Improved Two-Sample Comparisons for Bounded Data

Jianning Kong  Jason Parker
Shandong University  Michigan State University
Donggyu Sul
University of Texas at Dallas

Abstract

The goal of this paper is to establish new testing procedures for comparing two independent and bounded samples. The paper begins with a brief literature review of the best traditional tests available. These tests are then applied to two canonical examples of bounded data in public good games, revealing the problems of these traditional methods. Next, a new test for equal distributions is constructed using the sample moments, since the moments determine a bounded distribution up to uniqueness. When the underlying distributions of the two samples are beta, testing for equal first and second moments becomes equivalent to testing for equal distributions. This paper shows that the bootstrapped one-sample Kolmogorov-Smirnov test can be used for testing the null that the sample is drawn from the beta distribution. Next, a new central tendency test is developed using skewness. These newly-developed methods are found to perform well in the finite sample through Monte Carlo simulation and through application to the canonical examples.

Key Words: Distributional Comparisons, Beta Distribution, Bounded Data, Skewness, Rank Sum Test.

JEL Classification Numbers: C12, C16, C23
1 Introduction

This paper addresses the comparison of two independent and bounded samples. The latter restriction is a new addition to this literature that sharpens the statistical comparison, allowing researchers to test more accurately. In many situations, the data can be normalized to be bounded between zero and one. For example, in public good games and ultimatum games, the experimental data are all bounded on the unit interval. Auction data are also usually bounded in this way.

When data are bounded between zero and one, we will demonstrate that skewness can be another measure of central tendency since the skewness combines information from the mean and median. It is easy to show that the mean and median must be equal to 0.5 when bounded samples are symmetric. Otherwise, the distribution becomes asymmetric, and the mean may not be a good measure for central tendency. Meanwhile, the median may be a better measure for central tendency with an extremely asymmetrically distributed and unbounded sample, but with a bounded sample, the skewness can be an alternative measure for central tendency since by definition, skewness measures where the mass of a distribution is located.

Another interesting question is whether or not the underlying distribution of a bounded sample is a beta distribution. This question is interesting for two reasons. First, theoretical economists can use this information when they analyze the behavior of the subjects in experimental games. Many empirical and theoretical papers have used the beta distribution for their analyses (e.g., Bajari and Hortacşu, 2005; McKelvey and Palfrey, 1992). Second, two-sample comparisons are much more straightforward when the two samples are drawn from beta distributions. By estimating the two underlying parameters, $\alpha$ and $\beta$, various test statistics can be constructed and their tests performed. Since additional information about the distribution is used, these newly-designed methods have better power in the finite sample.

There are several goodness-of-fit tests available for testing whether or not a random variable is generated from a reference distribution. Among them, the Kolmogorov-Smirnov (KS hereafter) nonparametric test has been commonly used. When the underlying parameters are estimated, the critical value for the KS test is generally unknown and requires a bootstrap procedure. (See Babu and Rao, 2004; Meintanis and Swanepoel, 2007). We will show that
many laboratory experimental data sets are well characterized by a simple beta distribution.

Under the assumption that the underlying distributions of the two samples are beta, the
null hypothesis of equal distributions can be re-formulated by using the two beta parameters, \( \alpha \) and \( \beta \). Once these two parameters are estimated, each of the moments can also be estimated
since all of the moments are functions of \( \alpha \) and \( \beta \). The equality of means, medians, and
skewness tests are developed based on the estimates of these \( \alpha \) and \( \beta \), and their finite sample
performance is examined. For the case when the two samples are not beta distributed, we also
derive the limiting distributions of the proposed tests for equal distributions and skewness.

The rest of the paper consists of four sections. The next section provides a short literature
review of the two-sample comparison literature, particularly focusing on the Wilcoxon-Mann-
Whitney (1945,1947, WMW hereafter) rank sum test, which has been used to test for equal
distributions. Section 3 provides new test statistics and testing procedures for the comparison
of two samples under the boundedness condition; it also introduces the beta distribution.
In Section 4, the new tests are extended parametrically via the beta distribution to further
increase the power of the tests. Section 5 examines the finite sample properties of the new
tests by means of Monte Carlo simulation and provides an empirical demonstration. Section
6 concludes. Technical derivations and proofs are provided in appendices.

2 Preliminary and Canonical Empirical Examples

Here, we briefly discuss the statistical properties of the tests which are most popularly used in
practice. Then we demonstrate the problems of these tests by using real empirical examples.

2.1 Preliminary

We consider two continuous independent samples \( X = \{x_1, \ldots, x_m\} \) and \( Y = \{y_1, \ldots, y_n\} \)
where \( X \sim i.i.d. Q_X \) and \( Y \sim i.i.d. Q_Y \). Denote \( N = m + n \). The most popularly used tests
are the standard z-score, the WMW, and the KS tests. Here, we briefly review the latter
two tests.

The one-sample KS test is designed for detecting whether or not the sample of interest is
generated from a particular distribution. Let the empirical cumulative distribution function
be \( \hat{F}_n \) and its true cumulative distribution function (CDF) be \( F \). Then, the one-sample KS
The KS statistic, \( \hat{d}_n \), is defined as
\[
\hat{d}_n = \sup_x \left| \hat{F}_n(x) - F(x) \right|.
\]
Under the null hypothesis, the KS statistic, \( \hat{d} \), converges to zero in probability as \( n \to \infty \), and \( \sqrt{n} \hat{d} \) converges to the Kolmogorov distribution which does not depend on the function \( F \). However, in practice, the CDF of \( F(x) \) is usually evaluated with the estimated parameters. For example, in the case of the normal distribution, the mean and variance should be estimated from the sample. In this case, the critical value of \( \hat{d}_n \) must be obtained either by a numerical approximation or by bootstrap. We will use this one-sample KS test to detect whether or not the random samples are each generated from a beta distribution later. The power of the one-sample KS test is much better than that of the two-sample KS test, which we consider below.

The two-sample KS test is designed for testing whether or not the two samples are generated from the same probability density function. The two-sample KS statistic, \( \hat{d}_{n,m} \), is defined as
\[
\hat{d}_{n,m} = \sup_x \left| \hat{F}_n(x) - \hat{F}_m(x) \right|.
\]
The asymptotic critical value can be obtained by approximation to the beta distribution. See Zhang and Wu (2002) for a more detailed discussion. In practice, the two-sample KS test has not been used popularly due to its lack of power. See Marsh (2010) for a finite sample comparison of various tests.

In practice, empirical researchers have used the WMW’s rank sum test for testing the equality of two distributions. In fact, Mann and Whitney (1947) clearly states that the null hypothesis of the Wilcoxon (1945) test is the equality of two distributions. However, Chung and Romano (2013) recently showed that the null hypothesis of the WMW test is not equal distributions; instead the null should be written as
\[
H^w_0 : \Pr(X \leq Y) = 1/2.
\]
Note that \( H^w_0 \) can be interpreted as the median of the difference of the two samples is zero. Meanwhile, the null hypothesis of equal distributions is given by
\[
H^Q_0 : Q_X = Q_Y.
\]
Obviously, if \( H_0^Q \) is true, then \( H_0^w \) is true. However, the opposite is not true. Chung and
Romano (2013) pinpoint that

\[ H_0^w : \Pr (X \leq Y) = 1/2 \quad \leftrightarrow \quad H_0^Q : Q_X = Q_Y. \] (3)

Chung and Romano’s interpretation of \( H_0^w \) is statistically correct. Moreover Chung and Romano (2013) states that \( H_0^w \) can be rejected not because \( \Pr (X \leq Y) > 1/2 \) but because \( Q_X \neq Q_Y \).

Chung and Romano (2013)’s modified WMW test is given by

\[ S_w = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I \{x_i < y_j\} - (1/2) \]

\[ \frac{1}{\sqrt{\xi_x/m - \xi_y/n}}, \] (4)

where \( I \{\cdot\} \) is the indicator function such that \( I \{x_i < y_i\} = 1 \) if \( x_i < y_j \) and \( I \{x_i < y_i\} = 0 \) otherwise. \( \hat{\xi}_x \) and \( \hat{\xi}_y \) are defined as

\[ \hat{\xi}_x = \frac{1}{m-1} \sum_{i=1}^{m} \left( \frac{1}{n} \sum_{j=1}^{n} I \{y_j \leq x_i\} - \frac{1}{m} \sum_{i=1}^{m} \left[ \frac{1}{n} \sum_{j=1}^{n} I \{y_j \leq x_i\} \right] \right)^2, \]

\[ \hat{\xi}_y = \frac{1}{n-1} \sum_{j=1}^{n} \left( \frac{1}{m} \sum_{i=1}^{m} I \{x_i < y_j\} - \frac{1}{n} \sum_{j=1}^{n} \left[ \frac{1}{m} \sum_{i=1}^{m} I \{x_i < y_j\} \right] \right)^2. \]

Even though the WMW test cannot be used to test the equality of two distributions, it can detect the relative location difference between the two distributions effectively, as Lehmann (2009) pointed out. The finding by Chung and Romano (2013) becomes an issue for those who want to test for equal distributions in (2). For testing for equal distributions, one can use the two-sample KS test even though the power of the KS test is weak.

### 2.2 Canonical Empirical Example

The canonical example uses data from Nikiforakis and Normann (2008, NN hereafter) and Gächter, Renner, Sefton (2008, GRS hereafter), which are two of the most well-known and exemplary articles examining the effectiveness of punishment in public good games. Both experiments are based on a traditional repeated public good game. The individuals may punish the other group members by a punishment degree \( p \). The increase of degree \( p \) makes the punishment threat to free riders more damaging and thus more credible, and it may induce greater contribution, which is the measured quantity for each individual. In NN’s
study, $p$ varies on a scale from 0 to 4 (the corresponding experiments are denoted as $p0$, $p1$, $p2$, $p3$, $p4$), where 0 means there’s no punishment opportunity in the game. There are 24 individuals in each of these five experiments. In GRS’s design, there are four experiments with different $p$ and different numbers of rounds. For consistency, we only included the two experiments with the same number of rounds (10) as in NN’s, and we use $p0$ and $p3$ to denote these. Here, $p0$ has 60 individuals, and $p3$ has 45. One important goal of NN’s and GRS’s experiments is to test if the increase of $p$ has an impact on overall contribution.

