location-scale model, the line $\widehat{\delta}x + \widehat{\theta}$ would be sandwiched within the pair of curves $G_n^{-1}F_m(x) \pm e_\alpha/\{\sqrt{M}\;\widehat{g}(\widehat{\delta}x + \widehat{\theta})\}$ with asymptotic coverage probability $(1-\alpha)$. This pair of curves is referred to as a level $(1-\alpha)$ asymptotic *acceptance band* for the parametric estimate of the QCF represented by the straight line $\widehat{\delta}x + \widehat{\theta}$.

The proposed graphical test is to reject H_0' at level of significance α if the parametrically estimated curve represented by the straight line $\widehat{\delta}x + \widehat{\theta}$ does not fit into its level $(1-\alpha)$ asymptotic acceptance band. The p-value of the graphical test would be the smallest level of significance for which the parametrically estimated curve does not completely lie within its acceptance band.

5 Computational Study and Simulations

In this section, we compare the performance of the nonparametric confidence bands with that of the proposed acceptance band in terms of asymptotic relative efficiency and finite sample performance. Along with the proposed tests, we consider the existing graphical tests based on ASCB, S-band and W-band (see Section 3), through a Monte Carlo simulations.

In this study, we generate random samples from (i) some location shift models, (ii) some location-scale models and (iii) some non-location-scale models. Further, we use various combinations of distributions for the two populations connected by the above mentioned models. We denote by $N(\mu, \sigma^2)$ the normal distribution with mean μ and variance σ^2. Likewise, we denote by $L(\mu, \sigma)$, $Gam(\gamma, \zeta)$ and $C(\mu, \sigma)$ the Logistic distribution with location parameter μ and scale parameter σ (having variance $\pi^2\sigma^2/3$), the gamma distribution with scale parameter γ and shape parameter ζ (with mean ζ/γ and variance ζ/γ^2) and the Cauchy distribution with location parameter μ and scale parameter σ.

For study of finite sample performance, we consider sample sizes of 50, 100, 200 and 500 in each group and use the level of significance, $\alpha = 0.05$. We report the results based on $10,000$ simulation runs which corresponds to the standard error of (approximately) 0.0022 on the achieved significance level $\widehat{\alpha}$. Here, $\widehat{\alpha}$ is the number of simulation runs rejecting H_0, expressed as a fraction of the total number of runs. Thus, there is probability 0.95 that $\widehat{\alpha}$ would be less than $0.05 + 1.645 \times 0.0022$, that is, less than 0.0536.

5.1 Asymptotic Relative Efficiencies: Nonparametric Bands for Location Shift Model

We begin by computing the asymptotic relative efficiencies of the S-band and the W-band with respect to the proposed acceptance band for testing the location shift model. The asymptotic efficiencies reported in Table 2 are obtained by using the critical values for $m = n = 1,000$, and taking the supremum norm of the limiting Gaussian processes over the interval [0.025, 0.975] in case of each band. We choose $\alpha = 0.05$.

The critical value $K_{s,\alpha}$ is obtained from Appendix C of Klein and Moeschberger (2003, p.471) and $K_{w,\alpha}$ is determined based on 10,000 simulation runs according to Canner (1975), generating both X and Y-samples from $N(0,\ 1)$ distribution. The simulated value of $K_{w,0.05}$ is obtained as 3.1697, and $K_{s,0.05}$ is found as 1.3581.

Turning to the acceptance band, note that d_α is the $(1-\alpha)$th quantile of the supremum norm of F_L defined in (4). Its critical value depends on the true distribution G as well as the parameter $p^0 = F(x^0)$ used in computing $\widehat{\theta}$. In computing the asymptotic efficiencies, we use three choices of G: $N(0,\ 1)$, $L(0,\ 1)$ and $C(0,\ 1)$, and for each of these choices look for a suitable choice of p^0 from the interval $(0,\ 1)$. To determine the critical value d_α for each choice of G and p^0, we generate 10,000 sample paths, each path being evaluated at 2,000 uniform grid points over $[0,\ 1]$ (see Section 5.2). It was found through simulations (results omitted for brevity) that d_α is generally smaller (i.e., the acceptance band is narrower) when p^0 is around 1/2, for each of the three forms of G. We thus use $p^0 = 1/2$.

Table 2. Asymptotic relative efficiencies of the S-band and the W-band compared to the acceptance band ($\alpha = 0.05$, $m = n = 1,000$)

True G	d_α	ARE_1	ARE_2 with p equal to				
			0.1	0.2	0.3	0.4	0.5
$N(0,\ 1)$	1.0940	0.649	1.324	0.744	0.567	0.496	0.476
$L(0,\ 1)$	1.0792	0.631	1.288	0.724	0.552	0.483	0.464
$C(0,\ 1)$	1.0410	0.588	1.198	0.674	0.514	0.449	0.431

In Table 2, the values of ARE_2, which are symmetric about $p = 0.5$, are given only for $p \leq 0.5$. We see from Table 2 that the acceptance band is always better than the S-band. Furthermore, the acceptance band is also preferable with respect to the W-band (more efficient compared to the S-band) except for extreme quantiles ($p \notin (0.1, 0.9)$).

