Using goodness-of-fit tests to detect normality for mesokurtic data

Micaela Lee Sclanders

17140057

Supervisor: Dr P.J van Staden

Submitted in partial fulfilment of MSc Advanced Data Analytics

degree in the Faculty of Natural and Agricultural Sciences

Department of Statistics, University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

February 10, 2023

Abstract

The normality assumption is crucial in statistical inference and modelling. It is therefore important to determine if a sample comes from a normal distribution. Consequently, numerous goodness-of-fit hypothesis tests have been developed. The practical use of a specific test depends on its availability in the statistical software package considered. The field of application will also contribute to the choice of test. For instance, the Jarque-Bera test is popular in econometrics and finance.

Various power comparison and simulation studies of goodness-of-fit tests for normality can be found in the literature. Typically, these studies select alternative distributions whose levels of skewness and kurtosis deviate from that of the normal distribution. I.e., symmetric and asymmetric distributions exhibiting leptokurtosis or platykurtosis are included in these simulation studies. In this mini-dissertation, the focus is on mesokurtic distributions whose levels of skewness and kurtosis are equivalent to that of the normal distribution, but with different distributional shapes compared to the normal distribution.

Declaration

I, *Micaela Lee Sclanders,* declare that this mini-dissertation, submitted in partial fulfillment of the degree *MSc Advanced Data Analytics* , at the University of Pretoria, is my own work and has not been previously submitted at this or any other tertiary institution.

Acknowledgements

I would like to sincerely thank Beat Knechtle and Fabio Valeri for giving us access to the data on 100-km ultra-marathon race times. I would also like to thank the Postgraduate Scholarship Office of the University of Pretoria for their financial assistance. I thank Paul van Staden for his assistance and advice. And a special thanks to my family for their love and support.

We thank the NAS research ethic committee for approving our ethics application with ethics number NAS124/2019.

Contents

List of Figures

List of Tables

List of Abbreviations

AD	Anderson-Darling
BHEP	Baringhaus-Henze-Epps-Pulley
CDF	Cumulative distribution function
CvM	Cramér-von Mises
EDF	Empirical distribution function
JB	Jarque-Bera
KS	Kolmogorov-Smirnov
PDF	Probability density function
QF	Quantile function
SW	Shapiro-Wilk

Chapter 1

Introduction

Introductory example

To introduce the topic of testing for normality, an example on 100-km ultra-marathon race times is investigated. This example comes from an in-depth study on ultra-marathon runners' race times from various countries. This study was completed by Knechtle et al. (2018). The investigation carried out in this mini-dissertation focuses on only Hong-Kong 100-km marathon runners' race times. Note that all statistical and graphical analyses were carried out using Wolfram Mathematica, Version 12.3 (Wolfram Research, Champaign, IL, USA).

The first step in any statistical investigation is to plot the data. This is done in Figures 1.1, 1.2, and 1.3. To determine if a sample comes from a normal distribution, visually, a boxplot, histogram and Q-Q plot can be drawn and compared to the normal distribution. Figure 1.1, the boxplot for the race times of 100-km marathon runners from Hong-Kong, indicates that the data is symmetric around the sample mean, 20.1388. Figure 1.2 is the Q-Q plot which suggests that the data might be compatible with the normal distribution. Pearson's moment coefficients of skewness and kurtosis, in Figure 1.4, are -0.0716 and 2.8968, respectively. These sample values do not differ substantially from the normal distribution's theoretical moment coefficients of skewness and kurtosis of $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$, respectively. Hence, one may assume that the Hong-Kong runners' race times data is normally distributed.

With further investigation it is clear that the data is bimodal since it has two distinct peaks in the histogram, in Figure 1.3, and therefore cannot be normally distributed. If this is indeed the case and the data is not normally distributed, we expect all the goodness-of-fit tests performed to reject normality. However, as seen in the results in Figure 1.4, this is not the case as not all the tests reject normality. The discrepancies between the test results prompted this investigation into goodness-of-fit testing for normality, focusing on mesokurtic data and distributions whose levels of skewness and kurtosis are the same as that of the normal distribution, but with different distributional shapes and characteristics compared to the normal distribution.

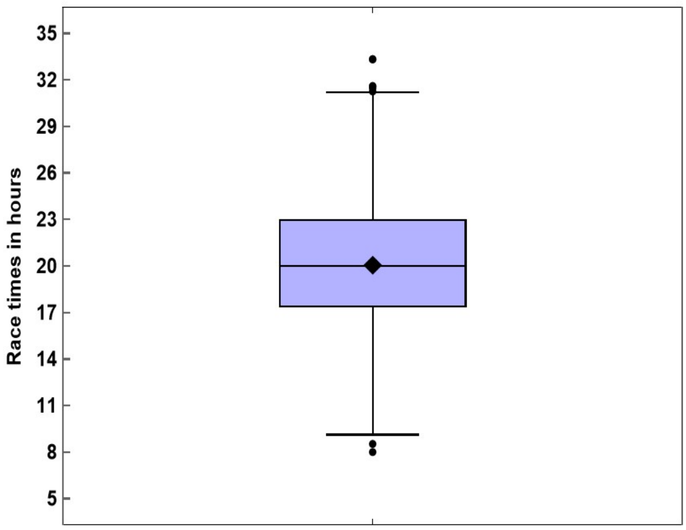


Figure 1.1: Boxplot for the race times of 100-km marathon runners from Hong-Kong

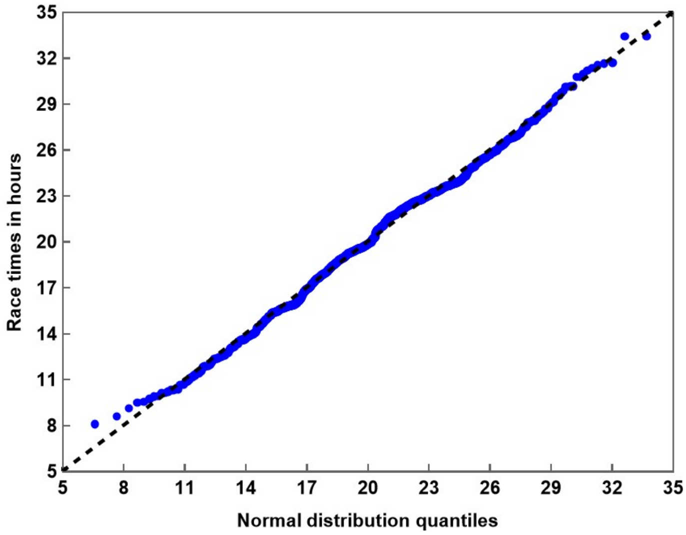


Figure 1.2: Q-Q plot for the race times of 100-km marathon runners from Hong-Kong

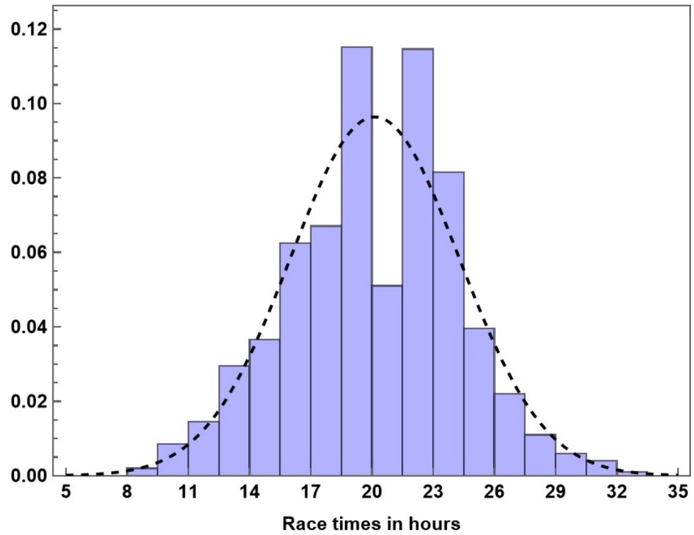


Figure 1.3: Histogram for the race times of 100-km marathon runners from Hong-Kong with normal density curve

Sample size: 1332

Mean: 20.1388

Standard deviation: 4.13892

Pearson's moment coefficient of skewness: -0.0715894

Pearson's moment coefficient of kurtosis: 2.89679

Test	Statistic	P-value
Anderson-Darling	2.75325	0.00000
Baringhaus-Henze	4.51030	0.00003
Cramér-von Mises	0.51793	0.00000
Jarque-Bera ALM	1.68976	0.42168
Kolmogorov-Smirnov	0.05112	0.00000
Kuiper	0.08382	0.00000
Mardia Combined	1.68976	0.42168
Mardia Kurtosis	-0.76889	0.44196
Mardia Skewness	1.13776	0.28613
Pearson χ^2	170.16200	0.00000
Shapiro-Wilk	0.99524	0.00033
Watson U^2	0.49924	0.00001

Figure 1.4: Descriptive statistics and goodness-of-fit tests for normality for the race times of 100-km marathon runners from Hong-Kong

Chapter 2

Literature review

Pearson (1900) introduced the first goodness-of-fit test known as the Pearson χ^2 test. Since the introduction of this test, many tests to assess the fit of distributions to data have been proposed and studied. Goodness-of-fit testing is now a prominent area of statistical inference. There are numerous normality tests that can be found in the literature. As new tests are introduced or existing tests modified, a simulation study is conducted to compare the power of the tests. Stephens (1974) conducted a simulation study on 5 goodness of fit tests, and tested modifications of these tests where the distribution is normal or exponential with known and unknown parameters. In the paper, Stephens (1974) explored a power comparison on 1000 Monte-Carlo samples to test for uniformity and normality. It was found that it is better to estimate the mean and variance from the data rather than to have the true mean and variable already available.

D'Agostino and Stephens (1986) is a book devoted to the presentation and discussion of goodness-of-fit techniques. It was written with the intention of putting together what had been done in the literature. Included in the book are methods of testing for specific well-known distributions, such as the normal distribution and the exponential distribution, and grouping of the tests according to techniques such as chi-squared tests, empirical distribution function tests, and tests based on probability plotting. Each test group is discussed in depth, tests such as the chi-squared test, Kuiper test, Watson U^2 test, Anderson-Darling test as well as some modifications of these tests are included. Once the tests were discussed, a power comparison study was conducted. Power results for symmetric distributions with non-normal kurtosis (platykurtic and leptokurtic) concluded that the R test and D'Agostino's D test are the most powerful tests. The Kolmogorov-Smirnov test and chi-squared test both have poor power and should not be used for testing normality for non normal distributions.

Romão et al. (2010) carried out a comprehensive power comparison study on 33 goodness-of-fit tests. The power comparison simulation study was based on 3 different sample sizes n ($n = 25$, $n = 50$ and $n = 100$) and several significance levels. Included in the study were symmetric, asymmetric, and modified normal distributions as alternative distributions. Romão et al. (2010) discussed the various power results considering different levels of comparison. Firstly, the levels of skewness and kurtosis, $\sqrt{\beta_1}$ and β_2 respectively, were observed in comparison to the normal distribution. It was found that for symmetric distributions, when β_2 is vastly different (higher or lower) to the kurtosis level of the normal distribution, tests perform better. As skewness increases so does the power of the tests for asymmetric distributions and finally, the effect of skewness and kurtosis levels for modified normal distributions is unidentified. Furthermore, considering the normality tests, the tests based on regression and correlation perform best for symmetric distributions and modified normal distributions. Tests based on the empirical distribution function perform best for asymmetric distributions.

Yap and Sim (2011) studied and compared 8 normality tests for symmetric short-tailed and long-tailed distributions and asymmetric distributions. In the simulation study, 10000 samples were used and a significance level of $\alpha = 0.05$ was considered. They concluded that the Shapiro-Wilk test is the most powerful test for asymmetric distributions, that for symmetric short-tailed (low kurtosis level) distributions the D'Agostino-Pearson test and Shapiro-Wilk test are the most powerful and lastly, for symmetric long-tailed (high kurtosis levels) distributions the Jarque-Bera, Shapiro-Wilk and Anderson-Darling have the most power.

More recently, Domański and Szczepocki (2020) wrote a power comparison study to include interesting and more recently developed tests using a simulation of 10000 samples for 5 sample sizes n ($n = 10$, $n = 20$, $n = 50$, $n = 100$ and $n = 500$) and significance levels $\alpha = 0.01, 0.05$ and 0.1. Symmetric short-tailed, symmetric long-tailed and asymmetric distributions were used in the power comparison. For the alternative distributions considered, the D'Agostino-Pearson test and the Lafaye de Micheaux's X_{APD} test are the most powerful tests for symmetric short-tailed distributions and the Robust Jarque-Bera test has the most power for symmetric long-tailed distributions. For asymmetric distributions, it is difficult to choose one test.

Anastasiou et al. (2020) evaluated goodness-of-fit tests using simulated data versus empirical data. They compared the performance of 10 tests for 10 different sample sizes for symmetric, asymmetric, and bimodal and contaminated alternative distributions to the normal distribution. The simulation study used 10000 simulations generated from the normal distribution for various sample sizes considering significance levels $\alpha = 0.01$ and $\alpha = 0.05$. The symmetric distributions resulted in the D'Agostino-Pearson Omnibus test having the highest power, followed by correlation tests. For asymmetric distributions, the Shapiro-Wilk test performs the best and the Skewness test performs the worst. Finally, for bimodal and contaminated alternative distributions, the Shapiro-Wilk test, Shapiro-Francia test and the D'Agostino-Pearson test have the highest power. Furthermore, Anastasiou et al. (2020) emphasized that the power of the test increases with increasing the sample size and the significance levels.

The most recent power study was conducted by Uyanto (2022). This is an extensive comparison of 50 normality tests for various samples sizes and a significance level of $\alpha = 0.05$. Symmetric and asymmetric distributions with bounded or infinite support were considered as alternative distributions. The simulation study was conducted in R using 10000 simulations for sample sizes n ($n = 10$, $n = 30$, $n = 50$, $n = 70$ and $n = 100$) for the alternative distributions and $\alpha = 0.05$ level of significance. The conclusion of the study was that the Robust Jarque-Bera and Gel-Mia-Gastwirth tests are the most powerful tests for symmetric distributions with infinite support, the 2nd Zhang-Wu test has the most power for asymmetric distributions on infinite support and for distributions with half-infinite support $(0, \infty)$ and for distributions with bounded support on $(0, 1)$ the 1st Zhang and 2nd Zhang-Wu tests are the most powerful.

When hypothesis tests for normality are conducted, one must take into account what software is being used along with the area of application. Yap and Sim (2011) mentioned the availability of normality tests in several software packages. This must be considered as software packages have built-in functions for selected goodness-of-fit tests and will not include all tests. In older literature, for example in Gan and Koehler (1990), the number of samples simulated in power comparison studies were much smaller due to computational strength. Nowadays, simulation procedures are not computationally expensive and much larger simulation sizes can be utilized such as in Uyanto (2022). Furthermore, tests have characteristics that make them useful in certain fields of research. For instance, the Jarque-Bera test is a popular test for testing normality of regression residuals which is useful in econometrics (Thadewald and Büning, 2007).

Chapter 3

Normality tests

Goodness-of-fit testing is a statistical procedure that makes assumptions about observed values. These tests assist in determining if a sample follows a normal distribution, which is the focus of this mini-dissertation. However, these tests can also check if categorical variables are related or if random samples are from the same distribution.