We report the WMW test statistics, two-sample $t$-test statistics and the p-value of the KS test, sample means ($\hat{\mu}$), medians ($\hat{\mu}_m$), and standard deviation ($\hat{\sigma}$) for the time series-averaged data in Table 1. The data are normalized to lie between 0 and 1.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Statistics</th>
<th>P-value</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td>WMW</td>
<td>z-score</td>
</tr>
<tr>
<td>$p0$</td>
<td>$p1$</td>
<td>-1.640</td>
<td>-2.444</td>
</tr>
<tr>
<td>$p1$</td>
<td>$p2$</td>
<td>-2.238</td>
<td>-2.179</td>
</tr>
<tr>
<td>$p2$</td>
<td>$p3$</td>
<td>-2.135</td>
<td>-2.954</td>
</tr>
<tr>
<td>$p3$</td>
<td>$p4$</td>
<td>-1.734</td>
<td>-1.289</td>
</tr>
<tr>
<td>GRS</td>
<td>$p0$</td>
<td>-1.647</td>
<td>-3.473</td>
</tr>
<tr>
<td></td>
<td>$p3$</td>
<td>-3.157</td>
<td>-3.473</td>
</tr>
</tbody>
</table>

The results in Table 1 provide interesting information to statisticians as to why the WMW test has been popularly used. In column 4, the WMW test rejects the null hypothesis of $H_0^w: \Pr(X \leq Y) = 1/2$ three out of five times at the 5% level, and rejects four times at the 10% level. As we pointed out before, the rejection of the WMW test is not a rejection of equal central tendency. The empirical explanation of the rejection of the WMW test is ambiguous.

Meanwhile, the z-score from the standard $t$-test rejects the null of equal means four out of five times for both the one-sided and two-sided tests at the 5% level. However, for a bounded sample, the sample mean is not a good measure of central tendency. Besides, the non-rejection of the null of equal means does not imply that the two samples share the same distribution. In other words, the standard $t$-test is not effective for detecting treatment
effects in the second-order or higher-order.

The sixth column shows the p-value of the KS test, where the null hypothesis is distributional equality. Surprisingly the KS test only rejects the null hypothesis two out of five times even at the 10% significance level. Even when the sample means are significantly different, the KS test does not reject the null that the distributions are the same. In other words, the power of the KS test is very weak. That’s why the KS test is not used very often in practice.

3 Practical Comparison of Two Samples

As demonstrated above, the power of the KS test is relatively weak. There are variants of the KS test such as those by Cramér-von Mises, Watson (1961) and Kuiper (1960) tests. However, all these variants are based on the comparison of the two cumulative distribution functions. If the two probability density functions (PDF) are the same, naturally the CDFs are also the same. Two-sample comparisons based on equal PDFs are statistically elegant but lack power. Here, we consider somewhat different statistical testing procedures.

3.1 Comparison of Moments

It is well-known that if two CDFs are the same, then all of the moments of the two samples must be identical. The opposite is true also as long as all moments exist and are finite. When the data are bounded this condition is clearly met, so it is possible to construct more powerful tests using the moments. Denote \( \mu_{ik} \) as the \( i \)th central moment of the \( k \)th sample for \( i > 1 \) and \( \mu_k \) as the first noncentral moment. Then the null \( \mathcal{H}_0^m \) becomes identical to \( \mathcal{H}_0^Q \) as long as \( \mu_{ik} \) exists.

\[ \mathcal{H}_0^m : \mu_{iX} = \mu_{iY} \text{ for all } i \iff \mathcal{H}_0^Q : Q_X = Q_Y \]

Of course, testing \( \mathcal{H}_0^m \) is not feasible at all. Instead, we suggest testing a subset of \( \mathcal{H}_0^m \). That is, we are interested in testing the following subsets of \( \mathcal{H}_0^m \).

\[ \mathcal{H}_0^\lambda : \mu_{iX} = \mu_{iY} \text{ for all } i \leq \lambda. \]

When \( \lambda = 1 \), the null of \( \mathcal{H}_0^\lambda \) becomes the null of the z-score test, \( \mathcal{H}_0^z \). When \( \lambda = 3 \), the null of \( \mathcal{H}_0^3 \) implies that the two samples share the same mean, variance and skewness. If the null hypothesis with \( \lambda \) moments is rejected, then the treated sample differs from the control.
sample ‘up to the \( \lambda \)th-order.’ For example, if the null with \( \lambda = 2 \) is rejected, we say it that there is a treatment effect up to the second-order.

This ‘subset approach’ is convenient but at the same time raises an important question: How should one choose a small value of \( \lambda \)? Of course, first, the answer depends on the research question. For example, if applied researchers care about the location of the distribution, then they should choose \( \lambda = 2 \) rather than \( \lambda = 1 \). There are different measures of statistical location: mean, median, mode and quantile mean. However, it is not rare to see that the median of \( X \) is greater than the median of \( Y \) even when the mean of \( X \) is smaller than the mean of \( Y \). When the two distributions are not symmetric, any such disagreement between differences in means and medians becomes an issue. In this case, the skewness becomes useful. Let \( m_k \) be the median of the \( k \)th variable. Then, Pearson’s nonparametric skewness is defined as

\[
S_{k, \text{non}} = 3 \frac{\bar{\mu}_k - m_k}{\sqrt{\mu_{2k}}},
\]

where \( \mu_k \) is the first noncentral moment. Since the median can be thought of as the weighted mean, the nonparametric skewness can be thought of as a function of the first and second moments. We will show later that this test with \( \lambda = 2 \) becomes much more powerful than the KS test.

The test statistic with \( \lambda = 2 \) is straightforward. The following statistic has the limiting distribution of \( \chi^2 \) with 2 degrees of freedom as \( n \) and \( m \to \infty \)

\[
\mathcal{M}_2 = [\hat{\mu}_X - \hat{\mu}_Y, \hat{\mu}_{2X} - \hat{\mu}_{2Y}] (\hat{\Xi}_X / m + \hat{\Xi}_Y / n)^{-1} [\hat{\mu}_X - \hat{\mu}_Y, \hat{\mu}_{2X} - \hat{\mu}_{2Y}]' \to^d \chi^2_2, \tag{5}
\]

where

\[
\hat{\Xi}_k = \begin{bmatrix} \hat{\mu}_{2k} & \hat{\mu}_{3k} \\ \hat{\mu}_{3k} & \hat{\mu}_{4k} - (\hat{\mu}_{2k})^2 \end{bmatrix}, \text{ for } k = X, Y,
\]

and the circumflex indicates a sample average. If the two samples are dependent, then the covariance matrix should be included in the inverse matrix, which is also straightforward. See Appendix B for the proof.

However, there are two problems of the use of \( \mathcal{M}_2 \). The first problem is a purely statistical issue. The asymptotic result in (5) holds well only with large \( n \) and \( m \). In fact, even when \( Q_X \) and \( Q_Y \) are truncated normal distributions, the exact distribution of \( \mathcal{M}_2 \) is unknown. Second, even though the subset approach may detect distributional differences, this approach is not particularly helpful in understanding changes in the shapes of the distributions across
different treatments, which is important to theoreticians. To overcome this issue, we consider the following alternative approach.

### 3.2 Approximation by Beta Distribution

Many lab experimental data are well-approximated by the beta distribution. For example, Figure 1 shows histograms of the level of contribution to the public account for three selected games in NN (2008). As the degree of punishment increases, evidently contribution to the public account also increases. This pattern can be well approximated by using the beta distribution.

![Figure 1: Histograms of the Contribution Level (NN, 2008)](image)

It is hard or even impossible to statistically show that these changes in the distributions across games are significant using a nonparametric method, including the WMW test or the subset approach. However, the beta approximation provides a simple but effective solution. The beta distribution has been popularly used to approximate an unknown distribution. For example, Henshaw (1966) and Zhang and Wu (2002) approximated the exact finite sample distributions of the Durbin-Watson and KS statistics by using the beta distribution. McDonald and Xu (1995) approximate the empirical distributions of income and stock prices. They report that the approximation is fairly accurate. In particular, Zhang and Wu (2002)
report that the beta approximation error is around the fourth digit difference even with a small sample size. McDonald and Xu (1995) report in detail how many distributions can be generated from the generalized beta distribution. See Figure 2 in McDonald and Xu (1995) for a detailed explanation. Such flexibility attracts many theoretical researchers to using the beta distribution in their studies. Among many examples, here we list a few references related to experimental studies: King and Lukas (1973), McKelvey and Palfrey (1992) and Nyarko, Schotter and Sopher (2006).

Beta random variables are bounded between zero and one. The density can be convex or concave. It can have a U-shape, L-shape or bell-shape, and it can be unimodal or bimodal. The density function is given by
\[ f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad \text{for } x \in [0, 1], \alpha > 0, \beta > 0, \]
where \( B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} \, dx \), which is the beta function. This beta distribution has many interesting properties which will be discussed below.

### 3.2.1 Null Spaces under the Beta Distribution

All moments including the median can be expressed as functions of \( \alpha \) and \( \beta \). For example, \( \mu_X = E(x) = \frac{\alpha_X}{(\alpha_X + \beta_X)}, \mu_{2X} = V(x) = \alpha_X \beta_X \left[ (\alpha_X + \beta_X)^2 (\alpha_X + \beta_X + 1) \right]^{-1}. \)

Hence, when the two samples share the same \( \alpha \) and \( \beta \) values, the two samples come from the same beta distribution. That is, the null hypothesis of equal distributions can be rewritten as
\[
\mathcal{H}_0^Q: \alpha_X = \alpha_Y \& \beta_X = \beta_Y \iff \mathcal{H}_0^Q: Q_X = Q_Y. \tag{6}
\]

Since the beta distribution can be characterized with only two parameters, \( \alpha \) and \( \beta \), the null of equal distributions can also be expressed as
\[
\mathcal{H}_0^\lambda: \mu_X = \mu_Y \& \mu_{2X} = \mu_{2Y} \text{ with } \lambda = 2 \iff \mathcal{H}_0^Q: Q_X = Q_Y. \tag{7}
\]

That is, if the first two central moments of the two samples are the same, then the PDFs of the two samples are identical.