5.2 Empirical Size and Power: Tests for Location Shift Model

Here, we compute the empirical probabilities of rejection of the null hypothesis (location shift model) for the existing and the proposed graphical tests for the following six combinations of F and G: (i) $N(0,1)$ and $N(1,1)$, (ii) $L(0, \sqrt{3}/\pi)$ and $L(1, \sqrt{3}/\pi)$, (iii) $N(0,1)$ and $N(1,3)$, (iv) $L(0, \sqrt{3}/\pi)$ and $L(1,1)$, (v) $N(0,1)$ and $L(0,1)$ and (vi) $Gam(1,2)$ and $L(1,1)$. The first two choices correspond to the null hypothesis, i.e., the computed probability is the estimated size. For the next four choices (of which, the choices (iii) and (iv) correspond to the location-scale model), the computed probability is the estimated power.

In order to compute the percentage point d_α, we initially simulate a number of sample paths from the Brownian bridge $W^0(v)$ over [0, 1] through the process $B(v) - vB(1)$, where $B(\cdot)$ is the standard Brownian motion over [0, 1]. We calculate each sample path on a grid of $n_0 + 1$ equidistant points between 0 and 1, i.e. on grid points $v_i = i/n_0$, $i = 0, 1, \ldots, n_0$, by using the expression

$$W^0(v_i) = \sum_{j=1}^{i} \sqrt{v_j - v_{j-1}}\; z_j - v_i \sum_{j=1}^{n_0} \sqrt{v_j - v_{j-1}}\; z_j$$

for $i = 1, 2, \ldots, n_0$, where $z_1, z_2, \ldots, z_{n_0}$ are independent standard normal variates. Obviously, $W^0(v_0)$ is 0 with probability 1. Then, we generate 2000 sample paths of the limiting Gaussian process (4) with the following approximation:

$$W^0(v_i) - \frac{\widehat{g}\{G_n^{-1}(v_i)\}}{\widehat{g}\{G_n^{-1}(p^0)\}} W^0(p^0),$$

where p^0 is taken as 1/2 (i.e., x^0 is the median of the X-sample) and $\widehat{g}$ is the kernel density estimator of g, with Gaussian kernel and bandwidth chosen by Silverman's rule of thumb.

Sample paths are calculated over a uniform grid of size $n_0 = 500$ over [0, 1], while the supremum is over [0.025, 0.975], to avoid instability of empirical QCF around end points.

Table 3 reports the achieved sizes of the existing and the proposed graphical tests at nominal level of significance 0.05, together with the noncontainment error rates of the QCF confidence bands, for choices (i) and (ii) of F and G. These error rates are the empirical probabilities of the confidence bands not containing the true QCF. It is seen that the proposed graphical test considerably improves over the existing graphical tests. The existing tests are found to be very conservative, although the coverage probabilities of the confidence bands appear to be reasonable. This contrast shows that their underutilization of the available level is not explained by laxity of the confidence bands; it is rather attributable to the philosophy of using confidence bands for the purpose of testing.

Table 3. Achieved sizes of some existing and the proposed graphical tests of the two-sample location shift hypothesis, at nominal level 0.05 (based on 10,000 simulation runs)

Choice of F and G	Sample Size (m, n)	Achieved sizes of Existing tests based on ASCB	S-band	W-band	Proposed Test	Error rates of the confidence bands ASCB	S-band	W-band
(i)	50, 50	0.002	0.000	0.000	0.033	0.060	0.031	0.038
	100, 100	0.001	0.000	0.000	0.032	0.051	0.033	0.040
	200, 200	0.000	0.000	0.000	0.031	0.042	0.034	0.040
	500, 500	0.000	0.000	0.000	0.033	0.037	0.039	0.040
(ii)	50, 50	0.008	0.000	0.000	0.054	0.076	0.028	0.036
	100, 100	0.003	0.000	0.000	0.052	0.060	0.030	0.039
	200, 200	0.001	0.000	0.000	0.043	0.048	0.035	0.042
	500, 500	0.000	0.000	0.002	0.043	0.046	0.048	0.046

Table 4 presents the powers of the graphical tests for data generated from the location-scale (choices (iii) and (iv)) and the non-location-scale models (choices (v) and (vi)). It is observed that the power of the proposed test is much higher than that of the existing tests for sample sizes less than 200 in each group.

Simulations from other combinations of distributions for F and G (not reported here for the sake of brevity) confirmed the findings reported above.

Table 4. Achieved powers of some existing and the proposed graphical tests of the two-sample location shift hypothesis, at nominal level 0.05 (based on 10,000 simulation runs)

Choice of F and G	Sample Size (m, n)	Achieved powers of Existing tests based on ASCB	S-band	W-band	Proposed Test
(iii)	50, 50	0.000	0.005	0.076	0.745
	100, 100	0.000	0.104	0.649	0.975
	200, 200	0.074	0.733	0.995	1.000
	500, 500	0.993	1.000	1.000	1.000
(iv)	50, 50	0.000	0.006	0.083	0.823
	100, 100	0.000	0.102	0.602	0.985
	200, 200	0.058	0.738	0.988	1.000
	500, 500	0.990	1.000	1.000	1.000
(v)	50, 50	0.000	0.002	0.051	0.790
	100, 100	0.000	0.046	0.534	0.978
	200, 200	0.011	0.525	0.987	1.000
	500, 500	0.908	0.999	1.000	1.000
(vi)	50, 50	0.004	0.001	0.008	0.783
	100, 100	0.002	0.009	0.203	0.984
	200, 200	0.001	0.258	0.932	1.000
	500, 500	0.329	0.995	1.000	1.000

5.3 Empirical Size and Power: Test for Location-Scale Model

The proposed graphical test for the hypothesis of location-scale has no competitor. We compute the empirical probabilities of rejection of this hypothesis, with data generated from the following five combinations of F and G: (a) $N(0,1)$ and $N(1,4)$, (b) $L(0, \sqrt{3}/\pi)$ and $L(1, 2/\pi)$, (c) $Gam(1,2)$ and $Gam(2,2)$ (d) $Gam(1,1)$ and $L(0,1)$, (e) $Gam(1,1/2)$ and $Gam(2,3)$. For the first three choices (which correspond to the null hypothesis), the empirical probability is the estimated size, while for the other two it is the estimated power.