The 11 goodness-of-fit tests for normality considered in this study are briefly presented below. Where applicable in the expressions given, let $X_1, X_2, \cdots, X_n$ denote a random sample size of n with $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ the order statistics of this sample and where $\bar{X}$, S, $\sqrt{b_1}$ and b_2 are the sample mean, sample standard deviation and Pearson's moment coefficients of skewness and kurtosis obtained from the sample. Let $Z_i = \Phi\left(\frac{X_{(i)}-\bar{X}}{S}\right)$ denote the cumulative distribution function (CDF) of the standard normal distribution.

3.1 Empirical distribution function and characteristic function tests

3.1.1 Kolmogorov-Smirnov test

The Kolmogorov-Smirnov (KS) test of Kolmogorov (1933) and Smirnov (1939) is a nonparametric, distribution-free test based on the empirical distribution function (EDF). The KS test statistic is given by

$$KS = \max\left(D^+, D^-\right) \tag{3.1}$$

where

$$D^+ = \max_{1 \le i \le n}\left\{\frac{i}{n} - Z_i\right\} \text{ and } D^- = \max_{1 \le i \le n}\left\{Z_i - \frac{i-1}{n}\right\}. \tag{3.2}$$

D^+ represents the largest positive deviation and D^- the largest negative deviation between the CDF of the hypothesised distribution and the EDF (Thas, 2010). This test can be applied to continuous distributions and is more sensitive near the centre of the distribution than around the tails (Lilliefors, 1967).

A well-known modification of this test, and the modification used by Wolfram Mathematica, is that of Lilliefors (1967). The KS test is used for simple hypotheses and the Lilliefors test is used for composite hypotheses. The test statistic can be calculated using KS, in Equation 3.1, and the sample mean and standard deviation are estimated from the data. This modification generally gives a more accurate approximation of the test statistic's distribution.

3.1.2 Cramér-von Mises test

The Cramér-von Mises (CvM) test, named after Cramér (1928) and von Mises (1931), is an alternative test to the Kolmogorov-Smirnov test. The CvM test assumes that the data comes from a continuous distribution. The test statistic of this distribution-free test is based on the EDF and given by

$$CvM = \frac{1}{12n} + \sum_{i=1}^{n} \left(\frac{2i-1}{2n} - Z_i \right)^2 . \tag{3.3}$$

The Cramér-von Mises test may be modified to the test statistic

$$CvM^* = CvM \left(1.0 + \frac{0.5}{n} \right),$$

as suggested by Stephens (1974). The reason for this modification is to improve the power of the test. However, the test statistic in Equation 3.3 is used by Wolfram Mathematica.

3.1.3 Anderson-Darling test

Anderson and Darling (1952, 1954) proposed the Anderson-Darling (AD) test statistic

$$AD = n \int_{-\infty}^{\infty} \left(F_n(x) - \Phi(x) \right)^2 \Psi(F(x)) dF(x),$$

which is based on the EDF, $F_n(x)$, and can be used to test how well a distribution fits a dataset. $\Phi(x)$ is the CDF of the standard normal distribution and $\Psi(F(x)) = [F(x)(1 - F(x))]^{-1}$ is the weight function. The weight function is used in order to emphasize deviations in the lower and upper tails. The AD test statistic can be reduced to the CvM test by taking $\Psi(F(x)) = 1$ and hence, is also an alternative of the KS test. The AD test gives more weight to the tails of the distribution compared to the CvM test. Hence, it is equally sensitive at the tails as it is at the centre. The AD test statistic given by Anderson and Darling (1954) that is considered in this study and that is used by Wolfram Mathematica is

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[ln(Z_i) + ln(1 - Z_{n-i+1})\right].$$

Stephens (1974) modified the AD test statistic to be

$$AD^* = AD\left[1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right],$$

which takes into account the sample size and is more powerful than AD.

3.1.4 Kuiper V test

The Kuiper test statistic, proposed by Kuiper (1960), is given by

$$V = D^+ + D^-,$$

where D^+ and D^- are given in Equation 3.1. The Kuiper test can be associated to the KS test since it uses D^+ and D^- but the Kuiper test measures the sum of D^+ and D^- whereas the KS test considers the maximum of the two. This test is equally sensitive in the tails as the centre, unlike the KS test. The Kuiper test is unique in testing for cyclic variations and testing the fit of circular probability distributions. In this study, the focus is on using the Kuiper test for testing for normality. Wolfram Mathematica uses V and assumes that the data comes from a continuous distribution.

3.1.5 Watson U^2 test

The Watson U^2 test, suggested by Watson (1961), is a robust test used as a general test for uniformity. The test statistic for the Watson U^2 test for normality is

$$U^2 = CvM - n\left(\bar{Z} - \frac{1}{2}\right)^2,$$

where CvM is the Cramér-von Mises test statistic, see Equation 3.3, and $\bar{Z} = \sum_{i=1}^{n}\frac{Z_i}{n}$. This U^2 test statistic is used by Wolfram Mathematica and assumes that the data comes from a continuous distribution.

The Watson U^2 test was modified to the test statistic

$$U^{2*} = U^2\left(1.0 + \frac{0.5}{n}\right),$$

by Stephens (1974) in order to improve the power.

3.1.6 Baringhaus-Henze-Epps-Pulley test

The Baringhaus-Henze-Epps-Pulley (BHEP) test, initially suggested by Epps and Pulley (1983), has the ability to detect neighbouring alternatives which converge to a normal distribution. The test statistic is

$$T_{EP} = \int_{-\infty}^{\infty} \left| \varphi_n(t) - \hat{\varphi_0(t)} \right|^2 dG(t),$$

where $\varphi_n(t)$ is the empirical characteristic function given by $n^{-1} \sum_{j=1}^{n} e^{(itX_j)}$, $\varphi_0(t)$ is the sample estimate of the characteristic function of the normal distribution given by $e^{(it\bar{X}-0.5S^2t^2)}$ and $G(t)$ is a weight function selected based on various criteria discussed by Epps and Pulley (1983). The test statistic can be written as

$$T_{EP} = 1 + \frac{n}{\sqrt{3}} + \frac{2}{n} \sum_{k=2}^{n} \sum_{j=1}^{k-1} e^{-\frac{(X_j - X_k)^2}{2S^2}} - \sqrt{2} \sum_{j=1}^{n} \exp\left\{ -\frac{(X_j - \bar{X})^2}{4S^2} \right\}.$$

Baringhaus and Henze (1988) extended the test statistic for multivariate normality. The $BHEP$ test statistic for univariate normality, used by Wolfram Mathematica, is given by

$$BHEP = n^{-2} \sum_{j=1}^{n} \sum_{k=1}^{n} \exp\left\{ \frac{-(X_j - X_k)^2}{2b^2S^2} \right\} - 2n^{-1}(1+b^{-2})^{-\frac{1}{2}} \sum_{j=1}^{n} \exp\left\{ \frac{-(X_j - \bar{X})^2}{2S^2(1+b^2)} \right\} + (1+2b^{-2})^{-\frac{1}{2}}.$$

In the BHEP test statistic b is the smoothing parameter. The performance of this test is very sensitive to the choice of b. Following Henze and Zirkler (1990), the smoothing parameter can be calculated as a function of the sample size using

$$b = \frac{1}{\sqrt{2}} \left(\frac{2d+1}{4} \right)^{\frac{1}{d+4}} n^{\frac{1}{d+4}}, \tag{3.4}$$

where $d \geq 1$ is the number of dimensions in the multivariate case. Thus, for the univariate case, $d = 1$ so that $b = 0.66757n^{\frac{1}{5}}$.

Because of the importance in the choice of b, an investigation on the effects of b on the BHEP test will be studied in Chapter 6. This is possible since Wolfram Mathematica allows users to either use the default method with b calculated by Equation 3.4 or to set the smoothing parameter to a value of their choice.

3.2 Moment-based tests

3.2.1 Jarque-Bera test

The Jarque-Bera (JB) test is a popular goodness-of-fit test for testing normality in econometrics and finance. This test is a moment-based test and was recognised because of the proposal by Jarque and Bera

(1987). The test statistic is

$$JB = \frac{n}{6}\left[(\sqrt{b_1})^2 + \frac{(b_2-3)^2}{4}\right],$$

where $\sqrt{b_1}$ is the sample skewness and b_2 the sample kurtosis.

Various modifications have been made to the JB test in order to increase its efficiency. The modification put forward by Doornik and Hansen (2008), makes use of a transformation of the skewness and kurtosis formulas. The modified test statistic is

$$JB_{DH} = \left[Z(\sqrt{b_1})\right]^2 + \left[\left[\left(\frac{\xi}{2a}\right)^{\frac{1}{3}} - 1 + \frac{1}{9a}\right](9a)^{\frac{1}{2}}\right]^2,$$

where ξ and a can be obtained by the following expressions:

$$\xi = 2k(b_2 - 1 - b_1),$$

$$k = \frac{(n+5)(n+7)(n^3+37n^2+11n-313)}{12(n-3)(n+1)(n^2+15n-4)},$$

and

$$a = \frac{(n+5)(n+7)\left[(n-2)(n^2+27n-70)+b_1(n-7)(n^2+2n-5)\right]}{6(n-3)(n+1)(n^2+15n-4)}.$$

Urzúa (1996) introduced a modification to the JB test by standardizing the skewness and kurtosis. The following test statistic is obtained

$$JB_{ALM} = \left[\frac{(\sqrt{b_1})^2}{var(b_1)} + \frac{(b_2 - E(b_2))^2}{var(b_2)}\right],$$

with

$$E(b_2) = \frac{3(n-1)}{(n+1)},$$

$$var(b_1) = \frac{6(n-2)(n+1)}{(n+1)(n+3)},$$

and

$$var(b_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}.$$

This modification, also known as the adjusted Lagrange Multiplier (ALM) Jarque-Bera test, is the version of the JB test used by Wolfram Mathematica.

A more recent modification is the robust Jarque-Bera test proposed by Gel and Gastwirth (2008). The reason for this modification is that the sample moments are known to be sensitive to outliers. This test uses a powerful estimate of the dispersion in the skewness and kurtosis, with the resulting test statistic given by

$$JB_R = \frac{n}{6}\left(\frac{m_3}{\left[\frac{\sqrt{\frac{\pi}{2}}}{n}\sum_{i=1}^{n}\left|X_i-\bar{X}\right|\right]^3}\right)^2 + \frac{n}{64}\left(\frac{m_4}{\left[\frac{\sqrt{\frac{\pi}{2}}}{n}\sum_{i=1}^{n}\left|X_i-\bar{X}\right|\right]^4}-3\right)^2,$$

with $m_3 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^3$, the third central moment, and $m_4 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^4$, the fourth central moment.

3.2.2 Mardia skewness and kurtosis tests

With the Mardia skewness test for normality, the measure of multivariate skewness proposed by Mardia (1970),

$$MS = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\widehat{D_{ij}}^3,$$

with $\widehat{D_{ij}} = (X_i-\bar{X})S^{-1}(X_j-\bar{X})$, is compared to the corresponding population counterpart.

Mardia (1970) also suggested

$$MK = \frac{1}{n}\sum_{i=1}^{n}\widehat{D_i^2}^2,$$

with $\widehat{D_i^2} = (X_i-\bar{X})S^{-1}(X_i-\bar{X})$ as a measure of multivariate kurtosis. With the Mardia kurtosis test this measure is again compared to the corresponding population counterpart. Since these measures were introduced, they have become standard characteristics of a multivariate distribution (Kollo, 2008).

Both the Mardia skewness and kurtosis tests can also be applied to univariate data as done in Wolfram Mathematica and hence in this study.

3.3 Other tests

3.3.1 Pearson's χ^2-test

The Pearson χ^2-test, proposed by Pearson (1900), is the oldest goodness-of-fit test for normality. The test statistic is represented as

$$P = \sum_{k=1}^{m}\frac{(O_k-E_k)^2}{E_k}.$$

where k represents the bin number, $O_k : k = 1, \cdots, m$ is the observed count and $E_k : k = 1, \cdots, m$ is the expected count. The bins are chosen to be equiprobable under the null hypothesis of normality. A disadvantage of this test is that it is not highly recommended for continuous distributions. This is because the test counts the number of observations in each bin rather than using the values of the observations.

Schorr (1974) found that the optimum number of bins is $m \leq 4\left(\frac{2n^2}{z_\alpha^2}\right)^{\frac{1}{5}}$, where z_α is the $100(1-\alpha)th$ percentile for the standard normal distribution. In order to determine the number of bins in the test statistic calculation, the formula $m = 2n^{\frac{2}{5}}$ is used (Moore, 1986). Wolfram Mathematica uses the default $m-3$ degrees of freedom for the χ^2 distribution to compute the p-values.

3.3.2 Shapiro-Wilk test

The Shapiro-Wilk (SW) test proposed by Shapiro and Wilk (1965) is a well-known, powerful test specifically introduced for testing normality and can be described as the proportion of a pair of variance estimates for the normal distribution. This test statistic is defined as

$$SW = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2},$$

where the coefficients $a_i : i = 1, \cdots, n$ are given by

$$\mathbf{a}' = \frac{\mathbf{m}'.\mathbf{V}^{-1}}{(\mathbf{m}'.\mathbf{V}^{-1}\mathbf{m})^{\frac{1}{2}}}.$$

where $\mathbf{m}'$ is the mean vector and $\mathbf{V}$ the covariance matrix of the standard normal distribution's order statistics.

The SW test was modified by Royston (1992) to be used for larger sample sizes. Wolfram Mathematica makes use of this modification.

Chapter 4

Mesokurtic distributions

Statistical distributions are important tools that are used to describe how values are dispersed in a dataset. Numerous distributions are considered and are used in the evaluation of the performance of normality tests in this simulation study. In the literature, most studies focus on leptokurtic and platykurtic distributions. The uniqueness of this study is the type of distributions that are included, these being mesokurtic distributions. A mesokurtic distribution has skewness and kurtosis levels that are the same as for the normal distribution, that is $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$ for Pearson's moment coefficients of skewness and kurtosis, respectively. In order to obtain mesokurtic distributions for the various distributions considered in this study, Pearson's moment coefficients of skewness and kurtosis were set to 0 and 3, respectively, and then the unknown location, scale and shape parameters were solved for in Wolfram Mathematica. By doing this we were able to obtain several distributions that are indeed mesokurtic.

4.1 Normal distribution

A long-standing, well-known distribution in statistics is the normal distribution. It was initially developed as an approximation to the binomial distribution in a pamphlet by De Moivre (1733). The normal distribution is often referred to as the Gaussian distribution in literature, because Gauss made significant contributions towards its development (Gauss, 1809). This continuous distribution is bell-shaped and symmetric around the mean, showing that data points around the mean appears more frequently than data points far from the mean. The normal distribution has probability density function (PDF), $f(x)$, and CDF, $F(x)$, given respectively by,

$$f(x) = \left(\frac{1}{\beta\sqrt{2\pi}}\right) e^{-\frac{(x-\alpha)^2}{(2\beta^2)}}, \quad -\infty < x < \infty,$$

and

$$F(x) = \Phi\left(\frac{x-\alpha}{\beta}\right), \quad -\infty < x < \infty,$$

where α is the mean and β is the standard deviation. Note that these location and scale parameters of the normal distribution are typically denoted by μ and σ in the literature. However, to allow for comparable notation of parameters across the distributions considered, we will use α and β.