Next, the standard z-score test is commonly used for the null hypothesis of equal means. From simple and direct calculation, it is clear that
\[
\mathcal{H}_0^\lambda: \alpha_X/\beta_X = \alpha_Y/\beta_Y \iff \mathcal{H}_0^*: \mu_X = \mu_Y. \tag{8}
\]
since equal means implies that $\alpha_X / (\alpha_X + \beta_X) = \alpha_Y / (\alpha_Y + \beta_Y)$.

It is important to note that comparing means is not the only way to measure changes in the locations of distributions.

### 3.2.2 Location Shift and Skewness under the Beta Distribution

Consider three location parameters: the mean, median and mode. Usually the location of a distribution shifts to the right or left if one of these location parameter shifts right or left. When a distribution is symmetric, all three measures are identical, so a change in one location parameter implies a corresponding shift in the distribution. However, when a distribution is not symmetric, it is not always the case that all three location parameters move in the same direction. Under such a circumstance, it is very ambiguous to judge whether or not a distribution has changed location.

As we discussed before, the notion of skewness is useful when the distribution is bounded. Broadly, there are three different measures of skewness:

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moments</td>
<td>$\mu_3 / (\mu_2)^{3/2}$</td>
</tr>
<tr>
<td>Nonparametric</td>
<td>$3 (\mu - m) / \sqrt{\mu_2}$</td>
</tr>
<tr>
<td>Quantile</td>
<td>$[F (u) + F (1 - u) - 2F (0.5)] / [F (u) - F (1 - u)]$</td>
</tr>
</tbody>
</table>

where $u$ stands for the arbitrary quantile. Under the beta distribution, the skewness based on the moments has a one-for-one relationship with the nonparametric skewness and has a positive relationship overall with the quantile skewness. Accordingly and without loss of generality, we choose to use the definition of skewness based on the moments in this paper, and denote it as $S$. By definition, the skewness is measuring where the mass of the distribution is concentrated relative to the mean. If the skewness is positive (negative), the mass of the distribution is located on the right (left) side of the mean. Hence, when the distribution is bounded between zero and one, the notion of the skewness can be used as a measure of the location of the distribution. It is important to note that the skewness cannot be one of the location parameters when the distribution is unbounded. For example, under the beta distribution, the skewness can be rewritten as a function of the mean and variance. That is,

$$S = \frac{2 (1 - 2 \mu) \sigma}{\mu (1 - \mu) + \sigma^2}.$$
It is easy to show that with a fixed variance, the skewness has a one-for-one relationship with the mean. That is,

$$\frac{\partial S}{\partial \mu} = -2 \frac{\sigma (2\sigma^2 + 2\mu^2 - 2\mu + 1)}{(\sigma^2 - \mu^2 + \mu)^2} < 0 \text{ for } 0 < \mu < 1,$$  \hspace{1cm} (9)

since $\sigma^2 < \mu (1 - \mu)$ by the Bhatia-Davis inequality. The inequality in (9) implies that the skewness decreases as the mean increases. In other words, the mass of the distribution moves to the right.

Next, with a fixed mean, as the variance increases, the skewness changes also.

$$\frac{\partial S}{\partial \sigma} = 2 \frac{(2\mu - 1) (\sigma^2 + \mu^2 - \mu)}{(\sigma^2 - \mu^2 + \mu)^2} = \begin{cases} > 0 & \text{if } \mu < 0.5 \\ < 0 & \text{if } \mu > 0.5 \end{cases}.$$ \hspace{1cm} (10)

The inequality in (10) states that with a fixed mean, as the variance increases, the mass of the distribution moves to the left if $S > 0$ or the right if $S < 0$.

Note that the relationship between the median and the skewness is very similar but the formulas are somewhat complicated, so they are omitted here. We have verified their relationship using direct calculation. That is, the skewness has a one-for-one relationship with the median when the variance is fixed.

Some illustrations follow. For example, consider the following case where $\alpha_X = 2$, $\beta_X = 4$, $\alpha_Y = 3$, and $\beta_Y = 6$. Both samples share the same mean of 0.33 but don’t share either the same variance or skewness. The skewness for $\{y_i\}$ becomes 0.41 but that for $\{x_i\}$ is 0.47. Hence the series $\{y_i\}$ is less right skewed compared to the series $\{x_i\}$. Therefore, the mass of the empirical distribution of $\{y_i\}$ shifts to the right of that of $\{x_i\}$. Figure 2 shows this case explicitly. Another example is where $\alpha_X = 1.5$, $\beta_X = 3.833$, $\alpha_Y = 2$, and $\beta_Y = 5.333$. Here, both samples have the same median of 0.25 but the skewness, mean, and variance are all different. The skewness for $\{y_i\}$ becomes 0.63, but that for $\{x_i\}$ is 0.67. The distribution of $\{y_i\}$ shifts to the right of that of $\{x_i\}$ here as well. Figure 3 shows this case. Likewise, both samples can have the same mode when shifting occurs. Consider when $\alpha_X = 2$, $\beta_X = 3$, $\alpha_Y = 3$, and $\beta_Y = 4.333$. Here, the modes for both distributions are the same, occurring at 0.333, but the skewness of $\{x_i\}$ is 0.286 whereas the skewness of $\{y_i\}$ is 0.229. Figure 4 shows this case explicitly. Notice that in this case, the tail is far more important than than the highest mass point. Around the highest mass point, both distributions are very similar, so the tail becomes the only important difference. $\{y_i\}$ has a smaller right tail, so
\{y_i\} has shifted left of the distribution of \{x_i\}, the opposite direction of the other cases. Two distributions can also have the same variances when shifting occurs.

Therefore, shifting provides new information about how the density changes. This new information is fundamentally important because of its economic intuition: Suppose there are two groups, a treatment and a control for different punishment degrees in a public good game. As stated above, distributions of the contributions typically change when the degree of punishment increases. If the empirical distribution from the treatment group has shifted to the right of that from the control group, then most individuals in the treatment group are making higher contributions, while a small number of treated individuals are making much lower contributions. Thus, shifting compares how the treated individuals play relative to their own group using the control group as a ‘base-line’ for this comparison.

$$J_x = 2, \quad K_x = 4$$

$$J_y = 3, \quad K_y = 6$$

Figure 2: Shifting Densities with Equal Means
Figure 3: Shifting Densities with Equal Medians

Figure 4: Shifting Densities with Equal Modes
4 Estimation and Testing

Before we discuss new testing procedures using the beta distribution, we provide estimation results for the two parameters, \( \alpha \) and \( \beta \). Then, we discuss how to compare two independent samples effectively.

4.1 Estimation

There are two potentially viable approaches for estimating \( \alpha \) and \( \beta \) : method of moments and maximum likelihood estimation (MLE). Although MLE could have smaller variance, in practice MLE has not commonly been used because of two serious problems. First, the maximum likelihood estimators have no closed form. In general, some maximum likelihood approximation algorithms commonly fail for the beta distribution due to the presence of local maxima. This makes the maximum likelihood solution difficult, though not impossible, to implement (for details, see Gupta and Nadarajah p. 230). Second, lab experiments typically have many individuals who play absolute zero or one. Playing absolute zero or one should not be a serious problem because the players may only be playing approximately beta. Nonetheless, any \( x \) value of zero or one makes the likelihood function uniformly zero for all choices of \( \alpha \) and \( \beta \), so maximum likelihood estimation becomes impossible.

The method of moments estimator is rather simple and has been used widely in practice. Let \( \hat{\mu}_X \) and \( \hat{\sigma}_X^2 \) be the sample mean and variance which are defined by

\[
\hat{\mu}_X = \frac{1}{m} \sum_{i=1}^{m} x_i, \quad \text{and} \quad \hat{\sigma}_X^2 = \frac{1}{m-1} \sum_{i=1}^{m} (x_i - \hat{\mu}_X)^2.
\]

Then the method of moments estimators are given by

\[
\hat{\alpha}_X = \hat{\mu}_X \frac{(1 - \hat{\mu}_X) - \hat{\sigma}_X^2}{\hat{\sigma}_X^2}, \quad \hat{\beta}_X = (1 - \hat{\mu}_X) \frac{\hat{\mu}_X (1 - \hat{\mu}_X) - \hat{\sigma}_X^2}{\hat{\sigma}_X^2}.
\] (11)

It is straightforward to show that as \( m \to \infty \) the limiting distributions are given by

\[
\begin{bmatrix}
\sqrt{m} (\hat{\alpha}_X - \alpha) \\
\sqrt{m} (\hat{\beta}_X - \beta)
\end{bmatrix} \xrightarrow{d} \mathcal{N}(0, \Omega_X).
\]

See Appendix A for the detailed proof.
4.2 Testing Procedures

Here we explain a two-step procedure detailing how to compare two independent samples, and provide an asymptotic analysis of the proposed tests. The first null hypothesis to test is the null of equal distributions, which is a pre-test for the location change test. If the two samples share the same distribution, the second null hypothesis of the location change is not needed. The null hypothesis of equal distributions is given in (6), and the test statistic \( B \) is defined as

\[
B = \left[ \hat{\alpha}_X - \hat{\beta}_Y, \hat{x}_X - \hat{y}_Y \right] \left( \hat{\Omega}_x/m + \hat{\Omega}_y/n \right)^{-1} \left[ \hat{\alpha}_X - \hat{\beta}_Y, \hat{x}_X - \hat{y}_Y \right]',
\]

where \( \hat{\Omega}_x \) is provided in Appendix A. The following theorem reveals the asymptotic properties of this method.