For the computation of cut-off, we approximate the limiting Gaussian process (7) by

$$W^0(v_i) - \frac{\widehat{g}\{G_n^{-1}(v_i)\}}{\widehat{g}\{G_n^{-1}(p_2)\}} W^0(p_2)$$

$$- \left(\frac{F_m^{-1}(v_i) - F_m^{-1}(p_2)}{F_m^{-1}(p_3) - F_m^{-1}(p_1)} \right) \left[\frac{\widehat{g}\{G_n^{-1}(v_i)\}}{\widehat{g}\{G_n^{-1}(p_3)\}} W^0(p_3) - \frac{\widehat{g}\{G_n^{-1}(v_i)\}}{\widehat{g}\{G_n^{-1}(p_1)\}} W^0(p_1) \right],$$

$\widehat{g}$ being a kernel density estimator and $p_i = i/4$, $i = 1, 2, 3$, and generate sample values of its supremum with parameters chosen as in Section 5.2, to compute the quantile e_α.

Table 5 shows the achieved size and power of the proposed test for varying sample sizes at nominal levels 0.05 and 0.10, for data generated from the five models. Each entry is based on 10,000 replications of the process. Achieved sizes of the test (presented in columns

Table 5. Achieved size and power of the proposed graphical test of the two-sample location-scale hypothesis, at nominal levels 0.05 and 0.10 (based on 10,000 simulation runs)

Level (α)	Sample Size (m, n)	Data generated from				
		Location-scale Model (Size)			Non-location-scale Model (Power)	
		(a)	(b)	(c)	(d)	(e)
0.05	50, 50	0.055	0.075	0.048	0.677	0.523
	100, 100	0.054	0.076	0.048	0.960	0.910
	200, 200	0.052	0.064	0.044	0.999	0.999
	500, 500	0.050	0.050	0.044	1.000	1.000
0.10	50, 50	0.081	0.104	0.072	0.753	0.629
	100, 100	0.092	0.106	0.075	0.976	0.946
	200, 200	0.088	0.101	0.073	1.000	0.999
	500, 500	0.089	0.092	0.075	1.000	1.000

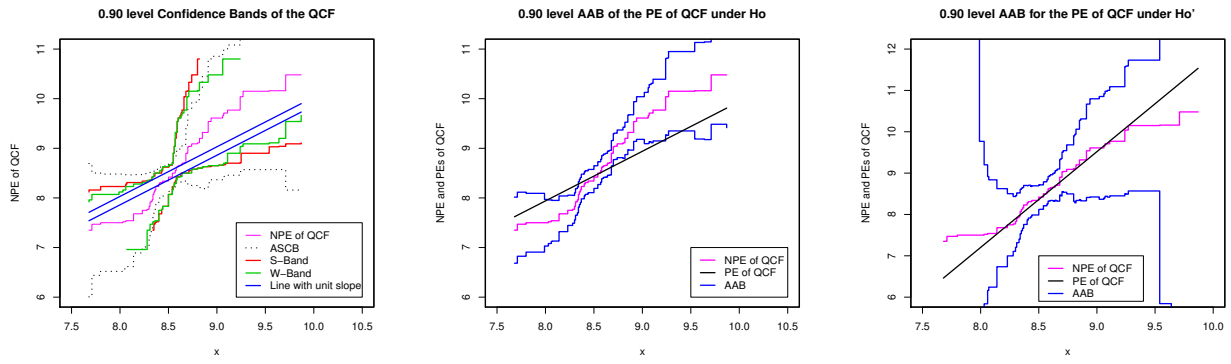
three to five) appear to be in line with the nominal level of significance except for model (b) at sample size 200 or smaller and $\alpha = 0.05$. Values of Power presented in the last two columns of Table 5 are found to be reasonable under the chosen conditions.

6 Data Analytic Illustration

For illustration we choose an astronomical data set collected during the 1761 transit of Venus over Sun. It consists of 158 parallax angles (in seconds of a degree) of the Sun, the angles subtended by the Earth's radius, measured in an experiment from different observatories of the world. Group 1 consists of 95 angles based on a comparison of observations at a single observatory with a long list of others, while Group 2 consists of 63 angles based on

pairwise comparisons of seven observatories with nine others. The two samples correspond to the two groupings of eight data sets as quoted by Stigler (1977). Fisher (1983) found no difference between the two groups. We will re-examine this conclusion.

Figure 2(a) exhibits plot of the nonparametric estimate of the QCF together with its 90% ASCB, 90% S-band and 90% W-band. It also shows the graphs of two extreme straight lines with unit slope and different intercepts, which can be drawn within all the three bands. Evidently, neither test can reject the null hypothesis of location shift at level 0.10, although the nonparametric estimate of the QCF has noticeable deviation from a straight line. Many straight lines, including the one through the origin with unit slope, lies well within the three bands, indicating that these level 0.90 bands fail to detect any difference between F and G. Fisher (1983) used the S-band to conclude that F and G do not differ. On the contrary, the parametric estimate of the QCF under the location shift model does not fit into its 90% acceptance band (see Figure 2(b)), indicating rejection of the null hypothesis of location shift by the proposed test at 0.10 level of significance. A follow-up analysis shows that the p-value of the test is 0.0005 and the estimate of the location parameter is -0.06.