The normal distribution's inverse CDF, also known as the quantile function (QF), is

$$Q(p) = \alpha + \sqrt{2}\beta erf^{-1}(2p-1), \; 0 < p < 1,$$

where erf is the error function. The r^{th} order uncorrected moment can be found using the following equation,

$$E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx.$$

The normal distribution is a bell-shaped, symmetric distribution with infinite support, $(-\infty; \infty)$. In this study we consider the normal distribution in its standard form. The standard normal distribution has location parameter $\alpha = 0$ and scale parameter $\beta = 1$. The standard normal distribution will be included in the simulation studies in Chapters 5 and 6 to enable comparisons with the other mesokurtic distributions. Wolfram Mathematica has the normal distribution as a built-in model and all that has to be done is to set the parameters $\alpha = 0$ and $\beta = 1$.

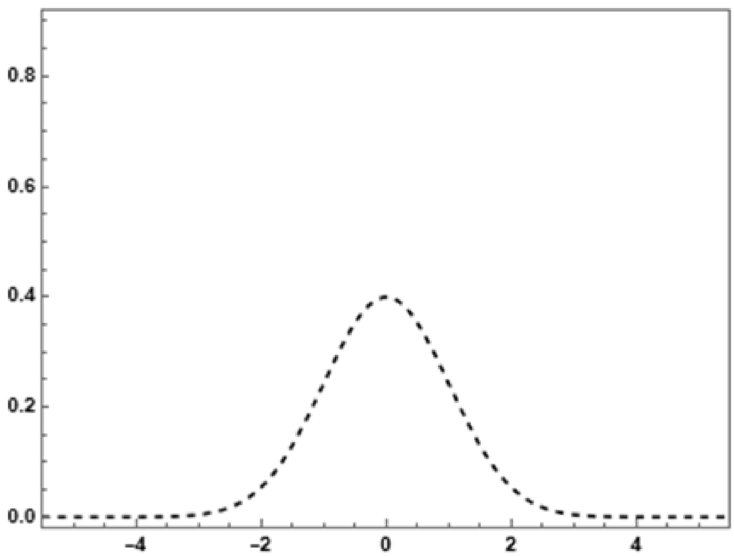


Figure 4.1: Probability density curve of the standard normal distribution

Figure 4.1 shows the probability density curve of the standard normal distribution. This density curve will be used as a reference curve when studying the alternative mesokurtic distributions.

4.2 Generalized secant hyperbolic pi distribution

The hyperbolic secant distribution, proposed by Baten (1934) and also studied by Talacko (1956), was later generalized by Vaughan (2002) to produce the generalized secant hyperbolic distribution. This continuous distributional family is symmetric, but includes members that are leptokurtic or platykurtic. Consequently, as explained by Vaughan (2002), the generalized secant hyperbolic distribution is useful for modelling data with heavy or thin tails.

If a random variable X has a generalized secant hyperbolic distribution, then the PDF, CDF and QF of X are given respectively by,

$$f(x) = \begin{cases} \frac{1}{2\beta\lambda}\frac{\sin(\lambda)}{\cos(\lambda)+\cosh(\frac{x-\alpha}{\beta})}, & -\pi < \lambda < 0, \\ \frac{1}{\beta}\frac{e^{-\frac{x-\alpha}{\beta}}}{\left(1+e^{-\frac{x-\alpha}{\beta}}\right)^2}, & \lambda = 0, \\ \frac{1}{2\beta\lambda}\frac{\sinh(\lambda)}{\cosh(\lambda)+\cosh\left(\frac{x-\alpha}{\beta}\right)}, & \lambda > 0, \end{cases}$$

$$F(x) = \begin{cases} \frac{1}{\lambda} arccot\left(\cot(\lambda)+\csc(\lambda)e^{(-\frac{x-\alpha}{\beta})}\right), & -\pi < \lambda < 0, \\ \frac{1}{1+e^{(-\frac{x-\alpha}{\beta})}}, & \lambda = 0, \\ \frac{1}{\lambda} arccoth\left(\coth(\lambda)+csch(\lambda)e^{(-\frac{x-\alpha}{\beta})}\right), & \lambda > 0, \end{cases}$$

$$Q(p) = \begin{cases} \alpha+\beta\log\left[\frac{\sin(\lambda p)}{\sin(\lambda(1-p))}\right], & -\pi < \lambda < 0, \\ \alpha+\beta\log\left[\frac{p}{1-p}\right], & \lambda = 0, \\ \alpha+\beta\log\left[\frac{\sinh(\lambda p)}{\sinh(\lambda(1-p))}\right], & \lambda > 0, \end{cases}$$

where α is the location parameter, $\beta > 0$ is the scale parameter and $\lambda > -\pi$ is a shape parameter. As with Vaughan (2002), our focus will be on the standard generalized secant hyperbolic distribution with $\alpha = 0$ and $\beta = 1$ so that $E(X) = 0$ and $Var(X) = 1$. The generalized secant hyperbolic distribution is always symmetric irrespective of the value of λ (Vaughan, 2002). That is, $\sqrt{\beta_1} = 0$. Klein and Fischer (2007) proved that λ is a kurtosis parameter, while Vaughan (2002) showed that Pearson's moment coefficient of kurtosis for the generalized secant hyperbolic distribution is

$$\beta_2 = \begin{cases} \frac{21\pi^2-9\lambda^2}{5\pi^2-5\lambda^2}, & -\pi < \lambda < 0, \\ 4.2, & \lambda = 0, \\ \frac{21\pi^2+9\lambda^2}{5\pi^2+5\lambda^2}, & \lambda > 0, \end{cases}$$

van Staden and Loots (2009) proposed using the generalized secant hyperbolic distribution to visually illustrate and explain the concept of kurtosis in introductory statistics courses by manipulating the value of β_2 and, by doing so, the value of λ.

The generalized secant hyperbolic distribution includes special cases such as the hyperbolic secant distribution (Baten, 1934; Talacko, 1956) when $\lambda = \frac{-\pi}{2}$ and the logistic distribution when $\lambda = 0$, while it approaches the uniform distribution for $\lambda \to \infty$ (Vaughan, 2002). The normal distribution is not a special case of the generalized secant hyperbolic distribution. But $\beta_2 = 3$ for $\lambda = \pi$ so that the generalized secant hyperbolic distribution is then mesokurtic. We will refer to this special case, to be used in the simulation studies in Chapters 5 and 6, as the generalized secant hyperbolic pi distribution.

Depending on the value of λ the support of the generalized secant hyperbolic distribution can be bounded or infinite. For $\lambda = \pi$ we have infinite support, $(-\infty; \infty)$. I.e., the generalized secant hyperbolic pi distribution does not only have the same levels of skewness and kurtosis as the standard normal distribution, but also shares the same support. Furthermore, as can be seen in Figure 4.2, both these distributions have bell-shaped density curves with the peak of the generalized secant hyperbolic pi distribution slightly lower around the mean, $\alpha = 0$. The Q-Q plot in Figure 4.2 (b) shows that there is also a slight difference between the two distributions' quantiles in the tails.

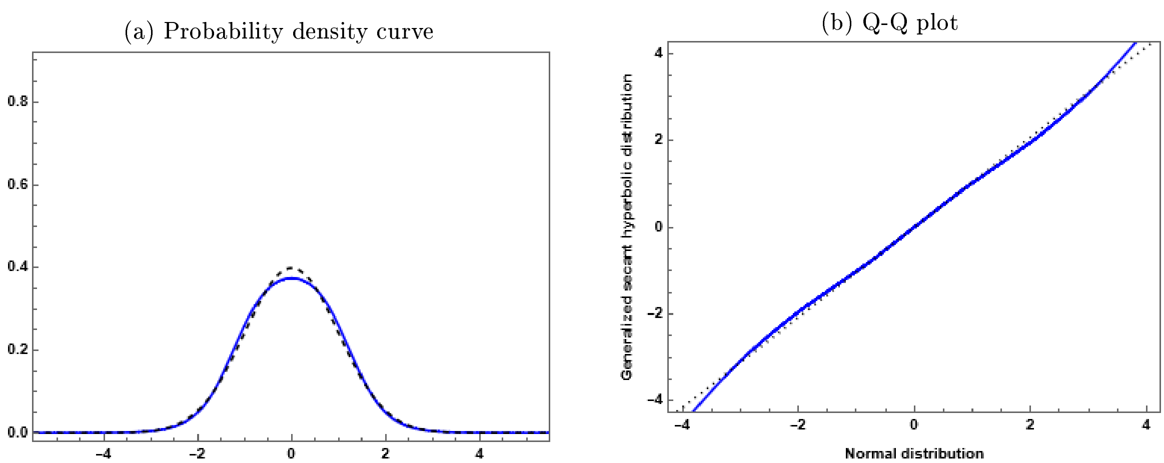


Figure 4.2: Probability density curves of the generalized secant hyperbolic pi distribution (solid line) and the standard normal distribution (dashed line) and Q-Q plot of the quantiles of the generalized secant hyperbolic pi distribution against the standard normal distribution quantiles

4.3 Quantile-based flattened logistic distribution

Gilchrist (2000) introduced a quantile-based model combining the distributional characteristics of the uniform and logistic distributions, resulting in a distribution with a flattened logistic shape for its density curve. This distribution has no closed-form expressions for its PDF or CDF and is defined in terms of its

QF by

$$Q(p) = \alpha + \beta \left[\ln \left(\frac{p}{1-p} \right) + \lambda p \right], \ 0 < p < 1,$$

where α is the location parameter, $\beta > 0$ is the scale parameter and $\lambda \geq 0$ is the shape parameter.

Sharma and Chakrabarty (2019) analysed this distribution in more detail and named it the quantile-based flattened logistic distribution. They derived various functions and properties for this distribution. In particular, if the random variable X has a quantile-based flattened logistic distribution, they showed that the mean and variance of X are respectively

$$E(X) = \alpha + \frac{\beta\lambda}{2}$$

and

$$Var(X) = \beta^2 \left(\lambda + \frac{\lambda^2}{12} + \frac{\pi^2}{3} \right),$$

while Pearson's moment coefficient of skewness is $\sqrt{\beta_1} = 0$ and Pearson's moment coefficient of kurtosis is

$$\beta_2 = \frac{\frac{7}{15}\pi^4 + 2\lambda\pi^2 + \frac{1}{6}\lambda^2\pi^2 + 2\lambda^2 + \frac{\lambda^3}{3} + \frac{\lambda^4}{80}}{\left(\lambda + \frac{\lambda^2}{12} + \frac{\pi^2}{3}\right)^2}.$$

The quantile-based flattened logistic distribution is always symmetric, but depending on the value of its shape parameter, λ, will be platykurtic, mesokurtic or leptokurtic. Specifically Sharma and Chakrabarty (2019) proved that the quantile-based flattened logistic distribution is mesokurtic for $\lambda = 2.892927$, and this special version of the distribution will be used in the simulation studies in Chapters 5 and 6. Note that in order for the distribution to have $E(X) = 0$ and $Var(X) = 1$, we set the location parameter to be $\alpha = -0.55145$ and the scale parameter to be $\beta = 0.38124$.

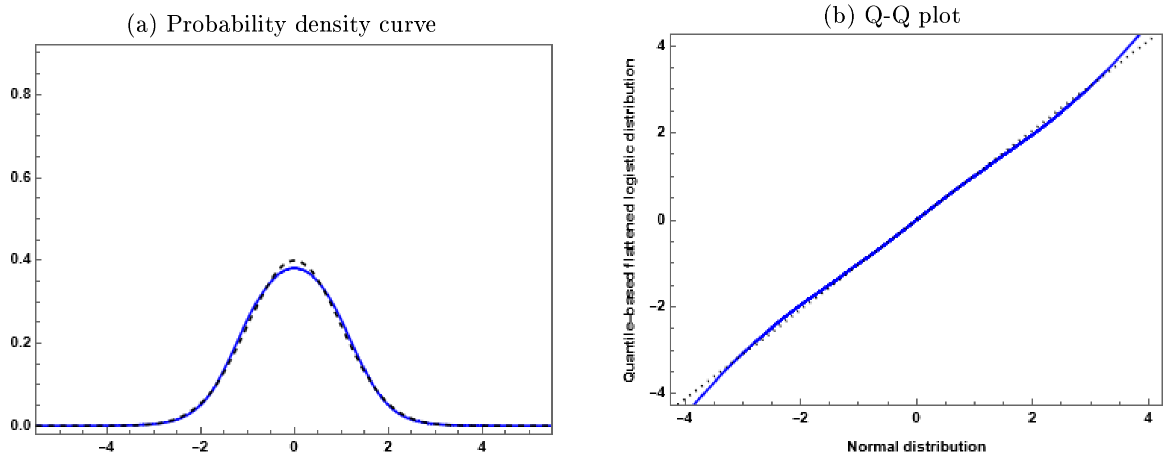


Figure 4.3: Probability density curves of the quantile-based flattened logistic distribution (solid line) and the standard normal distribution (dashed line) and Q-Q plot of the quantiles of the quantile-based flattened logistic distribution against the standard normal distribution quantiles

Figure 4.3 illustrates the density curve as well as the Q-Q plot for the quantile-based flattened logistic distribution with the above mentioned parameter values. As can be seen in Figure 4.3 (a), this distribution has a bell-shaped density curve similar to the standard normal distribution. Both distributions have infinite support, $(-\infty;\infty)$. The Q-Q plot in Figure 4.3 (b) shows a slight deviation between the tails of the two distributions' quantiles.

4.4 Tukey lambda distribution

The Tukey lambda distribution is a symmetric family of distributions named after John W. Tukey who introduced it (Tukey, 1960, 1962) as well as its shape parameter, lambda (λ). This quantile-based distribution is specified in terms of its QF,

$$Q(p) = \begin{cases} \alpha + \frac{\beta}{\lambda}\left[p^{\lambda} - (1-p)^{\lambda}\right], & \lambda \neq 0, \\ \alpha + \beta \log\left(\frac{p}{1-p}\right), & \lambda = 0, \end{cases} \tag{4.1}$$

for $0 < p < 1$. As before, α is the location parameter, $\beta > 0$ is the scale parameter and λ denotes the shape parameter. As with other quantile-based distributions, the PDF and CDF of the Tukey lambda distribution does not exist in closed form. For $\lambda = 0$ in Equation 4.1, the QF of the Tukey lambda distribution reduces to the QF of the logistic distribution.

Joiner and Rosenblatt (1971) gave expressions for the moments of the Tukey lambda distribution (also see Ramberg and Schmeiser (1972)). If a random variable X has a Tukey lambda distribution, then the r^{th} order moment of X, $E(X^r)$, only exists for $\lambda > -\frac{1}{r}$. The mean of X is $E(X) = \alpha$ for $\lambda > -1$ and the variance of X is

$$Var(X) = \frac{2\beta^2}{\lambda^2(2\lambda+1)}\left(1 - \frac{1}{2}\lambda B[\lambda,\lambda]\right),\ \lambda > -\frac{1}{2},$$

where

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt,\ a > 0,\ b > 0, \tag{4.2}$$

is the beta function. Since the Tukey lambda distribution is symmetric, Pearson's moment coefficient of skewness, $\sqrt{\beta_1}$, is zero. The Pearson's moment coefficient of kurtosis is

$$\beta_2 = \frac{2(2\lambda+1)^2(1 - 3\lambda B[3\lambda,\lambda] + 3\lambda B[2\lambda,2\lambda])}{(4\lambda+1)(\lambda B[\lambda,\lambda]-2)^2},\ \lambda > -\frac{1}{4}.$$

The Tukey lambda distribution has the ability to produce symmetric distributions with various distributional shapes depending on the value of λ. These include unimodal bell-shaped distributions with infinite support when $\lambda \leq 0$ and bounded support when $0 < \lambda < 1$, the uniform distribution for

$\lambda = 1$ and for $\lambda = 2$, U-shaped distributions for $1 < \lambda < 2$ and unimodal truncated distributions when $\lambda \geq 2$. This high flexibility of the Tukey lambda distribution in terms of distributional shape makes it a popular choice in simulation studies. For example, in power comparison studies to evaluate the goodness-of-fit tests for normality, the Tukey lambda distribution with different values of λ has been used by Seier (2002), Romão et al. (2010) and Yap and Sim (2011).