**Theorem 1 (Limiting Distribution of the \( B \) statistic)** Assume that \( x_1, \ldots, x_m \sim iid \) \( \text{Beta}(\alpha_X, \beta_X) \), \( y_1, \ldots, y_n \sim iid \text{Beta}(\alpha_Y, \beta_Y) \), and \((x_i, y_j)\) are independent of each other for all \( i \) and \( j \). Then, under \( H_0^B \) given in (6),

\[
B \xrightarrow{d} \chi^2_2,
\]

as \( n, m \to \infty \).

Theorem 1 is proved in Appendix A and shows that the equal distributions test is consistent as \( n, m \to \infty \). Usually the 20\% significance level is used for pre-testing. In other words, we recommend using 3.22 as the critical value for the beta distribution statistic \( B \), which is the 20\% significance level for the \( \chi^2_2 \) distribution.

The next null hypothesis is location change. As we discussed earlier, the null of equal locations is given by

\[
H_0^S : S_X = S_Y.
\]

The null hypothesis of equal locations or equal skewness can be tested by using the method of moments or Pearson’s skewness test. However as we will show, the power of the test of Pearson’s skewness test is much worse than the power of the test based on the beta distribution. The test statistic based on the beta distribution is given by

\[
Z_{\text{Skew}} = \frac{\hat{S}_{X}^{(\text{Beta})} - \hat{S}_{Y}^{(\text{Beta})}}{(J'X \Omega X J_X/m + J'Y \Omega Y J_Y/n)^{1/2}},
\]
where
\[
J_k = \frac{\left(\hat{\alpha}_k + \hat{\beta}_k\right)}{\left(\hat{\alpha}_k + \hat{\beta}_k + 2\right) \sqrt{\hat{\alpha}_k \hat{\beta}_k \left(\hat{\alpha}_k + \hat{\beta}_k + 1\right)}} \\
\times \left[-\left(\hat{\beta}_k + 1\right) \left(3\hat{\alpha}_k + \hat{\beta}_k + 2\right) \left(\hat{\alpha}_k + 1\right) \left(\hat{\alpha}_k + 3\hat{\beta}_k + 2\right)\right]^{'}
\]
and
\[
\hat{S}_k^{(\text{Beta})} = \frac{2 \left(\hat{\beta}_k - \hat{\alpha}_k\right) \sqrt{\hat{\alpha}_k + \hat{\beta}_k + 1}}{\sqrt{\hat{\alpha}_k \hat{\beta}_k \left(\hat{\alpha}_k + \hat{\beta}_k + 2\right)}} \text{ for } k = X, Y.
\]

The following theorem shows the asymptotic performance of this estimator.

**Theorem 2 (Limiting Distribution of the Skewness Statistic)** Assume that \(x_1, \ldots, x_m\) \sim iid Beta(\(\alpha_X, \beta_X\)), \(y_1, \ldots, y_n\) \sim iid Beta(\(\alpha_Y, \beta_Y\)), and \((x_i, y_j)\) are independent of each other for all \(i\) and \(j\). Then, under \(H_0^S\) given in (12),
\[
z_{\text{Skew}} \xrightarrow{d} \mathcal{N}(0, 1),
\]
as \(n, m \to \infty\).

Theorem 2 is proved in Appendix A and shows that the equal skewness test is consistent as \(n, m \to \infty\).

So far, we have assumed that the two samples are independent. Hence all the test statistics considered in this paper are valid only under this independence assumption. In public good games, it is well-known that the contribution levels for each subject are correlated over rounds. In other words, the proposed tests cannot be performed for the comparison of a pair of two samples across rounds due to serial dependence. However, theoretically this restriction can be relaxed by accounting for a dependence structure. In order to analyze paired data \((x_1, y_1), \ldots, (x_n, y_n)\) using the beta approach, a bivariate beta distribution must be assumed. This distribution is still an active area of research. Some of these distributions will lend themselves to this estimation better than others. For instance, El-Bassiouny and Jones (2005) provide a distribution which nests many other bivariate beta distributions, with joint density:
\[
f(x, y) = C \frac{x^{a/2-1} (1-x)^{b+d/2-1} y^{b/2-1} (1-y)^{(a+d)/2-1}}{(1-xy)^{(a+b)/2}} F \left(\frac{a + b}{2}, \frac{d - c}{2}; \frac{a + d}{2}; x \left(1 - y\right)\right),
\]
where $C$ is a constant defined so that the double integral equals unity and $0 < x, y < 1$. The moments of this distribution involve the generalized hypergeometric function, so the method of moments estimator will not have a closed form. Similarly, the five parameter bivariate beta distribution in Gupta and Wong (1985) has moments which involve the generalized hypergeometric function. Other choices include the three parameter bivariate beta distribution from Gupta and Wong (1985) with joint density:

$$f(x, y) = \frac{\Gamma(a + b + c)}{\Gamma(a) \Gamma(b) \Gamma(c)} x^{a-1} y^{b-1} (1 - x - y)^{c-1},$$

where $x + y \leq 1$ and $x, y > 0$ and where $\Gamma(x)$ is the gamma function. Because this density has only three parameters and because the inequality is restrictive, this density is not very general. Nadarajah and Kotz (2005) define the density

$$f(x, y) = x^{c-1} (y - x)^{b-1} y^{a_1 - c - b} (1 - y)^{b_1 - 1} \frac{B(a_1, b_1) B(c, b)}{B(a, b)},$$

where $0 \leq x \leq y \leq 1$. Because of the inequality, this distribution is also quite restrictive. If a general bivariate beta distribution with closed-form method of moments estimators cannot be developed, then approximation methods can be used with some existing bivariate beta distributions.

As we discussed earlier and will show shortly, the bootstrap one-sample KS test can be used for testing whether or not a random variable follows the beta distribution. Since this bootstrap one-sample KS test can be considered a pre-test, the 20% significance level should be used. However, either when the bootstrap one-sample KS test rejects the null of the beta distribution or when the two samples are not independent, the following moments test can be used up to $\lambda = 2$. The major drawback of this moment-based test, however, is that the rejection of the null hypothesis $H_0^\lambda$ in (7) implies the rejection of the equal distributions but the opposite does not hold. The moments test statistic for the equal distributions test is

$$B^* = n \left[ \hat{x} - \bar{y}, \hat{\sigma}_x^2 - \hat{\sigma}_y^2 \right] \hat{\Sigma}^{-1} \left[ \hat{x} - \bar{y}, \hat{\sigma}_x^2 - \hat{\sigma}_y^2 \right]'$$

where $\hat{\Sigma}$ is defined in Appendix B. Under the null of equal distributions, $B^*$ converges in distribution to the $\chi^2_2$-distribution as $n \to \infty$, which follows from Lemma 5 in Appendix B. The moments test statistic for the equal skewness test is

$$z_{\text{Skew}}^* = \sqrt{n} \frac{\hat{S}_X - \hat{S}_Y}{\sqrt{\hat{\omega}_X + \hat{\omega}_Y - 2\hat{\omega}_{X,Y}}}$$
where \( \hat{\omega}_X, \hat{\omega}_Y, \) and \( \hat{\omega}_{X,Y} \) are defined in Appendix B. Under the null of equal skewness, \( z_{\text{Skew}}^* \) converges in distribution to the standard normal distribution as \( n \to \infty \), which follows from Lemma 4 in Appendix B.

In the following section, a testing procedure for whether or not the samples follow beta distributions is proposed and the finite sample performance of the moments tests is also examined.

5 Monte Carlo Simulation and Empirical Example

This section consists of two subsections. The first subsection details the finite sample performance of the one-sample bootstrapped KS test, the WMW test, and the sequential testing procedure. The second subsection demonstrates the effectiveness of the suggested test statistics.

5.1 Monte Carlo Simulation

The tests are performed at the 5% nominal level. For all cases except for the bootstrapped one-sample KS test, 10,000 replications are used whereas for the cases of the bootstrapped one-sample KS test, 2,000 replications are used and for each replication 1,000 bootstrap replications are used. We use the following bootstrap procedure for the one-sample KS test.

Bootstrap Procedure for The One-Sample KS Test

**Step 1:** Estimate \( \hat{\alpha} \) and \( \hat{\beta} \) by using the method of moments. Obtain the one-sample KS statistic based on the estimated \( \hat{\alpha} \) and \( \hat{\beta} \).

**Step 2:** Generate \( n \) pseudo-beta random variables with \( \hat{\alpha} \) and \( \hat{\beta} \). Estimate \( \alpha \) and \( \beta \) using the pseudorandom variables. Obtain the one-sample KS statistic.\(^1\)

**Step 3:** Repeat Step 2 for 1,000 times. Obtain the critical value from the bootstrapped distribution and compare the KS statistic in Step 1 with this critical value. If the KS statistic in Step 1 is larger than the bootstrapped critical value, the null of the beta distribution is rejected.

\(^1\)As Babu and Rao (2004) and Meintanis and Swanepoel (2009) point out, the above parametric bootstrap procedure does not need to correct any bias in the method of moments estimation.
Table 2 displays the rejection rates of the one-sample KS test under the null hypothesis of the beta distribution. We consider five combinations of $\alpha$ and $\beta$ under the null hypothesis. The first column reports the rejection frequencies when $\alpha$ and $\beta$ are known, which is an infeasible one-sample KS test. The second column shows the rejection rates with estimated $\alpha$ and $\beta$, and the last column exhibits the rejection rate based on the bootstrapped critical values. For all cases, the infeasible KS test does not show any significant size distortion.
even in small samples \((n = 25)\). Meanwhile, as we discussed earlier, the one-sample KS test using the estimates of \(\alpha\) and \(\beta\) suffers from serious size distortion. The direction of the size distortion depends on the values of \(\alpha\) and \(\beta\). When \(\alpha < \beta < 1\), the one-sample KS test suffers from upward size distortion. Otherwise, the one-sample KS test suffers from downward size distortion. Once the bootstrap critical values are used, the size distortion is reduced in every case, and as \(n\) increases, the rejection rates converge to the nominal size of 5%.