(a) Confidence Bands (b) AAB under H_0 (c) AAB under H_0'

Figure 2: *Sun Data:* (a) Nonparametric estimate (NPE) of the QCF, and its 90% ASCB, 90% S-band and 90% W-band, together with two straight lines with unit slope passing through all the three bands. (b) NPE of the QCF and its parametric estimate (PE) under a location shift model, together with the 90% asymptotic acceptance band (AAB) of the PE. (c) NPE of the QCF and its PE under the location-scale model, together with the 90% AAB of the PE.

A straight line with slope greater than unity may fit the empirical QCF plot except at the two tails. Thus, one needs to check whether the two distributions differ by a location-scale shift. Figure 2(c) exhibits plots of the nonparametric estimate of the QCF and its parametric estimate under the location-scale model, together with the 90% acceptance band for the parametric estimate of the QCF. It is seen that the parametric estimate of the QCF represented by a straight line fits well into its 90% acceptance band. Therefore, the proposed graphical test does not reject the location-scale shift hypothesis at level 0.10 for the data. The p-value of the test turns out to be 0.8805, and the estimates of the location and scale parameters are obtained as -11.32 and 2.32, respectively.

7 Concluding Remarks

The graphical tests considered in this paper combine the descriptive value provided by a plot with the formality of an analytical test. In the proposed approach, a formal test is constructed by considering the difference between nonparametric and parametric estimates of a certain function of the two empirical distributions, and using its limit to construct an acceptance band for the parametric estimate. The acceptance band is a graphical analogue of cut-off values of a test statistic.

The goal of the existing graphical tests for the location shift hypothesis, based on confidence bands, is somewhat different from the purpose of those bands. A confidence band seeks to capture the *true* QCF (a straight line with unit slope and a particular intercept) with a specified coverage probability, while the corresponding graphical test seeks to capture *some* straight line with unit slope and arbitrary intercept, but not necessarily the *true* straight line. This is why the existing graphical tests are too conservative. In contrast, the acceptance band is designed to capture the parametric estimate of the true QCF with a specified asymptotic probability.

This idea of plugging the hole may also be used to improve other graphical tests based

on confidence band. For instance, one may construct better graphical tests for comparing transition probabilities in a Markov Chain model (Dabrowska and Ho, 2000), or for the proportional reversed hazards model in two samples (Sengupta and Nanda, 2011), also known as Lehman alternative in the literature. Besides, the proposed idea can be extended to the two-sample location or location-scale models with censored data (Lu et al., 1994), and also to comparison of several populations in both uncensored and censored situations (Nair, 1982).

Appendix A: Proofs of Theorems

Proof of Theorem 1:

Under H_0, the difference process can be written as

$$\begin{aligned} & \sqrt{M}\left(\widehat{D}_L(x) - D_L(x)\right) \\ = \; & \sqrt{M}\left(G_n^{-1}F_m(x) - G^{-1}F(x)\right) - \sqrt{M}\left(\hat{\theta} - \theta\right) \\ = \; & \sqrt{M}\left(G_n^{-1}F_m(x) - G^{-1}F(x)\right) - \sqrt{M}\left(G_n^{-1}F_m(x^0) - G^{-1}F(x^0)\right), \end{aligned}$$

which is a continuous function of the process $\sqrt{M}\left(G_n^{-1}F_m(x) - G^{-1}F(x)\right)$. Further, Doksum (1974) proved the weak convergence of the process $\sqrt{N}(\widehat{\Delta}(x) - \Delta(x))$, from which it can be easily shown that, under H_0:

$$\sqrt{M}\left(G_n^{-1}F_m(x) - G^{-1}F(x)\right) \xrightarrow{\mathcal{D}} \frac{W^0(F(x))}{g(G^{-1}F(x))}$$

in the space $D[a,\, b]$, where $\xrightarrow{\mathcal{D}}$ denotes weak convergence. Therefore, an application of the continuous mapping theorem completes the proof. □

Proof of Theorem 2:

First, we prove the uniform weak consistency of $\widehat{g}(x+\hat{\theta})$ on the compact set $[a,\, b]$.

By the triangle inequality, it is easy to write that

$$\begin{aligned}\sup_{x\in[a,\,b]}|\widehat{g}(x+\hat{\theta})-g(x+\theta)| &\leq \sup_{x\in[a,\,b]}|\widehat{g}(x+\hat{\theta})-\widehat{g}(x+\theta)|+\sup_{x\in[a,\,b]}|\widehat{g}(x+\theta)-g(x+\theta)| \\ &= \sup_{y\in[c,\,d]}|\widehat{g}(y+\hat{\theta}-\theta)-\widehat{g}(y)|+\sup_{y\in[c,\,d]}|\widehat{g}(y)-g(y)|,\ \text{under } H_0.\end{aligned}$$

Then, under H_0, for any $\epsilon>0$,

$$\begin{aligned}&P\left[\sup_{x\in[a,\,b]}|\widehat{g}(x+\hat{\theta})-g(x+\theta)|<\epsilon\right] \\ \geq\ & P\left[\left\{\sup_{y\in[c,\,d]}|\widehat{g}(y+\hat{\theta}-\theta)-\widehat{g}(y)|+\sup_{y\in[c,\,d]}|\widehat{g}(y)-g(y)|\right\}<\epsilon\right] \\ \geq\ & P\left[\sup_{y\in[c,\,d]}|\widehat{g}(y+\hat{\theta}-\theta)-\widehat{g}(y)|<\epsilon/2\cap\sup_{y\in[c,\,d]}|\widehat{g}(y)-g(y)|<\epsilon/2\right] \\ \geq\ & P\left[\sup_{y\in[c,\,d]}|\widehat{g}(y+\hat{\theta}-\theta)-\widehat{g}(y)|<\epsilon/2\right]+P\left[\sup_{y\in[c,\,d]}|\widehat{g}(y)-g(y)|<\epsilon/2\right]-1. \quad \text{(A.1)}\end{aligned}$$