In this mini-dissertation two mesokurtic distributions obtained from the Tukey lambda distribution for specific values of λ will be used in the simulation studies in Chapters 5 and 6. Firstly we will consider the Tukey lambda distribution with $\lambda = 0.1349$. For this value of λ, the Tukey lambda distribution gives an excellent approximation of the standard normal distribution. This is evident from the comparison of their density curves as well as their quantiles in Figure 4.4. Note that we set $\alpha = 0$ and $\beta = 0.68327$ so that $E(X) = 0$ and $Var(X) = 1$. Because $\lambda = 0.1349 > 0$, the distribution has finite support, $[-5.06454; 5.06464]$.

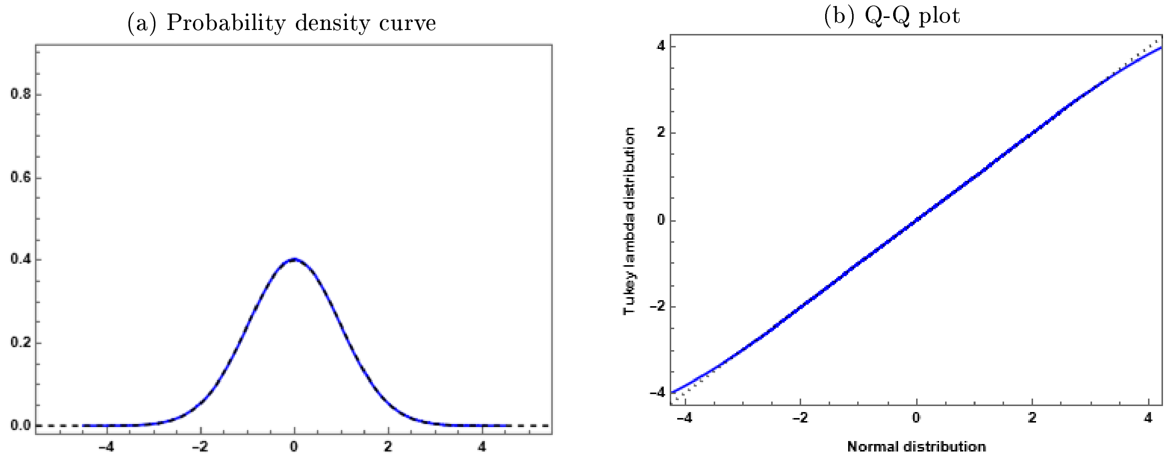


Figure 4.4: Probability density curves of the Tukey lambda distribution with $\lambda = 0.13491$ (solid line) and the standard normal distribution (dashed line) and Q-Q plot of the quantiles of the Tukey lambda distribution with $\lambda = 0.13491$ against the standard normal distribution quantiles

The second mesokurtic distribution obtained from the Tukey lambda distribution is for $\lambda = 5.2029$. For this distribution we let $\alpha = 0$ and $\beta = 12.4439$ so that we again have $E(X) = 0$ and $Var(X) = 1$. As indicated before, the Tukey lambda distribution gives unimodal truncated distributions when $\lambda \geq 2$. Hence, for $\lambda = 5.2029$ the Tukey lambda distribution is truncated with bounded support, $[-2.39173; 2.39173]$, as illustrated in Figure 4.5 (a). So even though this Tukey lambda distribution is mesokurtic, its density curve deviates substantially from the standard normal distribution's density curve with a much narrower and higher peak at the mean value. Furthermore, Figure 4.5 (b) shows that the two distributions' quantiles are also vastly different.

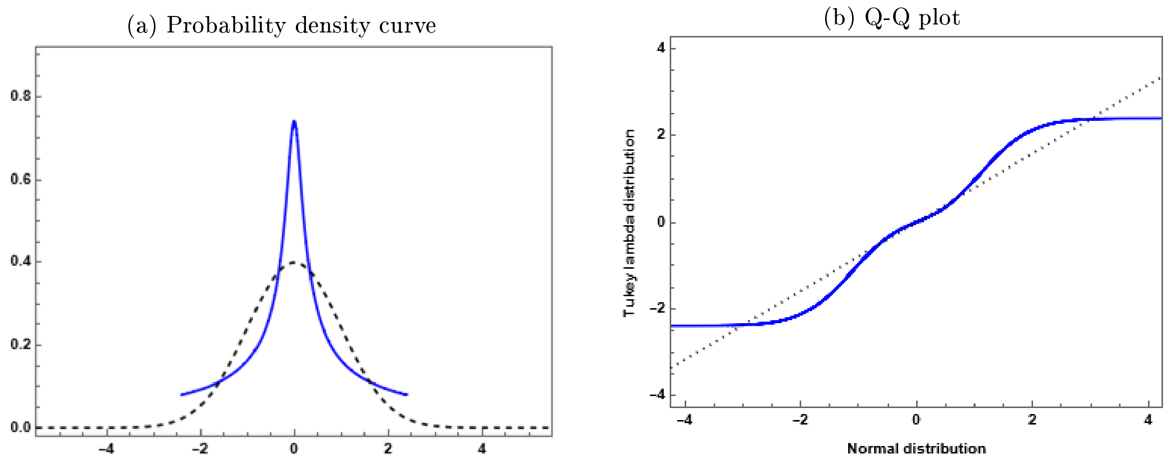


Figure 4.5: Probability density curves of the Tukey lambda distribution with $\lambda = 5.2029$ (solid line) and the standard normal distribution (dashed line) and Q-Q plot of the quantiles of the Tukey lambda distribution with $\lambda = 5.2029$ against the standard normal distribution quantiles

4.5 Generalized Tukey lambda distribution

Various asymmetric generalizations of the Tukey lambda distribution have been proposed in the literature. These include models by Ramberg and Schmeiser (1974), Freimer et al. (1988) and more recently van Staden et al. (2013). The generalized Tukey lambda distribution of Ramberg and Schmeiser (1972, 1974) was initially proposed by them to generate symmetric and asymmetric random variables. Ramberg et al. (1979) extended the use of this generalized Tukey lambda distribution to be used for fitting distributions to data. A detailed discussion of the generalized Tukey lambda distribution is beyond the scope of this mini-dissertation. See King (1999); Karian and Dudewicz (2010) and van Staden et al. (2013) for in-depth investigations into the characteristics of this model, specifically with respect to parameter space, distributional shape, and application to diverse fields of research.

In this mini-dissertation a specific mesokurtic member from the generalized Tukey lambda distribution of Ramberg et al. (1979) will be considered in the simulation study in Chapters 5 and 6. This quantile-based distribution has QF,

$$Q(p) = \alpha + \beta \left[p^{\lambda} - (1-p)^{\delta}\right], \; 0 < p < 1,$$

where α is the location parameter, $\beta > 0$ is the scale parameter and λ and δ are both shape parameters.

As first shown by Ramberg and Schmeiser (1974), if X is a random variable that has a generalized Tukey lambda distribution, then the r^{th} order moment of X only exists if $\lambda > -\frac{1}{r}$ and $\delta > -\frac{1}{r}$. If $\lambda > -\frac{1}{4}$ and $\delta > -\frac{1}{4}$, then the mean, the variance and the Pearson's moment coefficients of skewness and kurtosis of X are, as presented by van Staden et al. (2013),

$$E(X) = \alpha + \beta M_1,$$

$$Var(X) = \beta^2 \left(M_2 - M_1^2 \right),$$

$$\sqrt{\beta_1} = \frac{M_3 - 3M_1M_2 + 2M_1^2}{(M_2 - M_1^2)^{\frac{3}{2}}}$$

and

$$\beta_2 = \frac{M_4 - 4M_1M_3 + 6M_1^2M_2 - 3M_1^4}{\left(M_2 - M_1^2\right)^2},$$

where

$$M_1 = \frac{1}{\lambda+1} - \frac{1}{\delta+1},$$

$$M_2 = \frac{1}{2\lambda+1} - 2B[\lambda+1, \delta+1] + \frac{1}{2\delta+1},$$

$$M_3 = \frac{1}{3\lambda+1} - 3B[2\lambda+1, \delta+1] + 3B[\lambda+1, 2\delta+1] - \frac{1}{3\delta+1}$$

and

$$M_4 = \frac{1}{4\lambda+1} - 4B[3\lambda+1, \delta+1] + 6B[2\lambda+1, 2\delta+1] - 4B[\lambda+1, 3\delta+1] + \frac{1}{4\delta+1},$$

with $B(a, b)$ the beta function, given in Equation 4.2.

If $\lambda = \delta$, then $M_1 = M_3 = 0$ so that $\sqrt{\beta_1} = 0$. That is, for $\lambda = \delta$ the generalized Tukey lambda distribution reduces to the symmetric Tukey lambda distribution. When $\lambda \neq \delta$, the generalized Tukey lambda distribution is asymmetric. However, as indicated by Ramberg et al. (1980), there exist values for $\lambda \neq \delta$ for which $\sqrt{\beta_1} = 0$ even though the distribution is asymmetric.

We will consider a generalized Tukey lambda distribution with location parameter $\alpha = 1.093$, scale parameter $\beta = 3.0654$ and shape parameters $\lambda = 74.1984$ and $\delta = 1.70373$ for which $E(X) = 0$, $Var(X) = 1$, $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$. Figure 4.6 (a) shows that this member from the generalized Tukey lambda distribution is an asymmetric distribution with bounded support, $[-1.97240; 4.15841]$. This mesokurtic distribution is substantially different in distributional shape to that of the standard normal distribution. The Q-Q plot in Figure 4.6 (b) also indicates the difference between the two distributions' quantiles.

4.6 Schmeiser-Deutsch distribution

Schmeiser and Deutsch (1977) presented a flexible distributional family, known as the Schmeiser-Deutsch distribution, from which random variables can easily be generated through the distribution's quantile function. If a random variable X has a Schmeiser-Deutsch distribution, then the PDF, CDF and QF of X are given respectively by,

$$f(x) = \frac{\left| \frac{x-\alpha}{\beta} \right|^{-1+\frac{1}{\lambda}}}{\beta\lambda}, \ \alpha - \beta\delta^{\lambda} \leq x \leq \alpha + \beta\delta^{\lambda},$$

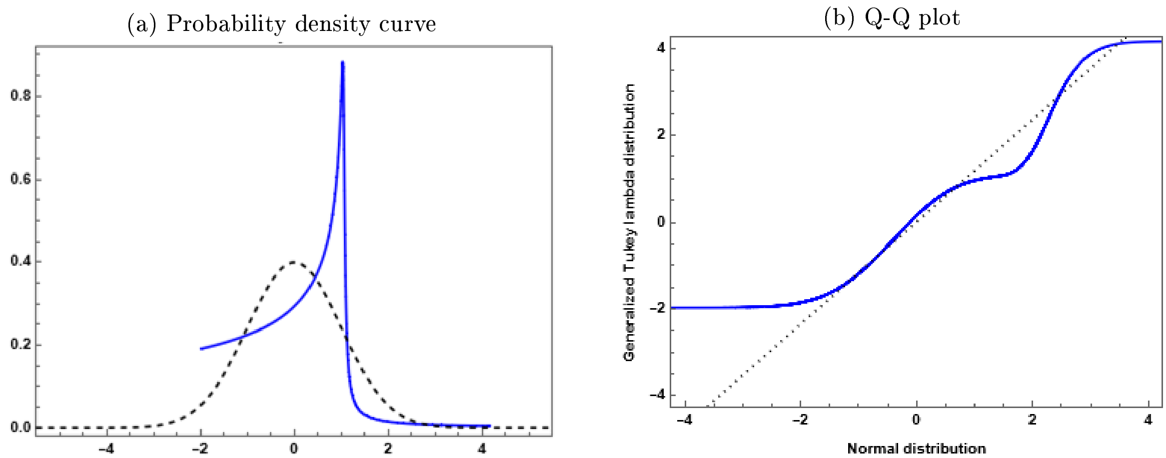


Figure 4.6: Probability density curves of the generalized Tukey lambda distribution (solid line) and the standard normal distribution (dashed line) and Q-Q plot of the quantiles of the generalized Tukey lambda distribution against the standard normal distribution quantiles

$$F(x) = \begin{cases} \delta - \left(\frac{\alpha - x}{\beta}\right)^{\frac{1}{\lambda}} & , \ \alpha - \beta\delta^{\lambda} \le x \le \alpha, \\ \delta + \left(\frac{x-\alpha}{\beta}\right)^{\frac{1}{\lambda}} & , \ \alpha \le x \le \alpha + \beta\delta^{\lambda}, \end{cases}$$

$$Q(p) = \begin{cases} \alpha - \beta(-p+\delta)^{\lambda}, & p \le \delta, \\ \alpha + \beta(p-\delta)^{\lambda}, & p > \delta, \end{cases}$$

where α is the location parameter, $\beta > 0$ is the scale parameter and λ and $0 \le \delta \le 1$ are both shape parameters.

Schmeiser and Deutsch (1977) showed that if X is a random variable that has a standard Schmeiser-Deutsch distribution with $\alpha = 0$ and $\beta = 1$, then the r^{th} order moment of X exists for $\lambda > -\frac{1}{r}$ and is given by

$$E(X^r) = \frac{(1-\delta)^{1+r\lambda}}{1+r\lambda} + \frac{(-1)^r\delta^{1+r\lambda}}{1+r\lambda}.$$

The four-parameter Schmeiser-Deutsch distribution included in this study has location parameter $\alpha = 0$, scale parameter $\beta = 10.9117$ and shape parameters $\lambda = 2.22474$ and $\delta = 0.5$ so that the distribution is mesokurtic with $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$. Schmeiser and Deutsch (1977) gave a brief comparison of this distribution and the standard normal distribution in terms of their CDFs. Figure 4.7 (a) compares their PDFs. The density curve of the Schmeiser-Deutsch distribution is infinite at its mean level of zero. Also, the distribution has truncated tails and bounded support, $[-2.33441; 2.33441]$. Consequently, the distributional shape of the Schmeiser-Deutsch distribution is completely different from the standard normal distribution's shape. The quantiles of the Schmeiser-Deutsch distribution and the standard normal distribution are also very different.