Table 3: Power of the One-Sample KS Test

\[\text{(5\% Nominal Size)}\]

<table>
<thead>
<tr>
<th>(\sigma_Y^2)</th>
<th>(n)</th>
<th>(\mu_Y = 0.5)</th>
<th>(\mu_Y = 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>25</td>
<td>0.067</td>
<td>1.000</td>
</tr>
<tr>
<td>0.1</td>
<td>50</td>
<td>0.255</td>
<td>1.000</td>
</tr>
<tr>
<td>0.1</td>
<td>100</td>
<td>0.593</td>
<td>1.000</td>
</tr>
<tr>
<td>0.1</td>
<td>200</td>
<td>0.931</td>
<td>1.000</td>
</tr>
<tr>
<td>0.5</td>
<td>25</td>
<td>0.244</td>
<td>0.954</td>
</tr>
<tr>
<td>0.5</td>
<td>50</td>
<td>0.980</td>
<td>1.000</td>
</tr>
<tr>
<td>0.5</td>
<td>100</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.5</td>
<td>200</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>0.150</td>
<td>0.680</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>0.984</td>
<td>0.999</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>200</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

To evaluate the power of the one-sample KS test, we consider a truncated normal distribution between zero and one as the alternative. Denote the truncated normal variable with mean \(\mu_Y\) and variance \(\sigma_Y^2\) between zero and one as

\[
Y \sim iid \mathcal{N} \left(\mu_Y, \sigma_Y^2\right)_{(0,1)}.
\]

We set \(\mu_Y = (0, 0.5)\) and \(\sigma_Y^2 = (0.1, 0.5, 1)\). Table 3 reports the power of the test. When \(\mu_Y = 0\), the distributions become asymmetric and have a \(U\)-shape. Meanwhile, when \(\mu_Y = 0.5\), the distribution becomes symmetric and will have three modes. When \(\mu_Y = 0\), the rejection rate increases as the variance \(\sigma_Y^2\) decreases or as the sample size increases.
When $\mu_Y = 0.5$, the rejection rate does not seem to depend much on the variance. However, as $n$ increases, the rejection rate approaches unity in all cases.

Next, we examine the finite sample performance of the equal distributions test and the equal skewness test. For all the following simulations, the data generating process (DGP) is given by

$$x_i \sim \text{Beta}(\alpha_X, \beta_X), \quad y_i \sim \text{Beta}(\alpha_Y, \beta_Y).$$

Under the null of equal distributions from (6), $\alpha_X = \alpha_Y$ and $\beta_X = \beta_Y$. Under the null of equal skewness from (12), $\text{Skew}(X) = \text{Skew}(Y)$. In both processes, the samples are independent, identically distributed, and they are independent of each other. In practice, the sample size for each of the two samples could be different. For simplicity, the sample size for both samples will be fixed to one number, $n$. Various cases were considered but only the following five cases are reported to save space.

<table>
<thead>
<tr>
<th>DGP</th>
<th>$X$</th>
<th>$Y$</th>
<th>$X$</th>
<th>$Y$</th>
<th>$X$</th>
<th>$Y$</th>
<th>$X$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha</td>
<td>1.70</td>
<td>1.70</td>
<td>9</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>1.8</td>
<td>7</td>
</tr>
<tr>
<td>Beta</td>
<td>5.54</td>
<td>5.54</td>
<td>3</td>
<td>1.35</td>
<td>2</td>
<td>1</td>
<td>3.27</td>
<td>13.67</td>
</tr>
<tr>
<td>Mean</td>
<td>0.24</td>
<td>0.24</td>
<td>0.75</td>
<td>0.69</td>
<td>0.75</td>
<td>0.75</td>
<td>0.36</td>
<td>0.34</td>
</tr>
<tr>
<td>Median</td>
<td>0.23</td>
<td>0.23</td>
<td>0.76</td>
<td>0.72</td>
<td>0.77</td>
<td>0.79</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.78</td>
<td>0.78</td>
<td>-0.60</td>
<td>-0.60</td>
<td>-0.69</td>
<td>-0.86</td>
<td>0.42</td>
<td>0.28</td>
</tr>
</tbody>
</table>

DGP 1 is used to demonstrate the size of the test of equal distributions. The distribution is constructed to resemble the p0 sample from NN. In DGP 2, the two distributions are different, yet both have equal skewness. Hence, when the equal distributions test is used on DGP 2, it demonstrates the power of the test, but when the equal skewness test is used, it demonstrates the size. DGP 3 shows the power of the tests when the samples have equal means but the median and skewness are different. DGP 4 is used to show the power of the tests when the two samples have equal medians. DGP 5 is used to demonstrate the power of the tests when the two parameters, mean, median and skewness of the two samples are all different.
Table 4: Test Rejection Rates (Nominal: 5%)

<table>
<thead>
<tr>
<th>DGP</th>
<th>$n$</th>
<th>WMW</th>
<th>KS</th>
<th>Moments Tests</th>
<th>Beta Dist. Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Equal Dist.</td>
<td>Equal Dist.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Skewness</td>
<td>Skewness</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>0.05</td>
<td>0.03</td>
<td>0.09</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.04</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>0.13</td>
<td>0.17</td>
<td>0.68</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.23</td>
<td>0.40</td>
<td>0.91</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.41</td>
<td>0.82</td>
<td>1.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.78</td>
<td>1.00</td>
<td>1.00</td>
<td>0.02</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>0.07</td>
<td>0.08</td>
<td>0.38</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.09</td>
<td>0.15</td>
<td>0.59</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.13</td>
<td>0.38</td>
<td>0.87</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.26</td>
<td>0.81</td>
<td>1.00</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>0.06</td>
<td>0.17</td>
<td>0.89</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.06</td>
<td>0.44</td>
<td>1.00</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.06</td>
<td>0.92</td>
<td>1.00</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.07</td>
<td>1.00</td>
<td>1.00</td>
<td>0.12</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>0.15</td>
<td>0.13</td>
<td>0.43</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.23</td>
<td>0.26</td>
<td>0.66</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.43</td>
<td>0.58</td>
<td>0.91</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.81</td>
<td>0.95</td>
<td>1.00</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Table 4 reports the probability of rejecting the null hypothesis for the WMW test, the two sample KS test, the method of moments tests and the beta distribution tests, where the latter two both include an equal distributions test and an equal skewness test.

Size and Power of Equal Distributions Test: With DGP 1, where the distributions of the two samples are the same, WMW has very accurate size. The size of the method of moments test decreases gradually from 0.09 to 0.06 as $n$ increases from 25 to 250. The beta
distribution test is mildly conservative when $n$ is small, but the size becomes close to 0.05 when $n$ is large. In terms of power of the test – as demonstrated using DGP 2 to DGP 5 – both equal distributions tests substantially dominates the WMW and KS tests and both reject perfectly when $n$ is greater than 100. When the moments test and the beta distribution test are compared, beta distribution test has lower power.

Size and Power of the Equal Skewness Test: Here, the distribution pre-test is performed at the 20% level, while the other tests are performed at the 5% level. DGP 2 shows the size of the equal skewness tests, while DGP 3 to DGP 5 demonstrate the power. In terms of size, the beta distribution test is quite accurate even when $n$ is small, whereas the method of moments test shows undersize distortion, rejecting in less than 5% of cases. Also, the equal skewness test using the beta distribution is quite powerful. With DGP 2 to DGP 5, the power of the moments test is quite low even when $n$ is large, while the power of the beta distribution test increases quickly.

5.2 Return to the Empirical Examples

Here we demonstrate the effectiveness of the suggested tests. Before we estimate $\alpha$ and $\beta$, we test whether or not each sample is drawn from the beta distribution. Using the bootstrapped one-sample KS test, we cannot reject the null of the beta distribution for all seven samples at the 5% level.

Table 5 reports the results of the equal distribution and skewness tests. Since it is recommended that we use the 20% significance level for the pre-test, the null hypothesis of equal distributions is rejected in all five cases, where the $B$—statistics are much larger than 3.22, the critical value for the $\chi^2$ distribution. Next, we perform the equal skewness tests. For the p3/p4 case, the z-score test did not reject the same mean at the 10% level. However, the $z_{\text{Skew}}$ statistic is much larger than 1.96, so the null of equal skewness is rejected at the 5% level. Hence the mass of the distribution in the $Y$ sample (p4) is located to the right of the mass of the distribution in the $X$ sample (p3).