By uniform weak consistency of $\widehat{g}$ which follows from its pointwise convergence in probability and uniform equicontinuity in probability, for every $\epsilon/2>0$ and $\eta/2>0$, there exists an $n_1(\epsilon,\,\eta)\in\mathbb{N}$ such that

$$P\left[\sup_{y\in[c,\,d]}|\widehat{g}(y)-g(y)|<\epsilon/2\right]>1-\eta/2,\qquad \forall n\geq n_1(\epsilon,\,\eta). \quad \text{(A.2)}$$

Further, by uniform stochastic equicontinuity (in probability) of $\widehat{g}$, for every $\epsilon/2>0$ and $\eta/2>0$, there exist a $\delta(\epsilon,\,\eta)>0$ and an $n_2(\epsilon,\,\eta)\in\mathbb{N}$ such that

$$P\left[\sup_{u,\,v\in[c,\,d]:\ |u-v|<\delta}|\widehat{g}(u)-\widehat{g}(v)|<\epsilon/2\right]>1-\eta/2,\qquad \forall n\geq n_2(\epsilon,\,\eta). \quad \text{(A.3)}$$

Moreover, since $\hat{\theta}$ is a consistent estimator of θ, for any given $\delta>0$ and $\eta>0$, there exists

an $n_3(\delta,\, \eta) \in \mathbb{N}$ such that

$$P\left[|\hat{\theta} - \theta| < \delta\right] > 1 - \eta, \qquad \forall m, n \geq n_3(\delta,\, \eta). \tag{A.4}$$

Then, for given any $\epsilon/2 > 0$ and $\delta > 0$, however small,

$$\begin{aligned}
& P\left[\sup_{y \in [c,\, d]} |\widehat{g}(y + \hat{\theta} - \theta) - \widehat{g}(y)| < \epsilon/2\right] \\
\geq\ & P\left[\sup_{y \in [c,\, d]} |\widehat{g}(y + \hat{\theta} - \theta) - \widehat{g}(y)| < \epsilon/2 \cap \{y + \hat{\theta} - \theta\} \in [c,\, d]\right] \\
\geq\ & P\left[\sup_{y \in [c,\, d]} |\widehat{g}(y + \hat{\theta} - \theta) - \widehat{g}(y)| < \epsilon/2 \cap \{y + \hat{\theta} - \theta\} \in [c,\, d] \cap |y + \overline{\hat{\theta} - \theta} - y| < \delta\right] \\
\geq\ & P\left[\sup_{u,\, v \in [c,\, d]:\ |u - v| < \delta} |\widehat{g}(u) - \widehat{g}(v)| < \epsilon/2\right] \\
>\ & 1 - \eta/2, \qquad \forall m \geq n_3(\delta,\, \eta) \text{ and } n \geq n_2(\epsilon,\, \eta) \vee n_3(\delta,\, \eta),
\end{aligned} \tag{A.5}$$

where the last inequality follows from the stochastic equicontinuity condition (A.3).

Therefore, (A.1) together with (A.2) and (A.5) implies that for every $\epsilon > 0$ and $\eta > 0$, there exists an $n_0(\epsilon,\, \eta) \in \mathbb{N}$ such that

$$P\left[\sup_{x \in [a,\, b]} |\widehat{g}(x + \hat{\theta}) - g(x + \theta)| < \epsilon\right] > 1 - \eta, \qquad \forall m, n \geq n_0(\epsilon,\, \eta).$$

which establishes the uniform weak consistency of $\widehat{g}(x + \hat{\theta})$ on $[a,\, b]$.

Finally, the proof of the convergence of the weighted supremum norm statistic in the theorem follows from Theorem 1 and the uniform weak consistency of $\widehat{g}(x + \hat{\theta})$, and from an application of the continuous mapping theorem. □

Proof of Theorem 3:

Under H_0', the difference process $\sqrt{M}\widehat{D}_{LS}(x)$ can be written as $\sqrt{M}(G_n^{-1}F_m(x) - G^{-1}F(x)) - x\sqrt{M}(\hat{\delta} - \delta) - \sqrt{M}(\hat{\theta} - \theta)$. Now, under H_0', $\sqrt{M}(\hat{\theta} - \theta)$ simplifies to $\sqrt{M}(G_n^{-1}F_m(x_2) - G^{-1}F(x_2)) - x_2\sqrt{M}(\hat{\delta} - \delta)$. Therefore, the difference process reduces to

$$\begin{aligned}\sqrt{M}\left(G_n^{-1}F_m(x) - G^{-1}F(x)\right) - (x - x_2)\sqrt{M}\left(\hat{\delta} - \delta\right) \\ -\sqrt{M}\left(G_n^{-1}F_m(x_2) - G^{-1}F(x_2)\right).\end{aligned} \tag{A.6}$$

Further, under H_0', $\sqrt{M}(\hat{\delta} - \delta)$ can be expressed as

$$\frac{\sqrt{M}\left(G_n^{-1}F_m(x_3) - G^{-1}F(x_3)\right) - \sqrt{M}\left(G_n^{-1}F_m(x_1) - G^{-1}F(x_1)\right)}{x_3 - x_1}.$$