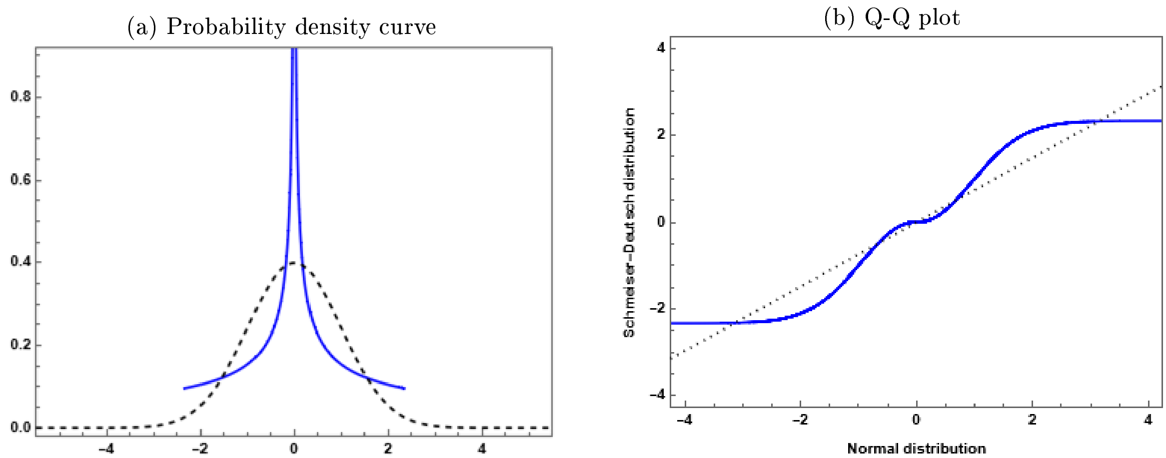


Figure 4.7: Probability density curves of the Schmeiser-Deutsch distribution (solid line) and the standard normal distribution (dashed line) and Q-Q plot of the quantiles of the Schmeiser-Deutsch distribution against the standard normal distribution quantiles

4.7 Two-tailed gamma distribution

In discussing whether kurtosis is a measure of bimodality, Hildebrand (1971) considered the two-tailed gamma distribution. If a random variable X has this distribution, then the PDF of X is given by

$$f(x) = \frac{\lambda^{\beta}}{2\Gamma(\beta)} \, |x|^{\beta-1} \, e^{-\lambda|x|}, \quad -\infty < x < \infty,$$

where $\beta > 0$ and $\lambda > 0$ are respectively scale and shape parameters. Note that a location parameter, α, can be included in the distribution. However, our focus will be on the two-tailed gamma distribution with $\alpha = 0$ so that $E(X) = 0$. The two-tailed gamma distribution has infinite support. When $\lambda > 1$, the two-tailed gamma distribution is bimodal.

The r^{th} order moment of X for the two-tailed gamma distribution is

$$E(X^r) = \frac{((1 + (-1)^r)\Gamma[r + \lambda]}{2\beta\Gamma[\lambda]},$$

where

$$\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt, \; a > 0, \tag{4.3}$$

is the gamma function. As indicated above, $E(X) = 0$. The variance of X is

$$Var(X) = \frac{\lambda(1 + \lambda)}{\beta^2}. \tag{4.4}$$

Because the two-tailed gamma distribution is symmetric around $E(X) = 0$, Pearson's moment coefficient of skewness is zero, $\sqrt{\beta_1} = 0$. Pearson's moment coefficient of kurtosis is given by

$$\beta_2 = \frac{6 + 5\lambda + \lambda^2}{\lambda + \lambda^2}. \tag{4.5}$$

As indicated by Hildebrand (1971), $\beta_2 = 3$ when

$$\lambda = \frac{1}{2}(1 + \sqrt{13}) = 2.30278.$$

This result also follows by setting Equation 4.5 equal to 3 and solving for λ. Furthermore, if we set Equation 4.4 equal to 1, i.e., $Var(X) = 1$ and solve for β, we get $\beta = 2.75782$. These values of β and λ are used in the simulation study in Chapters 5 and 6.

Figure 4.8 shows that the two-tailed gamma distribution is similar to the standard normal distribtuion at the tails but differs substantially at the mean since the two-tailed gamma distribution is bimodal. This can also be seen in the Q-Q plot.

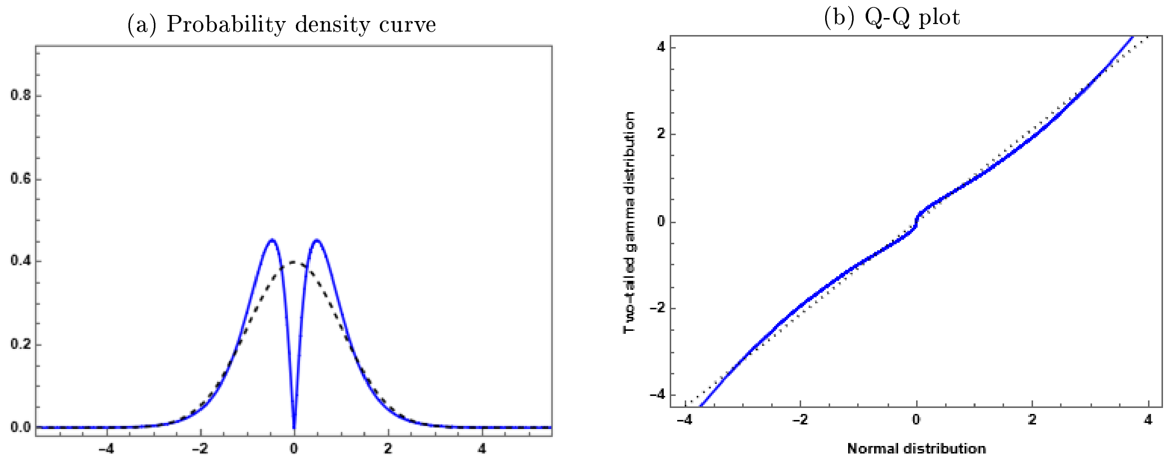


Figure 4.8: Probability density curves of the two-tailed gamma distribution (solid line) and the standard normal distribution (dashed line) and Q-Q plot of the quantiles of the two-tailed gamma distribution against the standard normal distribution quantiles

4.8 Burr type III and Burr type XII distributions

The Burr family of distributions is a system of 12 continuous distributions proposed by Burr (1942) in terms of their CDFs. Each member from this system is labelled as a numbered type. Among these different types, the Burr type XII distribution, often simply referred to as *the* Burr distribution, and the Burr type III distribution, often called the reciprocal Burr distribution, have become the most popular types in the literature. The reason is that these two distributions cover a wide range of values of skewness and kurtosis (Rodriguez, 1977; Tadikamalla, 1980) and are therefore useful for modelling data from different areas of applications.

First consider the Burr type XII distribution. For a random variable X that has this distribution, the PDF, CDF and QF of X is given by,

$$f(x) = \frac{\lambda\delta}{\beta}\left(\frac{x-\alpha}{\beta}\right)^{\delta-1}\left(1+\left(\frac{x-\alpha}{\beta}\right)^{\delta}\right)^{-\lambda-1},\ \alpha < x < \infty,$$

$$F(x) = 1 - \left(1+\left(\frac{x-\alpha}{\beta}\right)^{\delta}\right)^{-\lambda},\ \alpha < x < \infty,$$

$$Q(p) = \alpha + \beta\left((1-p)^{-\frac{1}{\lambda}} - 1\right)^{\frac{1}{\delta}},\ 0 < p < 1,$$

where α is the location parameter, $\beta > 0$ is the scale parameter and $\lambda > 0$ and $\delta > 0$ are both shape parameters.

Rodriguez (1977) gave expressions for the moments of the Burr type XII distribution (also see Tadikamalla (1980)). If the variable X has a standard Burr type XII distribution with $\alpha = 0$ and $\beta = 1$, then the r^{th} order moment of X only exists for $\lambda\delta > r$ and is given by

$$E(X^r) = \frac{L_r}{\Gamma[\lambda]}$$

where $L_r = \Gamma[\frac{r+\delta}{\delta}]\Gamma[\lambda - \frac{r}{\delta}]$ for $r = 1, 2, 3...$ and with $\Gamma(a)$ the gamma function given in Equation 4.3. Rodriguez (1977) then showed that Pearson's moment coefficients of skewness and kurtosis are respectively given by

$$\sqrt{\beta_1} = \frac{\Gamma^2[\lambda]L_3 - 3\Gamma[\lambda]L_2L_1 + 2L_1^3}{(\Gamma[\lambda]L_2 - L_1^2)^{\frac{3}{2}}}$$

and

$$\beta_2 = \frac{\Gamma^3[\lambda]L_4 - 4\Gamma^2[\lambda]L_3L_1 + 6\Gamma[\lambda]L_2L_1^2 - 3L_1^4}{(\Gamma[\lambda]L_2 - L_1^2)^2}.$$

In general for a variable X that has a Burr type XII distribution with parameters α and $\beta > 0$, the mean and variance are

$$E(X) = \alpha + \beta\frac{L_1}{\Gamma[\lambda]},$$

$$Var(X) = \frac{\beta^2}{\Gamma^2[\lambda]}\left[\Gamma[\lambda]L_2 - L_1^2\right].$$

Rodriguez (1977) and Tadikamalla (1980) both graphically illustrated the coverage of the $(\sqrt{\beta_1}, \beta_2)$ space by the Burr type XII distribution. Even though this distribution is asymmetric with half-infinite support, $[\alpha, \infty)$, it is possible to obtain $\sqrt{\beta_1} = 0$ for certain shape parameter values. This is for example the case when $\lambda = 6.15784$ and $\delta = 4.8737$. Furthermore, for these shape parameter values we have $\beta_2 = 3$ so that the Burr type XII distribution is mesokurtic. The Burr type XII distribution with these shape parameter values will be used in the simulation study in Chapters 5 and 6. Note that we set $\alpha = -3.97999$ and $\beta = 6.17331$ to get $E(X) = 0$ and $Var(X) = 1$. In Figure 4.9 (a) the density curve of this Burr type XII distribution with half-infinite support $[-3.97999; \infty)$ is compared to the standard normal distribution's density curve. Figure 4.9 (b) compares these distributions' quantiles. Despite the difference in their support, the Burr type XII distribution provides a good approximation of the standard normal distribtuion.

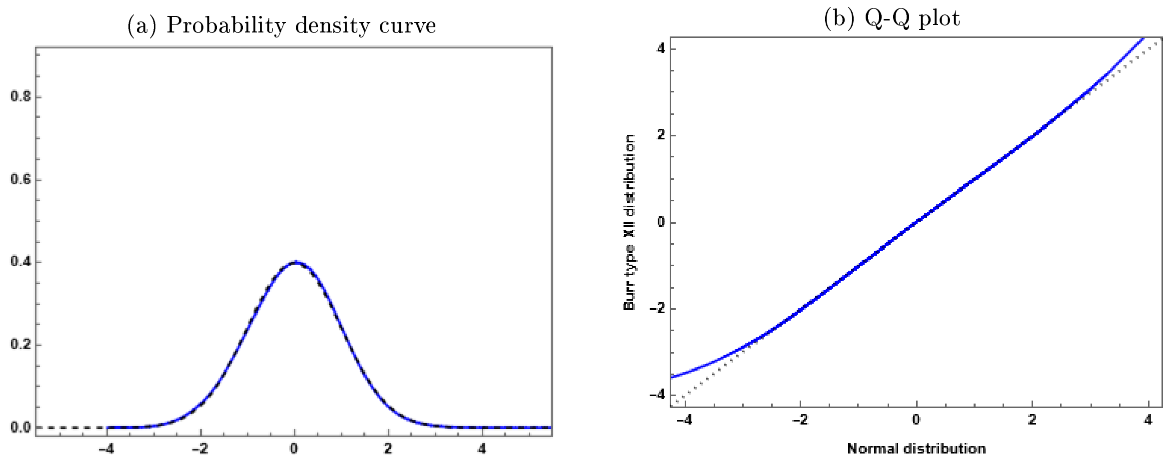


Figure 4.9: Probability density curves of the Burr type XII distribution (solid line) and the standard normal distribution (dashed line) and Q-Q plot of the quantiles of the Burr type XII distribution against the standard normal distribution quantiles

Consider next the Burr type III distribution. The PDF, CDF and QF of a random variable X that has this distribution are given by,

$$f(x) = \frac{\lambda\delta}{\beta}\left(\frac{x-\alpha}{\beta}\right)^{-\delta-1}\left(1+\left(\frac{x-\alpha}{\beta}\right)^{-\delta}\right)^{-\lambda-1}, \ \alpha < x < \infty,$$

$$F(x) = \left(1+\left(\frac{x-\alpha}{\beta}\right)^{-\delta}\right)^{-\lambda}, \ \alpha < x < \infty,$$

$$Q(p) = \alpha + \beta\left(\frac{p^{\frac{1}{\lambda}}}{1-p^{\frac{1}{\lambda}}}\right)^{\frac{1}{\delta}}, \ 0 < p < 1,$$

where α is the location parameter, $\beta > 0$ is the scale parameter and $\lambda > 0$ and $\delta > 0$ are both shape parameters.

Tadikamalla (1980) studied the moments of the Burr type III distribution and specifically the coverage of the $(\sqrt{\beta_1}, \beta_2)$ space by this distribution. For $\delta > r$ the r^{th} order moment exists and is given by,

$$E(X^r) = \frac{L_r}{\Gamma[\lambda]}.$$

where $L_r = \Gamma[\frac{\delta-r}{\delta}]\Gamma[\frac{r}{\delta}+\lambda]$ for $r = 1, 2, 3...$ and with $\Gamma(a)$ the gamma function given in Equation 4.3. Similarly to the Burr type XII distribution, the Pearson's moment coefficients of skewness and kurtosis can be shown to be respectively

$$\sqrt{\beta_1} = \frac{\Gamma^2[\lambda]L_3 - 3\Gamma[\lambda]L_2L_1 + 2L_1^3}{(\Gamma[\lambda]L_2 - L_1^2)^{\frac{3}{2}}}$$

and

$$\beta_2 = \frac{\Gamma^3[\lambda]L_4 - 4\Gamma^2[\lambda]L_3L_1 + 6\Gamma[\lambda]L_2L_1^2 - 3L_1^4}{(\Gamma[\lambda]L_2 - L_1^2)^2},$$

with the mean and variance given by respectively,

$$E(X) = \alpha + \beta\frac{L_1}{\Gamma[\lambda]}$$

and

$$Var(X) = \frac{\beta^2}{\Gamma^2[\lambda]}\left[\Gamma[\lambda]L_2 - L_1^2\right].$$

Like the Burr type XII distribution, the Burr type III distribution is asymmetric with half-infinite support, $[\alpha, \infty)$. But again, it is possible for $\sqrt{\beta_1}$ to be zero for certain values of the shape parameters. In particular, when we set $\lambda = 0.200681$ and $\delta = 10.5184$, then the resulting Burr type III distribution has $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$. We use these shape parameter values together with $\alpha = -2.61292$ and $\beta = 3.69681$, so that $E(X) = 0$ and $Var(X) = 1$, in the simulation study in Chapters 5 and 6.

Comparing Figure 4.10 with Figure 4.9, it is noticable that the density curve as well as the quantiles of the standard normal distribution deviates more from the selected Burr type III distribution than from the selected Burr type XII distribution. This is especially true in the lower tail.