Similarly, for the comparison between p0 and p1, the WMW test does not reject the null hypothesis $H^w_0$ at the 10% level for the two-sided test. However, both of the new beta test statistics, $B$ and $z_{\text{Skew}}$, reject their respective null hypotheses at the 10% level. For all the cases where both the WMW and z-score tests reject the null hypotheses, the $B$ and $z_{\text{Skew}}$ statistics also reject the null hypotheses as well.
Table 5: Canonical Examples Reexamined using the Beta Tests (Standard Errors)

<table>
<thead>
<tr>
<th></th>
<th>WMW</th>
<th>z-score</th>
<th>KS</th>
<th>$\hat{\alpha}_X$</th>
<th>$\hat{\beta}_X$</th>
<th>$\hat{\alpha}_Y$</th>
<th>$\hat{\beta}_Y$</th>
<th>B</th>
<th>$z_{Skew}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>-1.640</td>
<td>-2.444</td>
<td>0.109</td>
<td>1.697</td>
<td>5.544</td>
<td>0.713</td>
<td>1.059</td>
<td>6.422</td>
<td>1.655</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.51)</td>
<td>(1.77)</td>
<td>(0.21)</td>
<td>(0.33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p1</td>
<td>-2.238</td>
<td>-2.179</td>
<td>0.109</td>
<td>0.713</td>
<td>1.059</td>
<td>1.052</td>
<td>0.726</td>
<td>3.796</td>
<td>2.159</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.21)</td>
<td>(0.33)</td>
<td>(0.33)</td>
<td>(0.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p2</td>
<td>-2.135</td>
<td>-2.954</td>
<td>0.021</td>
<td>1.052</td>
<td>0.726</td>
<td>5.625</td>
<td>1.465</td>
<td>6.515</td>
<td>2.183</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.33)</td>
<td>(0.22)</td>
<td>(1.85)</td>
<td>(0.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p3</td>
<td>-1.734</td>
<td>-1.289</td>
<td>0.216</td>
<td>5.625</td>
<td>1.465</td>
<td>3.264</td>
<td>0.570</td>
<td>3.852</td>
<td>1.992</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.85)</td>
<td>(0.45)</td>
<td>(1.27)</td>
<td>(0.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRS</td>
<td>-3.157</td>
<td>-3.473</td>
<td>0.002</td>
<td>1.683</td>
<td>2.360</td>
<td>1.225</td>
<td>0.838</td>
<td>13.390</td>
<td>3.325</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.31)</td>
<td>(0.44)</td>
<td>(0.27)</td>
<td>(0.18)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6 Conclusion

In this paper, we began by summarizing the literature on two-sample comparisons. In particular, the WMW test, the traditional z-score test, and the KS test are the best at testing the null hypotheses of zero median difference, equal means, and equal distributions, respectively. We then used these three tests with two canonical examples of public good games, finding that the WMW and the z-score tests perform adequately, but that the KS test has low power in the finite sample.

Next, we devised two new approaches to the two-sample comparison problem, which are then used to develop new tests. The first new approach concerns testing for equal distributions. In the context of bounded data, the moments completely determine a distribution up to uniqueness, so it is possible to test for equal distributions by simply testing the moments. The second new approach concerns testing for location changes in bounded data. Since the mean, median, and mode are all free floating, it is somewhat common in bounded data that one moves right and another left, making these common measures of central tendency inadequate. We find that changes in skewness robustly convey much of the information about the overall movement of bounded densities. The derivations of these new tests are provided.
Afterward, we propose using the beta distribution to perform these tests parametrically to find important estimation results and increase the power of the tests. Since an approximate beta distribution cannot always be assumed, we advocate using a one-sample KS test for the beta distribution, which performs quite well in the finite sample. The two parameters for the beta distribution are estimated using the method of moments, because the maximum likelihood estimator commonly does not exist for experimental data. The limiting distribution of the beta distribution parameter estimators is provided. Distributional testing follows from knowing the limiting distribution of the parameter estimators, and the beta distribution equal skewness test is also provided.

Finally, we evaluate the finite sample performance of our newly developed methods through Monte Carlo simulations and through application to the canonical examples. We find that the new beta tests perform quite well in the finite sample from the Monte Carlo study. In the empirical example, for the tests of the null hypothesis that the samples are drawn from the beta distribution, each of the seven cases accepts the null at the 5% level, so it seems quite likely that the beta tests will perform reasonably well. In all five cases, we find that the new distribution test rejects the null of equal distributions at the recommended 20% level. We also find that the new skewness test rejects the null in four of the five cases at the 5% level.
References


Appendix A – Proofs for Beta Tests

Define $\mu_k$ as the non-central moments of beta random variable $x$, which is given by

$$\mu_k = E[x^k] = \prod_{i=0}^{k-1} \frac{\alpha_x + i}{\alpha_x + \beta_x + i}.$$

Next, define

$$\Omega_x = J^o \Omega_x^o J^o,$$

where

$$J^o = \begin{bmatrix} (\mu_2 - \mu_1^2)^{-1} & 0 & - (\mu_2 - \mu_1^2)^{-2} \mu_1 (\mu_1 - \mu_2) \\ 0 & (\mu_2 - \mu_1^2)^{-1} & - (\mu_2 - \mu_1^2)^{-2} (1 - \mu_1) (\mu_1 - \mu_2) \end{bmatrix},$$

and

$$\Omega_x^o = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix},$$

with

$$\begin{align*}
\sigma_{11} &= \tau_4 + 4 (\mu_2 - \mu_3) \mu_1^2 - 4 \mu_1 \mu_2^2, \\
\sigma_{12} &= \tau_1 - \tau_4 - \tau_2 + 6 \mu_2^2 \mu_3 + (1 + 2 \mu_2) \mu_1 \mu_2 - \mu_2^2, \\
\sigma_{13} &= -\tau_1 - 2 \tau_3 - 2 (1 + 2 \mu_1) \mu_1^2 \mu_2 + 2 \mu_1 \mu_3, \\
\sigma_{22} &= 2 \tau_1 + \tau_4 + 2 \tau_2 - (1 + 4 \mu_2) \mu_1^2 + (1 + \mu_2) \mu_2 - 2 \mu_3 + \mu_4, \\
\sigma_{23} &= \tau_1 + 2 \tau_3 + 2 (1 + 2 \mu_2) \mu_3^2 + (3 \mu_2 - 5 \mu_1) \mu_2 + \mu_3 - \mu_4, \\
\sigma_{33} &= 4 (3 \mu_2 - \mu_1^2) \mu_1^2 - 4 \mu_1 \mu_3 - 5 \mu_2^2 + \mu_4,
\end{align*}$$

where $\tau_1 = \mu_1 \mu_4 + \mu_2 \mu_3 - 2 \mu_3 \mu_2^2$, $\tau_2 = 3 \mu_1 \mu_3 + 2 \mu_1^3 - \mu_1 \mu_2$, $\tau_3 = 3 \mu_2^2 \mu_2 - 3 \mu_1 \mu_2^2 - 2 \mu_4^2$, $\tau_4 = \mu_2^3 - 4 \mu_2^4 + 4 \mu_1^2 \mu_2^2 + 8 \mu_1^3 \mu_2 + \mu_1^2 \mu_3 + 2 \mu_1 \mu_2 \mu_3$. $\Omega_y$ is similarly defined as $\Omega_x$ so the formula for $\Omega_y$ is omitted to save space.

For Lemma 1, we further define $n_s = n - s$, e.g. $n_1 = n - 1$, and $\bar{x}^k = (n^{-1} \sum_{i=1}^n x_i)^k$.

Lemma 1 (Expected Values of Beta Family Sample Moments) If $x_1, \ldots, x_n$ are drawn from a random sample beta distribution, then the expectations of the following powers
are:

\[
\begin{align*}
E[\bar{x}] &= \mu_1 \\
nE[\bar{x}^2] &= n_1\mu_1^2 + \mu_2 \\
n^2E[\bar{x}^3] &= n_1n_2\mu_1^3 + 3n_1\mu_1\mu_2 + \mu_3 \\
n^3E[\bar{x}^4] &= n_1n_2n_3\mu_1^4 + 6n_1n_2\mu_1^2\mu_2 + n_1(3\mu_2^2 + 4\mu_1\mu_3) + \mu_4 \\
E[(s_2^2 + \bar{x}^2)] &= \mu_2 \\
nE[(s_2^2 + \bar{x}^2)] &= n_1\mu_1\mu_2 + \mu_3 \\
n^2E[\bar{x}^2(s_2^2 + \bar{x}^2)] &= n_1n_2\mu_1^2\mu_2 + n_1(2\mu_1\mu_3 + \mu_2^2) + \mu_4 \\
n^3E[\bar{x}^3(s_2^2 + \bar{x}^2)] &= n_1n_2n_3\mu_1^3\mu_2 + 3n_1n_2(\mu_1^2\mu_3 + \mu_1\mu_2^2) + n_1(4\mu_2\mu_3 + 3\mu_1\mu_4) + \mu_5 \\
nE[(s_2^2 + \bar{x}^2)\mu_2] &= n_1\mu_2^2 + \mu_4 \\
n^2E[\bar{x}(s_2^2 + \bar{x}^2)] &= n_1n_2\mu_1\mu_2 + n_1(\mu_1\mu_4 + 2\mu_2\mu_3) + \mu_5 \\
n^3E[\bar{x}^2(s_2^2 + \bar{x}^2)] &= n_1n_2[\mu_1^2(n_3\mu_2^2 + \mu_4) + \mu_2(\mu_2^2 + 4\mu_1\mu_3)] + n_1(3\mu_2\mu_4 + 2\mu_3^2 + 2\mu_1\mu_5) + \mu_6
\end{align*}
\]

The proofs of Lemma 1 are straightforward hence omitted.

**Lemma 2 (Limiting Distribution of Method of Moments Components)**

If \(x_1, \ldots, x_n\) are drawn from a random sample beta distribution, then

\[
\sqrt{n} \left( [\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3]^T - [\theta_1, \theta_2, \theta_3]^T \right) \xrightarrow{d} \mathcal{N} \left( [0, 0, 0]^T, \Omega_\theta \right).
\]

where

\[
\theta_1 = \mu_1 (\mu_1 - \mu_2), \quad \theta_2 = (1 - \mu_1) (\mu_1 - \mu_2), \quad \theta_3 = \mu_2 - \mu_1^2,
\]

and

\[
\hat{\theta}_1 = (\bar{x}) \left( (1 - \bar{x}) - s_2^2 \right), \quad \hat{\theta}_2 = (1 - \bar{x}) \left( (1 - \bar{x}) - s_2^2 \right), \quad \hat{\theta}_3 = s_2^2.
\]