So, the difference process, under H_0', becomes

$$\begin{aligned}&\sqrt{M}\left(G_n^{-1}F_m(x) - G^{-1}F(x)\right) - \sqrt{M}\left(G_n^{-1}F_m(x_2) - G^{-1}F(x_2)\right) \\ -\left(\frac{x - x_2}{x_3 - x_1}\right)&\left[\sqrt{M}\left(G_n^{-1}F_m(x_3) - G^{-1}F(x_3)\right) - \sqrt{M}\left(G_n^{-1}F_m(x_1) - G^{-1}F(x_1)\right)\right],\end{aligned} \tag{A.7}$$

which is a continuous function of the process $\sqrt{M}\left(G_n^{-1}F_m(x) - G^{-1}F(x)\right)$. Now the process $\sqrt{M}\left(G_n^{-1}F_m(x) - G^{-1}F(x)\right)$ converges weakly to a mean zero Gaussian process $W^0(F(x))/g(G^{-1}F(x))$ in the space $D[a, b]$, which follows from Theorem 4.1 of Doksum (1974). Therefore, the result of the theorem immediately follows from an application of the continuous mapping theorem. □

Proof of Theorem 4:

Proof is analogous to the proof of Theorem 2.

By the triangle inequality,

$$\begin{aligned} & \sup_{x\in[a,\,b]} |\widehat{g}(\hat{\delta}x+\hat{\theta}) - g(\delta x+\theta)| \\ \leq & \sup_{x\in[a,\,b]} |\widehat{g}(\hat{\delta}x+\hat{\theta}) - \widehat{g}(\delta x+\theta)| + \sup_{x\in[a,\,b]} |\widehat{g}(\delta x+\theta) - g(\delta x+\theta)| \\ = & \sup_{y\in[c,\,d]} |\widehat{g}(y+x(\hat{\delta}-\delta)+(\hat{\theta}-\theta)) - \widehat{g}(y)| + \sup_{y\in[c,\,d]} |\widehat{g}(y) - g(y)|, \text{ under } H_0', \end{aligned}$$

which implies that for any $\epsilon > 0$, under H_0',

$$\begin{aligned} & P\left[\sup_{x\in[a,\,b]} |\widehat{g}(\hat{\delta}x+\hat{\theta}) - g(\delta x+\theta)| < \epsilon\right] \\ \geq & P\left[\left\{\sup_{y\in[c,\,d]} |\widehat{g}(y+x(\hat{\delta}-\delta)+(\hat{\theta}-\theta)) - \widehat{g}(y)| + \sup_{y\in[c,\,d]} |\widehat{g}(y) - g(y)|\right\} < \epsilon\right] \\ \geq & P\left[\sup_{y\in[c,\,d]} |\widehat{g}(y+x(\hat{\delta}-\delta)+(\hat{\theta}-\theta)) - \widehat{g}(y)| < \epsilon/2 \cap \sup_{y\in[c,\,d]} |\widehat{g}(y) - g(y)| < \epsilon/2\right] \\ \geq & P\left[\sup_{y\in[c,\,d]} |\widehat{g}(y+x(\hat{\delta}-\delta)+(\hat{\theta}-\theta)) - \widehat{g}(y)| < \epsilon/2\right] + P\left[\sup_{y\in[c,\,d]} |\widehat{g}(y) - g(y)| < \epsilon/2\right] - 1. \end{aligned} \tag{A.8}$$

Now, pointwise weak consistency of $\widehat{g}$ along with its uniform equicontinuity in probability implies the uniform weak consistency on $[c,\, d]$. Therefore, for every $\epsilon/2 > 0$ and $\eta/2 > 0$, there exists an $n_1(\epsilon,\, \eta) \in \mathbb{N}$ such that

$$P\left[\sup_{y\in[c,\,d]} |\widehat{g}(y) - g(y)| < \epsilon/2\right] > 1 - \eta/2, \qquad \forall n \geq n_1(\epsilon,\, \eta). \tag{A.9}$$

Again, by uniform equicontinuity in probability of $\widehat{g}$, for every $\epsilon/2 > 0$ and $\eta/2 > 0$, there exist a $\zeta(\epsilon,\, \eta) > 0$ and an $n_2(\epsilon,\, \eta) \in \mathbb{N}$ such that

$$P\left[\sup_{u,\,v\in[c,\,d]:\ |u-v|<\zeta} |\widehat{g}(u) - \widehat{g}(v)| < \epsilon/2\right] > 1 - \eta/2, \qquad \forall n \geq n_2(\epsilon,\, \eta). \tag{A.10}$$

Further, consistency of $\hat{\delta}$ and $\hat{\theta}$ implies that for any given $\zeta > 0$ and $\eta > 0$, there exists an $n_3(\zeta, \eta) \in \mathbb{N}$ such that for any fixed x

$$P\left[|x(\hat{\delta} - \delta) + (\hat{\theta} - \theta)| < \zeta\right] > 1 - \eta, \qquad \forall m, n \geq n_3(\zeta, \eta). \tag{A.11}$$