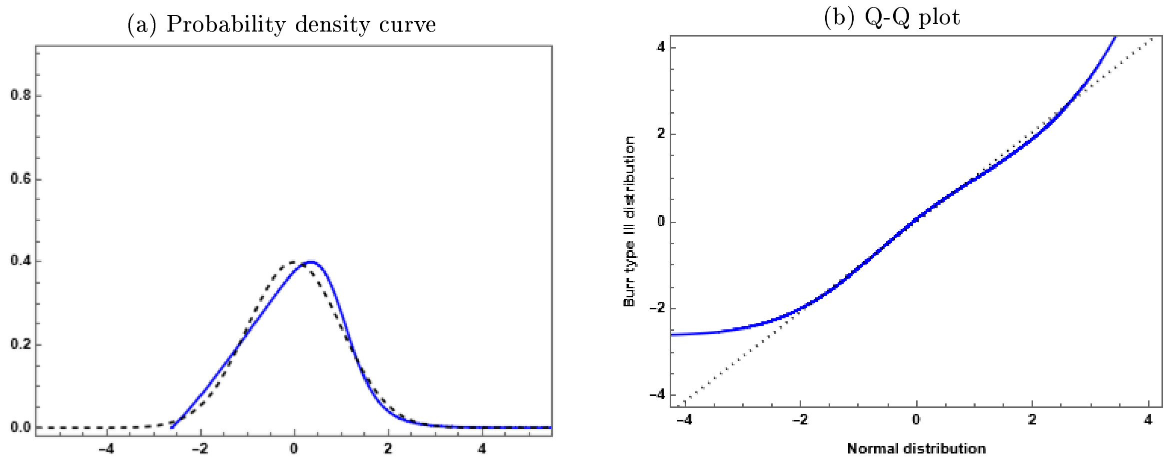


Figure 4.10: Probability density curves of the Burr type III distribution (solid line) and the standard normal distribution (dashed line) and Q-Q plot of the quantiles of the Burr type III distribution against the standard normal distribution quantiles

4.9 Davies distribution

The Davies distribution is a quantile-based distribution that was studied in detail by Hankin and Lee (2006). It first appeared under the name power-Pareto distribution in Gilchrist (2000). Like other quantile-based distributions, the Davies distribution has no closed form expression for either its PDF or its CDF. This distribution has QF given by,

$$Q(p) = \alpha + \beta \frac{p^{\lambda}}{(1-p)^{\delta}}, \ 0 < p < 1,$$

where α and $\beta > 0$ are the location and scale parameters respectively and where $\lambda \geq 0$ and $\delta \geq 0$ are shape parameters such that $\lambda + \delta > 0$. The r^{th} order moment of a random variable X that has a Davies distribution exists for $\delta < \frac{1}{r}$ and is given by

$$E(X^r) = B[1 - r\delta, 1 + r\lambda],$$

where $B[a, b]$ is the beta function given in Equation 4.2. Special cases of the Davies distribution appear in the literature, namely in Burr (1942) and in Johnson and Tadikamalla (1992). When $\lambda = \delta$ the scaled logistic distribution is obtained and if $\delta = 0$ the Burr distribution is obtained. Like the Burr type III and Burr type XII distributions the mean, variance and Pearson's moment coefficients of skewness and kurtosis can be shown to be

$$E(X) = \alpha + \beta B[1 - \delta, 1 + \lambda],$$

$$Var(X) = \beta^2 \left(B[1 - 2\delta, 1 + 2\lambda] - (B[1 - \delta, 1 + \lambda])^2 \right),$$

$$\sqrt{\beta_1} = \frac{B[1 - 3\delta, 1 + 3\lambda] - 3B[1 - 2\delta, 1 + 2\lambda]B[1 - \delta, 1 + \lambda] + 2(B[1 - \delta, 1 + \lambda])^2}{(B[1 - 2\delta, 1 + 2\lambda] - (B[1 - \delta, 1 + \lambda])^2)^{\frac{3}{2}}}$$

and

$$\beta_2 = \frac{B[1 - 4\delta, 1 + 4\lambda] - 4B[1 - 3\delta, 1 + 3\lambda]B[1 - \delta, 1 + \lambda]}{(B[1 - 2\delta, 1 + 2\lambda] - (B[1 - \delta, 1 + \lambda])^2)^2}$$
$$+ \frac{6B[1 - 2\delta, 1 + 2\lambda](B[1 - \delta, 1 + \lambda])^2 - 3(B[1 - \delta, 1 + \lambda])^4}{(B[1 - 2\delta, 1 + 2\lambda] - (B[1 - \delta, 1 + \lambda])^2)^2},$$

where $B[a, b]$ is the Beta function in Equation 4.2.

The Davies distribution is an asymmetric distribution with half infinite support, $[\alpha, \infty)$. However, for selected shape parameter values it is possible to get a Pearson's moment coefficient of skewness value of zero. An example is for shape parameters $\lambda = 0.340471$ and $\delta = 0.0819112$ for which we get $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$ so that the Davies distribution is indeed mesokurtic. Also, if we set $\alpha = -3.02177$ and $\beta = 3.65724$, then $E(X) = 0$ and $Var(X) = 1$. Figure 4.11 (a) shows that the Davies distributions density curve is similar to the standard normal distribution density curve. However, these distributions differ in their tails especially since this Davies distribution has half-infinite support, $[-3.02177; \infty)$.

(a) Probability density curve

(b) Q-Q plot

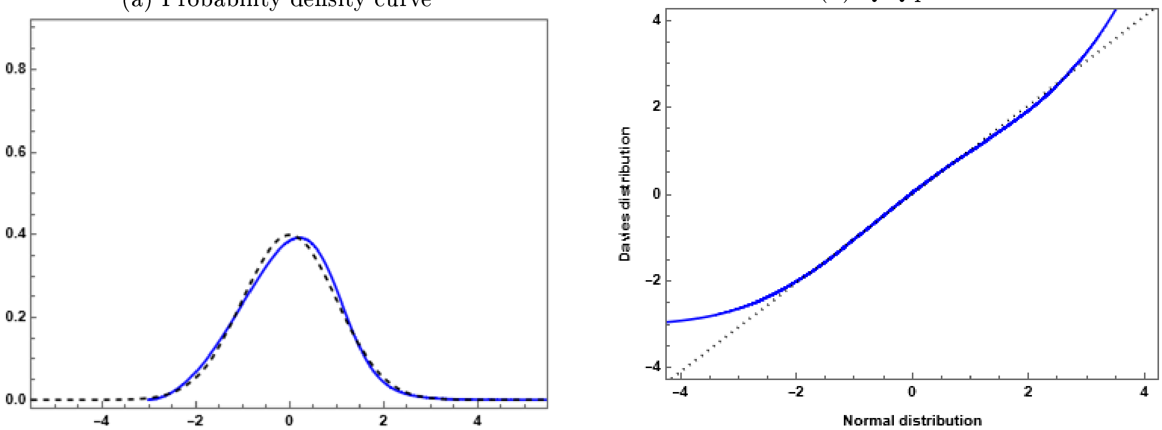


Figure 4.11: Probability density curves of the Davies distribution (solid line) and the standard normal distribution (dashed line) and Q-Q plot of the quantiles of the Davies distribution against the standard normal distribution quantiles

4.10 Summary of mesokurtic distributions

There are 11 mesokurtic distributions included in this mini-dissertation. To help compare the distributions' characteristics, a summary of their parameter values, see Table 4.1, and distributional properties, see Table 4.2, has been included. These distributions all have a mean value of 0, variance equal to 1, Pearson's moment coefficients of skewness and kurtosis values of 0 and 3, respectively, which is the same as for the standard normal distribution.

Table 4.1: Distributions and their location parameters (α), scale parameters (β) and shape parameters (λ, δ) included in the simulation study

Distribution	α	β	λ	δ	**References**
Normal	0	1			
Generalized secant hyperbolic pi	0	1	3.1416		Vaughan (2002)
Quantile-based flattened logistic	-0.5515	0.3812	2.8929		Sharma and Chakrabarty (2019)
Tukey lambda ($\lambda = 0.1349$)	0	0.6833	0.1349		Tukey (1960, 1962)
Tukey lambda ($\lambda = 5.2029$)	0	12.4439	5.2029		Tukey (1960, 1962)
Generalized Tukey lambda	1.093	3.0654	74.1984	1.7037	Ramberg et al. (1979)
Schmeiser-Deutsch	0	10.9117	2.2247	0.5	Schmeiser and Deutsch (1977)
Two-tailed gamma	0	2.7578	2.3028		Hildebrand (1971)
Burr type III	-2.6129	3.6968	0.2007	10.5184	Burr (1942)
Burr type XII	-3.9800	6.1733	6.1578	4.8737	Burr (1942)
Davies	-3.0218	3.6572	0.3405	0.0819	Hankin and Lee (2006)

Table 4.2: Comparison of the distributional properties of the standard normal distribution with the properties of the distributions used in the simulation study

Distribution	**Support**	**Unimodal**	**Symmetric**	**Truncated**	**Quantile-based**
Normal	Infinite	Yes	Yes	No	No
Generalized secant hyperbolic pi	Infinite	Yes	Yes	No	No
Quantile-based flattened logistic	Infinite	Yes	Yes	No	Yes
Tukey lambda ($\lambda = 0.1349$)	Bounded	Yes	Yes	No	Yes
Tukey lambda ($\lambda = 5.2029$)	Bounded	Yes	Yes	Yes	Yes
Generalized Tukey lambda	Bounded	Yes	No	Yes	Yes
Schmeiser-Deutsch	Bounded	Yes	Yes	Yes	No
Two-tailed gamma	Infinite	No	Yes	No	No
Burr type III	Half-infinite	Yes	No	No	No
Burr type XII	Half-infinite	Yes	No	No	No
Davies	Half-infinite	Yes	No	No	Yes

We can see in Table 4.2 that even though all these distributions are mesokurtic, some distributions have distributional shapes and properties that are substantially different to the standard normal distribution. Based on their characteristics, the mesokurtic distributions can be grouped as follow:

Group 1:

- Generalized secant hyperbolic pi distribution
- Quantile-based flattened logistic distribution
- Tukey lambda distribution with $\lambda = 0.13491$

Group 2:

- Tukey lambda distribution with $\lambda = 5.2029$
- Generalized Tukey lambda distribution
- Schmeiser-Deutsch distribution

Group 3:

- Two-tailed gamma distribution

Group 4:

- Burr type III distribution
- Burr type XII distribution
- Davies distribution

Group 1 consists of unimodal, symmetric distributions on infinite or bounded support. These distributions closely approximate the normal distribution. The distributions in Group 2 are unimodal, truncated distributions on bounded support. The two-tailed gamma distribution from Group 3 is the only bimodal distribution considered in the simulation study. Finally, Group 4 includes distributions that are unimodal, asymmetric distributions with half-infinite support.

Chapter 5

Simulation study and comparison of goodness-of-fit tests

The simulation study was conducted using Wolfram Mathematica, Version 12.3 (Wolfram Research, Champaign, IL, USA). The objective of the simulation study is to evaluate the power of the normality tests, discussed in Chapter 3, for the different mesokurtic distributions, discussed in Chapter 4. The number of simulations used are 50000 and the comparison study is carried out for sample sizes $n = 20$, $n = 50$ and $n = 100$. Furthermore, significance levels 1%, 5% and 10% are considered for each sample size. Take note that the power of a test is the proportion of times the null hypothesis is correctly rejected for 50000 repetitions. The simulation study includes the distribution of the p-values and a comparison of the goodness-of-fit tests for the various distributions.

5.1 Algorithm for obtaining parameter values for mesokurtic distributions

In order to obtain parameter values so that a distribution is indeed mesokurtic, Algorithm 1 is used. This algorithm is used for each distribution in Chapter 4 to obtain their parameter values so that the distribution has Pearson's moment coefficients of skewness and kurtosis, equal to 0 and 3, respectively.

Algorithm 1 Parameter values algorithm

1. Find the general formulas for the mean, variance and Pearson's moment coefficients of skewness and kurtosis for the distribution.
2. Set Pearson's moment coefficients of skewness and kurtosis equal to 0 and 3, respectively.
3. Solve simultaneously for the unknown shape parameter values of the distribution using the formulas for Pearson's moment coefficients of skewness and kurtosis from step 1.
4. Set the variance equal to 1.
5. Solve for the unknown scale parameter value using the formula for the variance from step 1 as well as the calculated shape parameter values from step 3.
6. Set the mean equal to 0.
7. Finally solve for the unknown location parameter value using the formula for the mean from step 1 together with the calculated shape and scale parameter values in steps 3 and 5.

To illustrate how Algorithm 1 is used, we show how to find the parameter values for the Davies distribution.

1. The general formulas for the mean, variance and Pearson's moment coefficients of skewness and kurtosis are in Equations 4.9 to 4.9.
2. Set $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$.
3. Solving simultaneously we get $\lambda = 0.340471$ and $\delta = 0.0819112$.
4. Set $Var(X) = 1$.
5. Solving for the unknown scale parameter value gives $\beta = 3.65724$.
6. Set $E(X) = 0$.
7. Solving for the location parameter value we obtain $\alpha = -3.02177$.

See the Appendix (Wolfram Mathematica code) for details to access the code implementing Algorithm 1 for the Davies distribution.

Table 5.1: Power comparison of normality tests (1%)

Distributions	**n**	**KS**	**CvM**	**AD**	**V**	**U^2**	**BHEP**	**JB**	**MS**	**MK**	**P**	**SW**
Normal	20	0.9	0.9	1.0	1.0	0.9	0.6	1.0	0.6	0.6	1.2	1.0
	50	1.0	0.9	0.9	1.0	1.0	0.7	1.0	1.0	1.2	1.0	1.0
	100	1.1	1.0	1.0	1.1	1.0	0.8	1.0	1.1	1.5	1.1	1.0
Generalized secant hyperbolic pi	20	0.9	1.0	1.0	1.1	1.0	0.6	1.2	0.8	0.7	1.2	1.1
	50	1.2	1.1	1.2	1.3	1.2	0.8	1.9	1.7	2.1	1.2	1.7
	100	1.4	1.4	1.5	1.7	1.5	1.1	2.5	2.0	3.0	1.4	2.2
Quantile-based flattened logistic	20	0.9	1.0	1.0	1.1	1.0	0.6	1.2	0.8	0.7	1.3	1.1
	50	1.1	1.1	1.1	1.2	1.1	0.7	1.8	1.6	2.0	1.1	1.6
	100	1.3	1.2	1.3	1.4	1.3	1.0	2.4	1.9	2.8	1.3	2.0
Tukey lambda ($\lambda = 0.13491$)	20	0.9	0.9	0.9	1.0	1.0	0.6	1.0	0.6	0.5	1.2	0.9
	50	1.0	1.0	0.9	1.0	1.0	0.7	0.8	0.8	1.0	1.0	0.9
	100	1.1	1.0	1.1	1.1	1.0	0.8	0.7	0.8	1.0	1.1	0.9
Tukey lambda ($\lambda = 5.2029$)	20	3.8	4.6	4.0	5.2	5.2	2.3	0.9	0.5	0.6	4.4	2.0
	50	11.3	14.7	12.9	17.6	17.3	8.6	0.1	0.2	0.3	12.3	4.0
	100	29.9	40.3	37.6	44.8	44.9	33.0	0.0	0.1	0.1	33.9	16.4
Generalized Tukey lambda	20	6.8	11.0	12.6	10.5	11.9	5.7	2.0	1.1	1.2	10.8	10.3
	50	24.2	39.1	47.6	38.6	41.5	36.6	3.3	2.3	4.2	38.2	46.5
	100	*58.0*	**75.5**	**87.1**	**78.4**	**78.5**	**80.1**	4.0	3.6	6.9	**76.1**	**90.2**
Schmeiser-Deutsch	20	12.2	13.6	10.7	16.8	15.9	4.1	1.1	0.6	0.8	17.2	4.6
	50	39.6	46.1	38.8	*53.2*	*50.7*	18.3	0.2	0.2	0.3	*57.5*	14.5
	100	**79.6**	**87.0**	**83.1**	**90.5**	**89.7**	*62.4*	0.0	0.1	0.2	**93.9**	*56.5*
Two-tailed gamma	20	3.0	2.7	2.4	3.4	3.1	1.0	1.3	1.0	0.8	3.6	1.9
	50	8.9	8.0	6.2	10.8	9.4	3.1	2.3	2.2	2.4	8.1	3.9
	100	23.2	23.2	17.3	30.3	27.0	10.5	3.1	2.9	3.5	23.4	7.8
Burr type III	20	1.3	1.4	1.5	1.4	1.5	0.9	1.2	0.8	0.8	1.5	1.3
	50	2.1	2.7	3.0	2.4	2.7	2.3	2.2	1.7	2.5	1.8	2.9
	100	3.8	5.3	5.8	4.5	5.2	5.1	3.5	2.4	4.0	2.7	5.7
Burr type XII	20	0.9	1.0	1.0	1.0	1.0	0.7	1.0	0.6	0.6	1.2	1.0
	50	1.1	1.0	1.0	1.0	1.0	0.8	1.0	1.0	1.2	1.0	1.1
	100	1.2	1.1	1.1	1.2	1.1	0.8	1.2	1.1	1.6	1.1	1.2
Davies	20	1.0	1.1	1.2	1.1	1.1	0.8	1.1	0.8	0.7	1.3	1.2
	50	1.4	1.6	1.7	1.5	1.6	1.3	1.9	1.5	2.1	1.3	2.0
	100	2.0	2.4	2.7	2.2	2.4	2.1	2.9	2.0	3.3	1.7	3.3