**Proof of Lemma 2**

By using Lemma 1, it is straightforward to show that:

\[
\begin{align*}
E[\hat{\theta}_1] &= \frac{n_1}{n} (\mu_1^2 - \mu_1\mu_2) + \frac{1}{n} (\mu_2 - \mu_3), \\
E[\hat{\theta}_2] &= (\mu_1 - \mu_2) + \frac{n_1}{n} (\mu_1\mu_2 - \mu_1^2) + \frac{1}{n} (\mu_3 - \mu_2), \\
E[\hat{\theta}_3] &= \frac{(n + 1)}{n} \mu_2 - \frac{n_1}{n} \mu_1^2.
\end{align*}
\]
and

\[
\begin{align*}
n \text{Var} \left[ \hat{\theta}_1 \right] &= \mu_2^3 - 4 \mu_1^4 - 4 \mu_1^2 \mu_2^2 - 4 \mu_1 \mu_2^3 + 4 \mu_1^2 \mu_2 + 4 \mu_1 \mu_3^2 - 4 \mu_1^2 \mu_3 + 8 \mu_1^3 \mu_2 + \\
& \quad + \mu_1^2 \mu_4 + 2 \mu_1 \mu_2 \mu_3 + \mathcal{O} \left( n^{-1} \right), \\
n \text{Var} \left[ \hat{\theta}_2 \right] &= \mu_2 - 2 \mu_3 + \mu_4 - \mu_1^2 + 4 \mu_1^3 + \mu_2^2 - 4 \mu_1^4 + \mu_3^2 - 4 \mu_1^2 \mu_2^2 \\
& \quad - 2 \mu_1 \mu_2 + 6 \mu_1 \mu_3 - 2 \mu_4 - 2 \mu_2 \mu_3 - 4 \mu_1^2 \mu_2 - 4 \mu_1^2 \mu_3 + \\
& \quad + 8 \mu_1^3 \mu_2 + \mu_1 \mu_2 \mu_3 + \mathcal{O} \left( n^{-1} \right), \\
n \text{Var} \left[ \hat{\theta}_3 \right] &= -4 \mu_1^4 + 12 \mu_1^2 \mu_2 - 4 \mu_3 \mu_1 - 5 \mu_2^2 + \mu_4 + \mathcal{O} \left( n^{-1} \right), \\
n \text{Cov} \left[ \hat{\theta}_1, \hat{\theta}_2 \right] &= 4 \mu_1^4 - \mu_1^2 - 2 \mu_1^3 - \mu_2^2 + 4 \mu_1^2 \mu_2^2 + 2 \mu_1 \mu_2 - 3 \mu_1 \mu_3 + 6 \mu_1 \mu_4 + \\
& \quad + \mu_2 \mu_3 + 2 \mu_1 \mu_2 + 4 \mu_1^2 \mu_3 - 8 \mu_1^3 \mu_2 - \mu_2 \mu_4 - 2 \mu_1 \mu_2 \mu_3 + \\
& \quad + \mathcal{O} \left( n^{-1} \right), \\
n \text{Cov} \left[ \hat{\theta}_1, \hat{\theta}_3 \right] &= 4 \mu_1^4 + 2 \mu_1 \mu_3 - \mu_1 \mu_4 - \mu_2 \mu_3 + 6 \mu_1 \mu_2^2 - 8 \mu_1^2 \mu_2 + 2 \mu_1^2 \mu_3 - \\
& \quad - 4 \mu_1^3 \mu_2 + \mathcal{O} \left( n^{-1} \right), \\
n \text{Cov} \left[ \hat{\theta}_2, \hat{\theta}_3 \right] &= -4 \mu_1^4 + 4 \mu_1^3 \mu_2 + 2 \mu_1^3 + 6 \mu_1^2 \mu_2 - 2 \mu_3 \mu_2 - 6 \mu_1 \mu_2^2 - 5 \mu_1 \mu_2 + \\
& \quad + \mu_4 \mu_1 + 3 \mu_2^2 + \mu_3 \mu_2 + \mu_3 - \mu_4 + \mathcal{O} \left( n^{-2} \right).
\end{align*}
\]

Multiplying by \( n \) and taking the limit as \( n \to \infty \) yields the variance-covariance matrix expressed above. By taking transformations, we can apply the Lindeberg-Levy central limit theorem and the statement is proved. □

**Lemma 3 (Limiting Distribution of the Method of Moments Estimator)** If \( x_1, \ldots, x_n \) are drawn from a random sample beta distribution with parameters \( \alpha_x \) and \( \beta_x \), then

\[
\sqrt{n} \left( \left[ \hat{\alpha}_x, \hat{\beta}_x \right] - [\alpha_x, \beta_x] \right) \xrightarrow{d \, \mathcal{L}} \mathcal{N} ( [0,0]', \Omega_x ).
\]

**Proof of Lemma 3**

Let \( g_1, g_2 : \mathbb{R}^3 \to \mathbb{R} \) be defined by:

\[
g_1 (\theta_1, \theta_2, \theta_3) = \frac{\theta_1}{\theta_3}, \quad g_2 (\theta_1, \theta_2, \theta_3) = \frac{\theta_2}{\theta_3}.
\]
Differentiating, we find that:

\[
\frac{\partial g_1 (\theta_1, \theta_2, \theta_3)}{\partial x} = \frac{1}{\theta_3}, \quad \frac{\partial g_1 (\theta_1, \theta_2, \theta_3)}{\partial y} = 0, \quad \frac{\partial g_1 (\theta_1, \theta_2, \theta_3)}{\partial \theta_3} = -\frac{\theta_1}{\theta_3^2};
\]

\[
\frac{\partial g_2 (x, y, z)}{\partial x} = 0, \quad \frac{\partial g_2 (x, y, z)}{\partial y} = \frac{1}{\theta_3}, \quad \frac{\partial g_2 (x, y, z)}{\partial z} = -\frac{\theta_2}{\theta_3^2}.
\]

The Jacobian, \( J_x \), consists of the above partial derivatives evaluated at the following values:

\[
\begin{align*}
\theta_1 &= \lim_{n \to \infty} \mathbb{E} [m_1^2 - m_1 m_2] = \mu_1 (\mu_1 - \mu_2), \\
\theta_2 &= \lim_{n \to \infty} \mathbb{E} [m_1 (1 - m_1) - (1 - m_1) m_2] = (1 - \mu_1) (\mu_1 - \mu_2), \\
\theta_3 &= \lim_{n \to \infty} \mathbb{E} [m_2 - m_1^2] = \mu_2 - \mu_1^2.
\end{align*}
\]

From here, we can use the delta method and Lemma 2 to find the limiting distribution of \( (\hat{\alpha}_x, \hat{\beta}_x)' \), and the lemma is proved. \( \square \)

### 6.1 Proof of Theorem 1

For simplicity, assume that \( n = m \). From Lemma 3 and the independence of \( x_i \) and \( y_j \) for all \( i \) and \( j \), it follows that

\[
\sqrt{n} \left( (\hat{\alpha}_X - \hat{\alpha}_Y, \hat{\beta}_X - \hat{\beta}_Y) - (\alpha_X - \alpha_Y, \beta_X - \beta_Y) \right) \xrightarrow{d} \mathcal{N} \left( (0, 0)', \Omega_X + \Omega_Y \right)
\]

as \( n \to \infty \). Hence,

\[
\left( (\hat{\alpha}_X - \hat{\alpha}_Y, \hat{\beta}_X - \hat{\beta}_Y) - (\alpha_X - \alpha_Y, \beta_X - \beta_Y) \right) \left( \hat{\Omega}_x/m + \hat{\Omega}_y/n \right)^{-1/2} \xrightarrow{d} \mathcal{N} \left( (0, 0)', \mathbf{I}_2 \right),
\]

as \( n, m \to \infty \). Thus,

\[
\mathcal{B} \xrightarrow{d} \chi^2_2,
\]

as \( n, m \to \infty \). From the null hypothesis, we know that \( \alpha_X = \alpha_Y \) and \( \beta_X = \beta_Y \). Therefore

\[
\mathcal{B} \xrightarrow{d} \chi^2_2,
\]

as \( n, m \to \infty \). \( \square \)
Proof of Theorem 2

The result for the skewness test can be proved by applying the delta method. The skewness of sample \( k \) is estimated by

\[
\hat{S}_{k}^{(\text{Beta})} = \frac{2(\hat{\beta}_k - \hat{\alpha}_k) \sqrt{\hat{\alpha}_k + \hat{\beta}_k + 1}}{\sqrt{\hat{\alpha}_k} \hat{\beta}_k (\hat{\alpha}_k + \hat{\beta}_k + 2)} \quad \text{for } k = X, Y.
\]

The Jacobian for the beta method of moments skewness is,

\[
J_k = \frac{(\alpha_k + \beta_k)}{(\alpha_k + \beta_k + 2) \sqrt{\alpha_k} \beta_k \sqrt{\alpha_k + \beta_k + 1}} \times \left[ -\frac{(\beta_k + 1)(3\alpha_k + \beta_k + 2)}{\alpha_k} \frac{(\alpha_k + 1)(\alpha_k + 3\beta_k + 2)}{\beta_k} \right]'.
\]

The skewness test statistic is defined as

\[
z_{\text{Skew}} = \frac{\hat{S}_{X}^{(\text{Beta})} - \hat{S}_{Y}^{(\text{Beta})}}{(J'_{X} \Omega_{X} J_{X}/m + J'_{Y} \Omega_{Y} J_{Y}/n)^{1/2}}.
\]

Hence, under the null that the distributions have equal skewness,

\[
z_{\text{Skew}} \rightarrow^d N(0, 1),
\]

and the theorem is proved. \( \square \)

Appendix B – Proofs for Nonparametric Testing

Throughout Appendix B, denote \( \mu_{i,j} \) and \( \bar{\mu}_{i,j} \) as the \((i, j)\)th non-central and central comoment of \((X, Y)\) respectively. So \( \bar{\mu}_{i,j} = E \left[ (x - \mu_x)^i (y - \mu_y)^j \right] \), and since the two samples are not assumed to be independent, this cannot be further reduced to a product of two expectations. \( \bar{x} \) are \( \hat{\sigma}_x^2 \) are the usual sample mean and sample variance for the \( x = (x_1, \ldots, x_n) \).