Then, for given any $\epsilon/2 > 0$ and $\zeta > 0$, it follows from (A.10) and (A.11) that,

$$\begin{aligned}
& P\left[\sup_{y\in[c,\,d]} |\widehat{g}(y + x(\hat{\delta} - \delta) + (\hat{\theta} - \theta)) - \widehat{g}(y)| < \epsilon/2\right] \\
\geq\ & P\left[\sup_{y\in[c,\,d]} |\widehat{g}(y + x(\hat{\delta} - \delta) + (\hat{\theta} - \theta)) - \widehat{g}(y)| < \epsilon/2 \cap \{y + x(\hat{\delta} - \delta) + (\hat{\theta} - \theta)\} \in [c,\, d]\right] \\
\geq\ & P\left[\sup_{y\in[c,\,d]} |\widehat{g}(y + x(\hat{\delta} - \delta) + (\hat{\theta} - \theta)) - \widehat{g}(y)| < \epsilon/2 \cap \{y + x(\hat{\delta} - \delta) + (\hat{\theta} - \theta)\} \in [c,\, d]\right. \\
& \qquad\qquad \left. \cap |y + x(\hat{\delta} - \delta) + (\hat{\theta} - \theta) - y| < \zeta\right] \\
\geq\ & P\left[\sup_{u,\,v\in[c,\,d]:\ |u-v|<\zeta} |\widehat{g}(u) - \widehat{g}(v)| < \epsilon/2\right] \\
>\ & 1 - \eta/2, \qquad \forall m \geq n_3(\zeta, \eta) \text{ and } n \geq n_2(\epsilon, \eta) \vee n_3(\zeta, \eta).
\end{aligned} \tag{A.12}$$

Therefore, (A.8), (A.9) and (A.12) together implies the uniform weak consistency of the estimator $\widehat{g}(\hat{\delta}x + \hat{\theta})$ of the function $g(\delta x + \theta)$ on $[a, b]$.

Then, the convergence result of the theorem follows from Theorem 3, the uniform weak consistency of $\widehat{g}(\hat{\delta}x + \hat{\theta})$, and from an application of the continuous mapping theorem. □

Appendix B: An Illustration of Efficiency of Acceptance Band

A graphical illustration of the efficiency results based on simulated data generated from a location-scale model is presented in Figure B.1 . We choose $m = n = 100$ and $\alpha = 0.05$. We generate X and Y-samples from $N(0,\, 1)$ and $N(1,\, 2)$ distributions, respectively. Figure B.1 exhibits the plot of the empirical QCF together with its 95% ASCB, 95% S-band and 95% W-band. It also shows the graph of the parametric estimate of the QCF under the location

shift model, represented by the straight line, along with its 95% asymptotic acceptance band. The acceptance band is found narrower than the W-band except near the tails, as expected asymptotically. As far as the testing is concerned, a straight line with unit slope and arbitrary intercept can easily pass through all the confidence bands, whereas the parametric estimate does not fit into its asymptotic acceptance band. This illustrates the less conservative nature of the proposed graphical test compared to the existing ones.

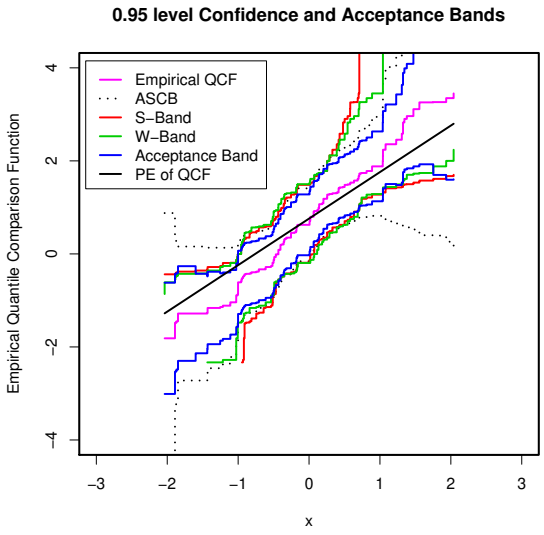


Figure B.1: Empirical QCF plot for data simulated from location-scale model, together with its 95% ASCB, 95% S-band and 95% W-band, and the plot of the parametric estimate (PE) of the QCF under location shift model with its 95% asymptotic acceptance band.

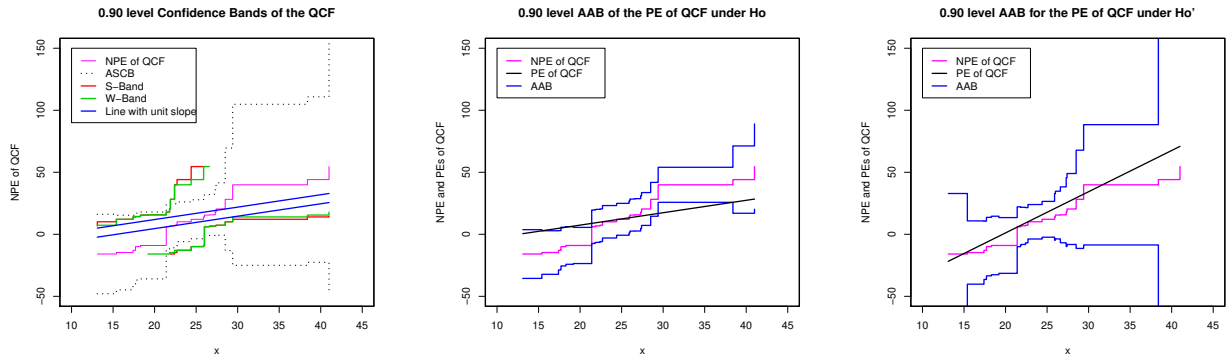
Appendix C: An Additional Data Analytic Illustration

We illustrate the proposed graphical methods through analysis of another real data set which consists of weight gains (in grams) of 45 seventy-day-old rats in an experiment conducted at California Primate Research Center, University of California, Davis, for investigating the undesirable effects of ozone on weight gain in rats. The control group (Group 1) consists of 23 rats kept in an ozone-free environment for seven days; while the treatment group (Group 2) consists of 22 rats kept in an ozone environment for seven days.