Table 5.2: Power comparison of normality tests (5%)

Distributions	**n**	**KS**	**CvM**	**AD**	**V**	**U^2**	**BHEP**	**JB**	**MS**	**MK**	**P**	**SW**
Normal	20	4.8	4.9	4.9	4.8	5.0	4.8	4.9	2.8	1.3	4.9	4.9
	50	5.1	5.1	5.1	5.1	5.1	5.5	5.0	4.1	2.5	5.2	5.1
	100	5.2	5.1	5.1	5.2	5.1	5.5	5.0	4.5	3.1	5.2	4.8
Generalized secant hyperbolic pi	20	5.0	5.3	5.3	5.5	5.5	5.0	5.0	3.1	1.5	5.0	5.5
	50	5.7	6.0	6.2	6.1	6.1	6.4	6.2	5.3	3.5	5.7	6.8
	100	6.2	6.7	6.9	6.9	6.9	7.3	7.3	6.3	5.3	6.1	8.0
Quantile-based flattened logistic	20	4.9	5.2	5.1	5.3	5.3	4.9	5.0	3.0	1.5	5.0	5.4
	50	5.5	5.7	5.8	5.8	5.7	6.1	6.0	5.1	3.4	5.4	6.4
	100	5.8	6.1	6.3	6.3	6.3	6.7	7.0	6.1	5.0	5.8	7.4
Tukey lambda ($\lambda = 0.13491$)	20	4.8	4.9	4.9	4.8	5.0	4.8	4.9	2.7	1.3	4.9	4.8
	50	5.1	5.1	5.0	5.1	5.1	5.5	4.6	4.0	2.2	5.1	4.9
	100	5.2	5.1	5.1	5.1	5.1	5.5	4.4	4.2	2.6	5.2	4.5
Tukey lambda ($\lambda = 5.2029$)	20	13.0	14.4	13.0	15.1	15.3	10.9	5.1	2.4	1.4	12.9	8.6
	50	29.3	34.2	31.6	36.6	37.2	28.1	1.9	1.3	1.0	29.2	17.4
	100	*56.8*	*65.6*	*64.1*	*67.6*	*69.1*	*63.9*	0.8	0.7	0.6	*56.0*	47.2
Generalized Tukey lambda	20	21.5	28.8	32.4	27.3	30.0	25.5	7.9	5.3	2.7	21.7	34.2
	50	50.0	*62.8*	*72.3*	*62.5*	*65.1*	*65.8*	12.6	9.8	10.1	*59.9*	**75.8**
	100	**82.5**	**90.3**	**96.6**	**91.9**	**91.8**	**94.7**	29.7	13.5	25.2	**89.2**	**98.4**
Schmeiser-Deutsch	20	28.9	30.3	26.1	34.1	32.7	15.4	5.8	2.5	1.8	32.9	15.6
	50	*65.1*	*69.4*	*63.9*	*73.9*	*73.0*	43.7	2.2	1.0	1.3	**75.8**	41.6
	100	**93.5**	**96.3**	**95.3**	**97.1**	**97.2**	**86.9**	0.9	0.5	0.8	**97.9**	**87.1**
Two-tailed gamma	20	12.5	12.3	10.9	13.4	13.2	7.7	5.2	3.5	1.5	11.8	9.3
	50	26.5	25.8	21.7	29.3	28.3	15.9	7.1	6.4	4.0	25.0	14.9
	100	*50.7*	*52.9*	44.2	*58.0*	*56.7*	34.9	8.7	8.1	7.1	49.7	25.1
Burr type III	20	5.9	6.5	6.5	6.4	6.5	6.2	4.4	2.8	1.5	5.7	6.3
	50	8.7	9.9	10.4	9.3	11.2	10.0	5.1	4.6	3.5	7.6	10.0
	100	12.8	15.8	17.1	14.4	16.2	29.6	6.8	6.1	6.1	9.7	16.7
Burr type XII	20	4.7	5.0	4.8	4.8	4.9	4.9	4.9	2.8	1.3	4.9	4.9
	50	5.1	5.2	5.2	5.2	5.2	5.7	4.7	4.0	2.4	5.2	5.1
	100	5.5	5.4	5.4	5.4	5.3	5.8	4.9	4.4	3.1	5.3	5.1
Davies	20	5.3	5.5	5.4	5.5	5.6	5.4	4.5	2.7	1.4	5.2	5.5
	50	6.5	7.0	7.3	6.7	7.1	7.8	5.0	4.4	3.2	6.1	7.4
	100	8.3	9.3	9.8	8.8	9.3	10.6	6.2	5.7	5.1	6.9	10.3

Table 5.3: Power comparison of normality tests (10%)

Distributions	**n**	**KS**	**CvM**	**AD**	**V**	**U^2**	**BHEP**	**JB**	**MS**	**MK**	**P**	**SW**
Normal	20	9.7	9.9	9.8	9.9	9.7	10.5	10.0	5.9	2.0	12.3	10.1
	50	10.3	10.2	10.1	10.1	10.2	11.4	10.0	8.2	3.9	10.2	10.0
	100	10.3	10.1	10.2	10.3	10.1	11.3	10.0	9.1	5.6	10.3	9.7
Generalized secant hyperbolic pi	20	10.1	10.6	10.6	10.7	10.6	10.8	9.6	6.1	2.2	12.7	11.1
	50	11.3	11.9	12.0	11.8	12.0	13.0	11.0	9.3	5.4	11.2	12.9
	100	12.3	12.9	13.2	13.2	13.2	14.5	12.7	11.3	9.4	12.0	14.5
Quantile-based flattened logistic	20	10.0	10.3	10.3	10.5	10.3	10.7	9.6	6.1	2.2	12.6	10.8
	50	11.0	11.3	11.4	11.3	11.4	12.5	10.8	9.1	5.2	10.8	12.2
	100	11.6	11.9	12.2	12.1	12.2	13.3	12.3	10.9	8.8	11.4	13.6
Tukey lambda ($\lambda = 0.13491$)	20	9.7	9.9	9.8	9.9	9.6	10.5	10.1	5.9	2.0	12.3	9.9
	50	10.3	10.2	10.0	10.1	10.1	11.4	9.8	8.0	3.6	10.2	9.6
	100	10.3	10.1	10.1	10.3	10.1	11.3	9.4	8.7	4.9	10.3	9.1
Tukey lambda ($\lambda = 5.2029$)	20	21.4	23.2	21.7	23.9	24.1	18.9	11.3	5.1	2.3	25.4	15.8
	50	41.9	47.0	44.7	48.7	49.8	41.3	5.6	3.1	2.0	41.0	30.4
	100	*70.1*	**77.2**	**76.5**	**77.8**	**79.8**	**76.7**	2.9	2.1	1.6	*68.1*	*65.7*
Generalized Tukey lambda	20	33.6	42.0	46.5	40.3	42.8	40.4	14.4	10.5	3.9	36.0	49.8
	50	*64.1*	*74.1*	**82.9**	*74.1*	**75.9**	**78.0**	39.0	18.4	23.7	*70.2*	**86.2**
	100	**90.9**	**94.8**	**98.7**	**95.7**	**95.6**	**97.7**	46.8	22.9	38.7	**93.8**	**99.4**
Schmeiser-Deutsch	20	40.9	42.1	37.8	45.7	44.2	24.5	12.1	5.1	2.8	49.0	25.9
	50	**76.6**	**80.2**	**76.3**	**82.8**	**82.6**	*57.7*	5.8	2.5	2.5	**83.4**	*59.5*
	100	**96.9**	**98.3**	**98.0**	**98.6**	**98.8**	**93.5**	3.0	1.4	2.1	**98.9**	**94.9**
Two-tailed gamma	20	22.1	22.3	20.1	23.3	22.9	16.2	9.6	6.8	2.3	24.2	17.8
	50	40.3	40.5	35.2	43.5	42.9	27.8	12.7	10.9	7.1	38.2	25.8
	100	*65.9*	*68.9*	*60.9*	*72.2*	*72.2*	*50.8*	16.3	13.5	13.3	*65.2*	39.4
Burr type III	20	11.7	12.6	12.5	12.2	12.5	12.9	8.5	5.5	2.1	13.8	12.4
	50	15.5	17.3	17.9	16.6	17.6	19.9	9.3	8.3	5.1	14.0	17.5
	100	20.9	25.0	26.6	23.3	25.6	29.6	13.1	10.6	10.8	17.5	26.5
Burr type XII	20	9.7	10.0	9.9	10.0	9.8	10.6	9.9	5.8	2.0	12.3	10.0
	50	10.4	10.4	10.3	10.2	10.3	11.5	9.7	8.0	3.8	10.3	10.0
	100	10.6	10.6	10.6	10.5	10.6	11.9	9.7	8.9	5.4	10.6	10.0
Davies	20	10.5	11.0	11.0	10.8	10.9	11.3	8.9	5.5	2.1	12.9	11.0
	50	12.4	13.1	13.4	12.8	13.2	11.5	9.2	8.2	4.5	11.8	13.3
	100	14.8	16.4	17.1	15.6	16.6	19.1	11.3	10.0	8.6	13.3	17.5

5.2 Results from the simulation study for goodness-of-fit tests

The results from the simulation study are tabled in Tables 5.1, 5.2 and 5.3 for significance levels of 1%, 5% and 10% respectively. Each tabulated value gives the percentage of times the null hypothesis of normality is incorrectly rejected in the case of the normal distribution and is correctly rejected for all the other mesokurtic distributions. I.e., for the normal distribution the tabulated values are the empirical type I errors, while for the other distributions the values are the empirical powers. Empirical power values between 50% and 74% in the tables are italicized and values greater and equal to 75% are in bold.

Section 5.8 graphically illustrates the distributions of the simulated p-values for each goodness-of-fit test and mesokurtic distribution for sample sizes $n = 20$, $n = 50$ and $n = 100$ by presenting their probability histograms. We used 20 bins of size 0.05 will be used when plotting these probability histograms. Therefore, these probability histograms are useful to show the empirical type I errors for the normal distribution and the empirical power for the other mesokurtic distributions of the goodness-of-fit tests at 5% and 10% significance levels - it is not possible to observe the empirical type I errors or empirical power at a 1% level in these graphs.

5.3 Empirical type I errors

As indicated before, the tabulated values for the normal distribution in Tables 5.1, 5.2 and 5.3 are the empirical type I errors, which should correspond to the respective significance levels. For most of the goodness-of-fit tests this is the case. Exceptions include the BHEP test, which is conservative at a 1% level for all sample sizes (i.e., the empirical type I errors are less than 1%) and is liberal for $n = 50$ and $n = 100$ at 5% and 10% levels (the empirical type I errors are more than the significance levels). The Mardia skewness and kurtosis tests are conservative at a 1% significance level for $n = 20$ and at 5% and 10% levels for all sample sizes. The Pearson χ^2 test is liberal for $n = 20$ at a 10% significance level.

For the normal distribution, the probability histogram of each goodness-of-fit test in Figures 5.1 and 5.2 should be uniform at 0.05. This is again the case for most of the goodness-of-fit tests. However, this is clearly not true for the Mardia skewness and kurtosis tests which, as indicated in the previous paragraph, are conservative. The other test whose probability histograms are not uniform is the Pearson χ^2 test, in particular for $n = 20$. The p-values for this test are unevenly distributed from 0 to 1. The reason for this is that a small number of distinct p-values were generated due to the use of bins. Finally, it is interesting to note that, although uniform in general over the intervals of p-values, the probability histograms of the BHEP test have increased probabilities greater than 0.05 for p-values between 0.95 and 1. This implies that the probabilities are slightly less than 0.05 at some of the other intervals of p-values.

5.4 Power against Group 1 distributions

Because the distributions in Group 1 closely approximates the standard normal distribution, it is not surprising that all the normality tests have extremely low power against these alternatives - see Tables 5.1, 5.2 and 5.3. In most cases the power against the Group 1 distributions is equivalent to the empirical type I errors of the standard normal distribution. Only for larger sample sizes ($n = 50$ and $n = 100$) the power is slightly larger than the empirical type I errors. However, for the Tukey lambda distribution with $\lambda = 0.13491$, the moment-based tests as well as the SW test have power less than the empirical type I errors at all significance levels and sample sizes.

The probability histograms in Figures 5.3 to 5.8 for the Group 1 distributions are similar to the probability histograms for the standard normal distribution in Figures 5.1 and 5.2. This is again due to the close approximation of the standard normal distribution by these mesokurtic distributions.

5.5 Power against Group 2 distributions

The distributions in Group 2 differ drastically from the standard normal distribution with respect to distributional shape and support. All these distributions are truncated and thus have bounded support. Hence, the goodness-of-fit tests should be more powerful when testing for normality against the Group 2 distributions compared to the tests applied to the Group 1 distributions in the previous section.

We see in Tables 5.1, 5.2 and 5.3 as well as Figures 5.9 to 5.14 that the EDF, BHEP, Pearson χ^2 and SW tests indeed have high power against the Group 2 distributions. For the Tukey lambda distribution with $\lambda = 5.2029$ and the Schmeiser-Deutsch distribution, which are both symmetric distributions, the EDF tests have higher power than the BHEP and SW tests at all sample sizes. Specifically, the Kuiper and Watson U^2 tests perform the best for the Tukey lambda distribution with $\lambda = 5.2029$. Among all the tests, the Pearson χ^2 test is the most powerful for the Schmeiser-Deutsch distribution. For the generalized Tukey lambda distribution, which is asymmetric, the SW test has the highest power.