Lemma 4 (General Limiting Distribution of the Mean and Variance) Where \( \{(x_i, y_i)\} \) is a size \( n \) random sample from a bivariate distribution with finite fourth moments, under the null that \( x \) and \( y \) have the same mean and variance,

\[
\sqrt{n} \left[ \bar{x} - \bar{y}, \hat{\sigma}_x^2 - \hat{\sigma}_y^2 \right] \rightarrow^d N \left( (0, 0)', \Xi \right),
\]

33
where

\[ \Xi = \begin{bmatrix}
\mu_{2,0} + \mu_{0,2} - 2\mu_{1,1} & \mu_{3,0} + \mu_{0,3} - \mu_{1,2} - \mu_{2,1} \\
\mu_{3,0} + \mu_{0,3} - \mu_{1,2} - \mu_{2,1} & \mu_{4,0} - \mu_{2,0} + \mu_{0,4} - \mu_{0,2} - 2\mu_{2,2} + 2\mu_{0,2}\mu_{2,0}
\end{bmatrix} \]

Proof of Lemma 4

It is well-known that

\[ \text{E} [\bar{x}] = \mu_{1,0}, \quad \text{E} [\sigma^2_x] = \bar{\mu}_{2,0}, \]

and

\[ n\text{Var} [\bar{x}] = \bar{\mu}_{2,0}, \quad n\text{Var} [\sigma^2_x] = \bar{\mu}_{4,0} - \frac{n - 3}{n - 1}\mu^2_{2,0}. \]

It is also easy to derive:

\[ n\text{Cov} [\bar{x}, \bar{y}] = \mu_{1,1}, \quad n\text{Cov} [\bar{x}, \sigma^2_x] = \mu_{3,0}, \quad n\text{Cov} [\bar{x}, \sigma^2_y] = \mu_{1,2}. \]

The only tedious term is

\[ n\text{Cov} [\sigma^2_x, \sigma^2_y] = \bar{\mu}_{2,2} - \bar{\mu}_{0,2}\bar{\mu}_{2,0} + \frac{2}{n - 1}\mu^2_{1,1}. \]

Using the symmetric form, we know all the components of:

\[ \Xi = \lim_{n \to \infty} \begin{bmatrix}
n\text{Var} [\bar{x} - \bar{y}] & n\text{Cov} [\bar{x} - \bar{y}, \sigma^2_x - \sigma^2_y] \\
n\text{Cov} [\bar{x} - \bar{y}, \sigma^2_x - \sigma^2_y] & n\text{Var} [\sigma^2_x - \sigma^2_y]
\end{bmatrix} = \begin{bmatrix}
\mu_{2,0} + \mu_{0,2} - 2\mu_{1,1} & \mu_{3,0} + \mu_{0,3} - \mu_{1,2} - \mu_{2,1} \\
\mu_{3,0} + \mu_{0,3} - \mu_{1,2} - \mu_{2,1} & \mu_{4,0} - \mu_{2,0} + \mu_{0,4} - \mu_{0,2} - 2\mu_{2,2} + 2\mu_{0,2}\mu_{2,0}
\end{bmatrix} \]

And so, by the Lindeberg-Levy central limit theorem,

\[ \sqrt{n} [\bar{x} - \bar{y}, \sigma^2_x - \sigma^2_y]' \xrightarrow{d} \mathcal{N}_2 \left( [0,0]', \Xi \right). \]

While the normality does not immediately follow as we have to apply tedious transformations to use the Lindeberg condition, such transformations are beyond the scope of this paper, and do not provide any useful intuition.❑
Lemma 5 (General Limiting Distribution of the Skewness) Where \( \{(x_i, y_i)\}_{i=1,\ldots,n} \) is a random sample from a bivariate distribution with finite sixth moments, under the null that \( x \) and \( y \) have equal skewness,

\[
\sqrt{n} \left( \hat{S}_x - \hat{S}_y \right) \xrightarrow{d} \mathcal{N} (0, \omega_x + \omega_y - 2 \omega_{xy}) ,
\]

where

\[
\omega_x = \frac{1}{4 \mu_{2,0}^5} \left( 36 \mu_{2,0}^5 - 24 \mu_{4,0} \mu_{2,0}^3 + 35 \mu_{2,0}^2 \mu_{3,0}^2 + 4 \mu_{6,0} \mu_{2,0}^2 - 12 \mu_{5,0} \mu_{2,0} \mu_{3,0} + 9 \mu_{4,0} \mu_{3,0}^2 \right) ,
\]

and

\[
\omega_{xy} = \frac{1}{4 \mu_{2,0}^5} \left( 12 \mu_{2,0}^2 \mu_{3,1} - 18 \mu_{2,1} \mu_{3,0} - 18 \mu_{2,0} \mu_{0,3} \mu_{1,2} + 12 \mu_{0,3} \mu_{3,0}^2 \right) \\
+ \frac{1}{4 \mu_{2,0}^5} \left( -36 \mu_{2,0} \mu_{1,1} \mu_{2,0}^2 - 4 \mu_{0,2} \mu_{2,0} \mu_{3,3} + 6 \mu_{0,2} \mu_{3,0} \mu_{2,3} + 6 \mu_{2,0} \mu_{0,3} \mu_{3,2} \right) \\
+ \frac{1}{4 \mu_{2,0}^5} \left( -9 \mu_{0,3} \mu_{3,0} \mu_{2,2} + \mu_{0,2} \mu_{2,0} \mu_{0,3} \mu_{3,0} \right) .
\]

Proof of Lemma 5

By Taylor expanding around the expectations of the moment estimators, we obtain the approximation

\[
\hat{S}_x = \frac{\mu_{3,0}}{\mu_{2,0}^{3/2}} + \frac{1}{\mu_{2,0}^{3/2}} \left( \bar{m}_{3,0} - \mu_{3,0} \right) - \frac{3}{2 \mu_{2,0}^{5/2}} \left( \hat{\sigma}_x^2 - \mu_{2,0} \right) + O \left( n^{-1} \right) ,
\]

where \( \bar{m}_{3,0} = n^{-1} \Sigma_{i=1}^n (x_i - \bar{x})^3 \). Notice that \( \text{E}(\bar{m}_{3,0}) = \mu_{3,0} \). So it is obvious that \( \text{E}(\hat{\kappa}_3) = \kappa_3 + O \left( n^{-1} \right) \). So, the only thorny issue is deriving the variance-covariance matrix of \( [\hat{\kappa}_3, \hat{\kappa}_3'] \),

\[
\text{Var} [\hat{S}_x] = \frac{1}{\mu_{2,0}^3} \text{Var} [\bar{m}_{3,0}] + \frac{9}{4 \mu_{2,0}^2} \text{Var} [\hat{\sigma}_x^2] - \frac{3 \mu_{3,0}}{\mu_{2,0}^4} \text{Cov} [\bar{m}_{3,0}, \hat{\sigma}_x^2] + O \left( n^{-2} \right) .
\]

Although the calculation is rather tedious, it is straightforward to obtain the following,

\[
\text{Var} [\bar{m}_{3,0}] = n^{-1} \left( \mu_{6,0} - \mu_{3,0}^2 - 6 \mu_{2,0} \mu_{4,0} + 9 \mu_{3,0}^3 \right) + O \left( n^{-2} \right) ,
\]

\[
\text{Cov} [\bar{m}_{3,0}, \hat{\sigma}_x^2] = n^{-1} \left( \mu_{5,0} - 4 \mu_{2,0} \mu_{3,0} \right) + O \left( n^{-2} \right) .
\]
Hence,

\[
\text{Var} \left[ \hat{S}_x \right] = \frac{1}{4n\mu_2^0} \left( 36\mu_2^5 - 24\mu_4\mu_2^2 + 35\mu_2^3 + 4\mu_6 - 12\mu_5\mu_2^2 + 9\mu_4\mu_2^2 \right) + \frac{1}{4n\mu_2^0} \left( -12\mu_5\mu_2 + 9\mu_4\mu_2^2 \right) + O \left( n^{-2} \right).
\]

To find the covariance, we compute:

\[
\text{Cov} \left[ \bar{m}_{3,0}, \bar{m}_{0,3} \right] = \frac{1}{n} \left( \mu_{3,3} - \mu_3\mu_{3,0} - 3\mu_{2,0}\mu_{3,1} - 3\mu_{1,3}\mu_{2,0} + 9\mu_{0,2}\mu_{1,1}\mu_{2,0} \right) + O \left( n^{-2} \right),
\]

\[
\text{Cov} \left[ \bar{m}_{3,0}, \bar{\sigma}_y^2 \right] = \frac{1}{n} \left( \mu_{3,2} - \mu_2\mu_{3,0} - 3\mu_{1,2}\mu_{2,0} \right) + O \left( n^{-2} \right).
\]

So,

\[
\text{Cov} \left[ \hat{S}_x, \hat{S}_y \right] = \frac{1}{4n\mu_{5/2}^0\mu_{5/2}^0} \left( 12\mu_2^2\mu_{2,0}\mu_{3,1} - 18\mu_2^2\mu_{2,1}\mu_{3,0} - 18\mu_2^2\mu_{0,3}\mu_{1,2} + 12\mu_2^2\mu_{0,2}\mu_{1,3} \right) + \frac{1}{4n\mu_{5/2}^0\mu_{5/2}^0} \left( -36\mu_2^2\mu_1\mu_{2,0} - 4\mu_0\mu_2\mu_{2,0}\mu_{3,0} + 6\mu_0\mu_3\mu_{2,3} + 6\mu_0\mu_3\mu_{2,3} \right) + \frac{1}{4n\mu_{5/2}^0\mu_{5/2}^0} \left( -9\mu_{3,0}\mu_{2,0}\mu_{3,0} + \mu_{2,0}\mu_{2,0}\mu_{3,0} \right) + O \left( n^{-2} \right).
\]

Thus, under the null that \( X \) and \( Y \) have equal skewness,

\[
\sqrt{n} \left( \hat{S}_x - \hat{S}_y \right) \xrightarrow{d} \mathcal{N} \left( 0, \omega_x + \omega_y - 2\omega_{xy} \right),
\]

and the lemma is proved. \( \square \)