We denote the weight gains of rats in the control group and the treatment group by x and y, respectively. The data set is given in Doksum and Sievers (1976) with its analysis.

At the outset, we delete the outlier $x = -16.9$ from the analysis as identified by Doksum and Sievers (1976). Figures C.1, C.2 and C.3 exhibit plots similar to those in Figures 2(a), 2(b) and 2(c) of the paper, respectively. Two extreme straight lines with unit slope and different intercepts in Figure C.1 indicate that each of the three existing graphical tests is unable to reject the location shift hypothesis at level 0.10. The fact that the 90% S-band is unable to reject the null hypothesis had been reported by Doksum and Sievers (1976, p. 433), who also noted that the nonparametric estimate of the QCF "strongly indicates that it (the location shift model) does not hold". Moreover, the 90% W-band, even though found narrower than the corresponding S-band, can not reject the location shift hypothesis at 10% level of significance. The W-band is constructed using the simulated critical value obtained for the given combination of sample sizes as described in Canner(1975). In Figure C.2, the parametric estimate of the QCF under the location shift model is not contained within its 90% asymptotic acceptance band. Thus, the null hypothesis of the location shift is rejected by the proposed graphical test at level 0.1. Further analysis of the data provides p-value of the test as 0.0001 and the estimate of the location parameter as -12.60.

It transpires from the above analysis that the Q-Q plot of Group 2 against Group 1 is somewhat linear, though the slope is clearly different from one. Thus, it would be interesting to check whether the location-scale model holds. In Figure C.3, the parametric estimate of the QCF under the location-scale model is found to have been completely contained within its 90% acceptance band. Therefore, the proposed graphical test designed for checking the location-scale relationship between the two groups accepts the hypothesis of the location-scale model, at 10% level of significance. Subsequent analysis of the data set shows the estimates of the location and scale parameters as -65.29 and 3.32, respectively, and the p-value of the test as 0.9925.



(C.1) Confidence Bands (C.2) AAB under H_0 (C.3) AAB under H_0'

Figure: *Ozone Data:* (C.1) Nonparametric estimate (NPE) of the QCF, and its 90% ASCB, 90% S-band and 90% W-band, together with two straight lines with unit slope passing through all the three bands. (C.2) NPE of the QCF and its parametric estimate (PE) under a location shift model, together with the 90% asymptotic acceptance band (AAB) of the PE. (C.3) NPE of the QCF and its PE under the location-scale model, together with the 90% AAB of the PE.

References

Atkinson, A. C. (1985). *Plots, Transformations and Regression.* Oxford: Oxford University Press.

Canner, P. L. (1975). A Simulation Study of One- and Two-Sample Kolmogorov-Smirnov Statistics With a Particular Weight Function. *Journal of American Statistical Association*, 70, 209-211.

Dabrowska, D., Doksum, K., and Song, J.-K. (1989). Graphical Comparison of Cumulative Hazards for Two Populations. *Biometrika*, 76, 763–773.

Dabrowska, D., and Ho, W. (2000). Confidence Bands for Comparison of Transition Probabilities in a Markov Chain Model. *Lifetime Data Analysis*, 6, 5–21.

Doksum, K. A. (1974). Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case. *The Annals of Statistics*, 2, 267–277.

Doksum, K. A. (1977). Some Graphical Methods in Statistics. A Review and Some Extensions. *Statistica Neerlandica*, 31, 53–68.

Doksum, K. A., and Sievers, G. L. (1976). Plotting With Confidence: Graphical Comparisons of Two Populations. *Biometrika*, 63, 421–434.

Fisher, N. I. (1983). Graphical Methods in Nonparametric Statistics: A Review and Annotated Bibliography. *International Statistical Review*, 51, 25–58.

Klein, J. P., and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data, Springer Series in Statistics for Biology and Health.* New York: Springer.

Lu, H. H. S., Wells, M. T., and Tiwari, R. C. (1994). Inference for Shift Functions in the Two-Sample Problem with Right-Censored Data: With Applications. *Journal of the American Statistical Association*, 89, 1017–1026.

Nair, V. N. (1982). Q-Q Plots with Confidence Bands for Comparing Several Populations. *Scandinavian Journal of Statistics*, 9, 193–200.

Pearson, E. S., and Hartley, H. O. (1972). *Biometrika Tables for Statisticians, Vol. 2.*. Cambridge University Press.

Sahoo, S., and Sengupta, D. (2016). On Graphical Tests for Proportionality of Hazards in Two Samples. *Statistics in Medicine*, 35, 942–956.

Sengupta, D., and Nanda, A. K. (2011). The Proportional Reversed Hazards Regression Model. *Journal of Statistical Theory and Applications*, 18, 461–476.

Shorack, G. R., and Wellner, J. (1986). *Empirical Processes with Applications to Statistics.* New York: Wiley.

Silverman, B. W. (1998). *Density Estimation for Statistics and Data Analysis*. Boca Raton: Chapman & Hall/CRC.

Stigler, S. M. (1977). Do Robust Estimators Work with Real Data?. *The Annals of Statistics*, 5, 1055–1098.

Switzer, P. (1976). Confidence Procedures for Two-Sample Problems. *Biometrika*, 63, 13–25.

Thas, O. (2009). *Comparing Distributions*. Belgium: Springer.

Wilk, M. B., and Gnanadesikan, R. (1968). Probability Plotting Methods for the Analysis of Data. *Biometrika*, 55, 1–17.