The moment-based tests perform extremely poorly for the truncated distributions in Group 2. This is to be expected, because these tests are only based on the skewness and/or the kurtosis and not the distributional shape. For the two symmetric truncated distributions (Tukey lambda with $\lambda = 5.2029$ and Schmeiser-Deutsch) we see that the power of the moment-based tests decreases as the sample size increases, which is opposite to what is anticipated. The reason for this anomaly is that the sample Pearson's moment coefficients of skewness and kurtosis for the simulated data are closer to 0 and 3 respectively for larger sample sizes than for smaller sample sizes.

5.6 Power against Group 3 distributions

Group 3 consists of just the two-tailed gamma distribution which is bimodal with infinite support. The probability histograms depicting the distributions of the simulated p-values for this distribution are given in Figures 5.15 and 5.16.

As with the distributions in Group 2, the moment-based tests again have by far the worst performance in terms of power. The BHEP and SW tests also have lower power compared to the EDF and Pearson's χ^2 tests. None of the goodness-of-fit tests have powers more than 75% irrespective of sample size or significance level considered. The most powerful test for the two-tailed gamma distribution is the Kuiper test.

5.7 Power against Group 4 distributions

Similar to the distributions from Group 1, the distributions in Group 4, in particular the Burr type XII distribution, closely approximate the standard normal distribution, but not as well as the Group 1 distributions (the reason for this is that the Group 4 distributions have half-infinite support). Consequently, as can be seen in Tables 5.1 to 5.3 and in Figures 5.17 to 5.22, the power of the normality tests is rather low against Group 4 alternatives.

In general the power is the highest when testing against the Burr type III distribution and the lowest against the Burr type XII distribution. For none of the tests or distributions a power of more than 30% is observed. There is no test that consistently performs best over all alternatives from Group 4 and for all different sample sizes. However, as before, the moment-based tests clearly perform the worst among the normality tests.

5.8 Appendix

In this appendix the distributions of the simulated p-values for the various goodness-of-fit tests and mesokurtic distributions are plotted using probability histograms.

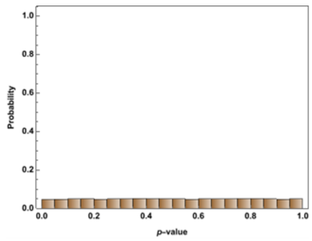
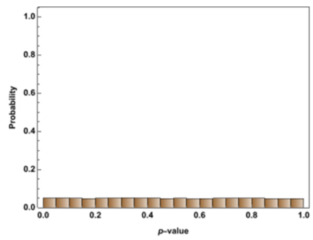
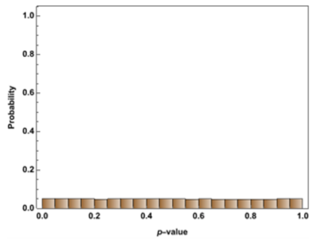
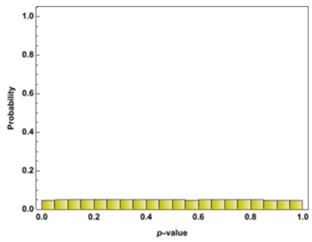
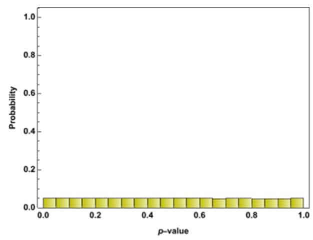
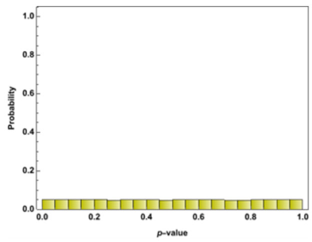
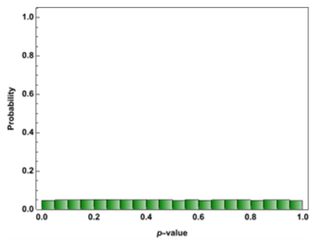
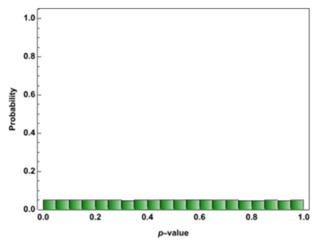
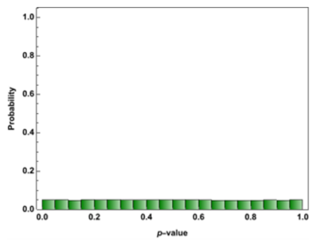
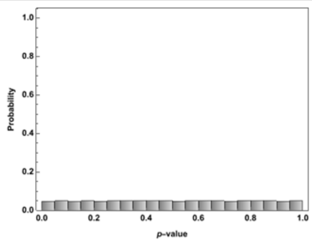
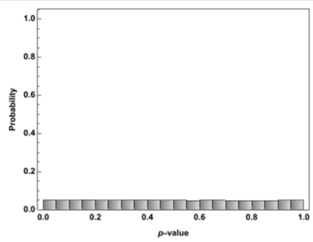
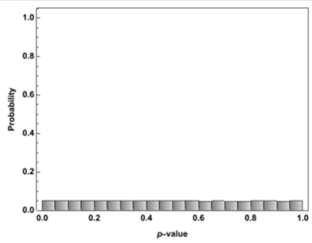
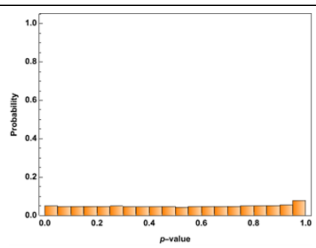
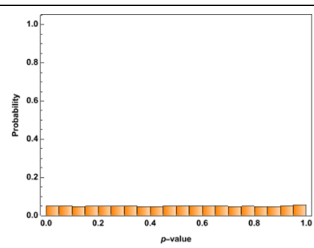
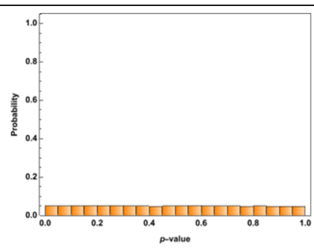
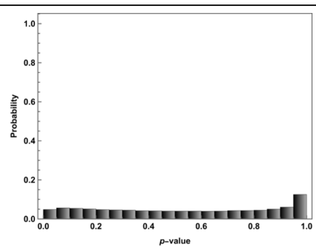
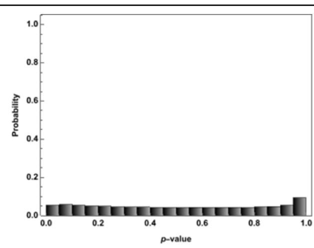
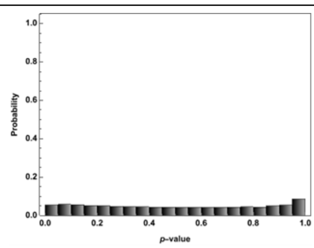
Test	$n = 20$	$n = 50$	$n = 100$
KS			
CvM			
AD			
V			
U^2			
BHEP			

Figure 5.1: Power comparison of EDF and BHEP goodness-of-fit tests for the normal distribution

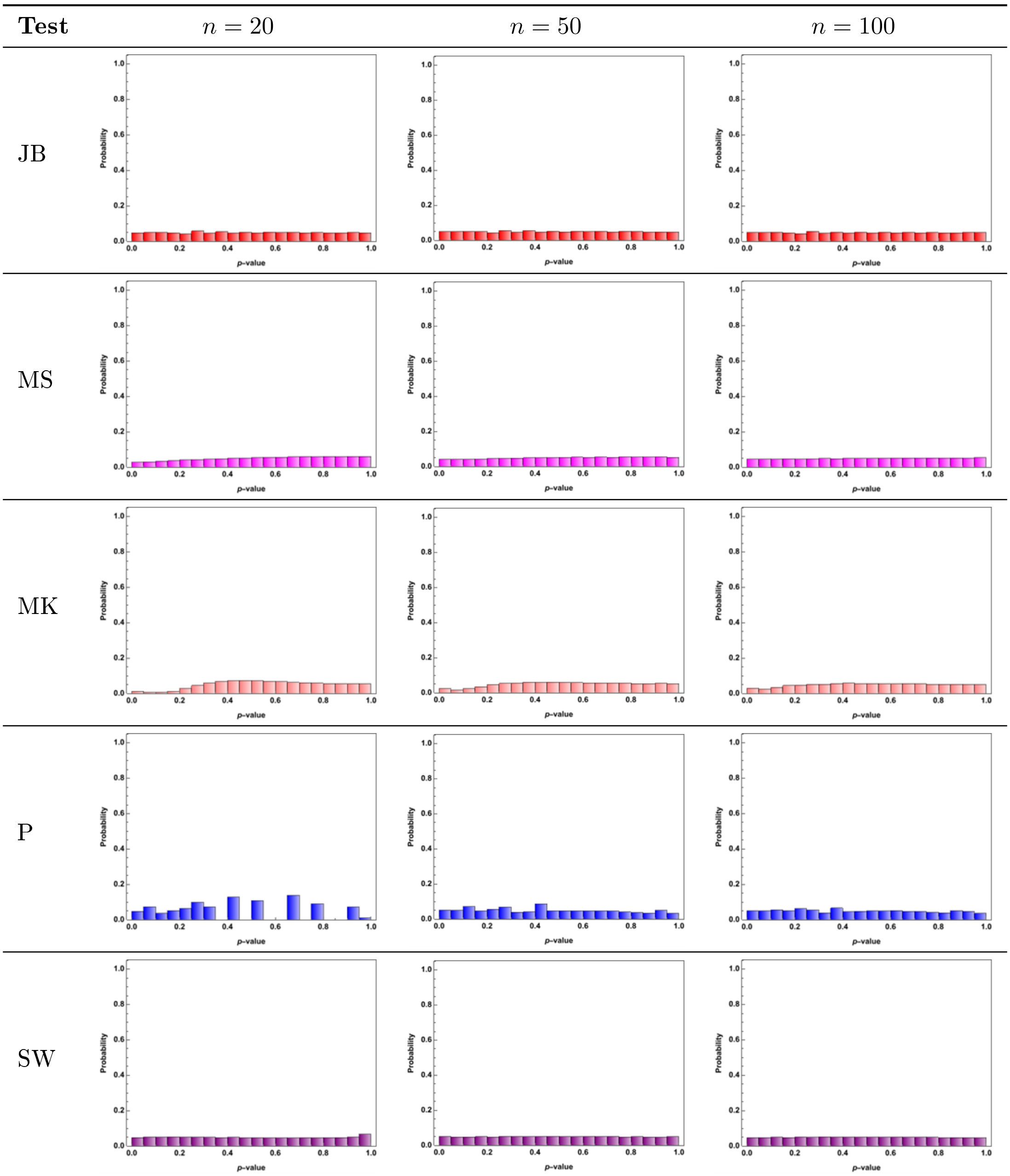


Figure 5.2: Power comparison of moment-based, Pearson χ^2 and SW goodness-of-fit tests for the normal distribution

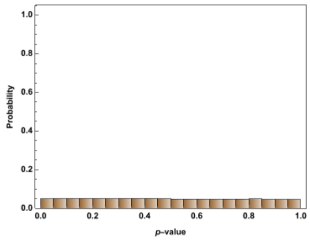
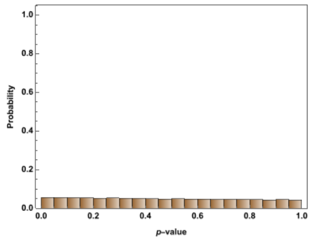
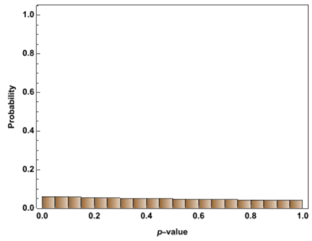
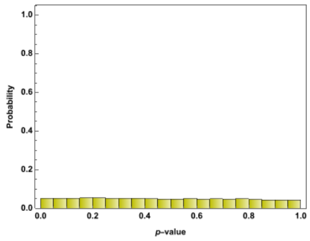
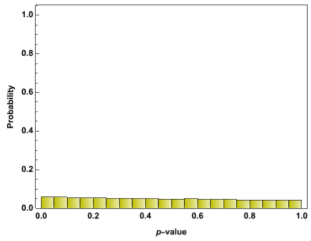
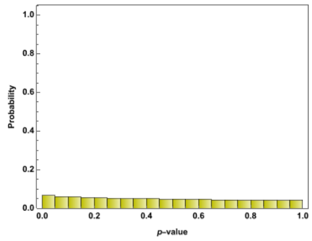
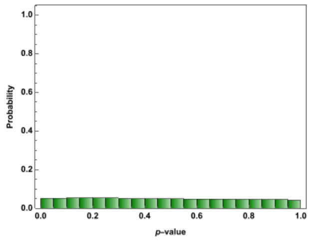
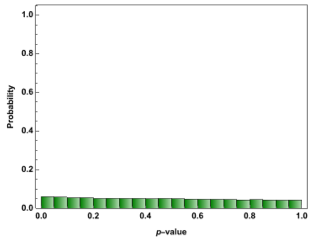
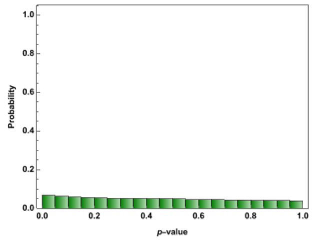
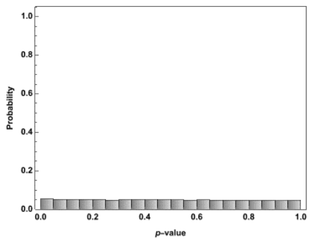
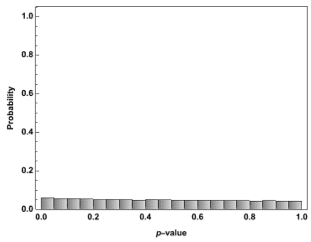
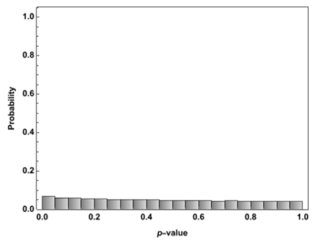
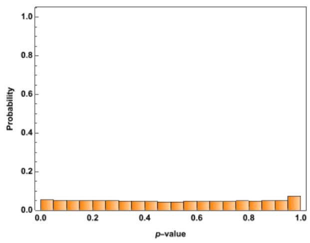
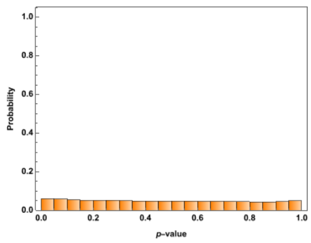
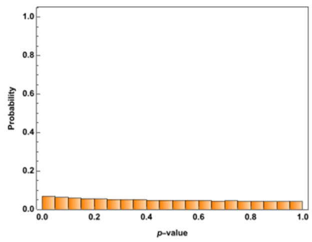
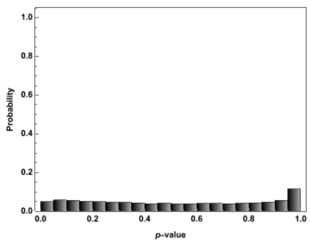
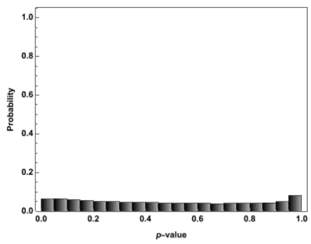
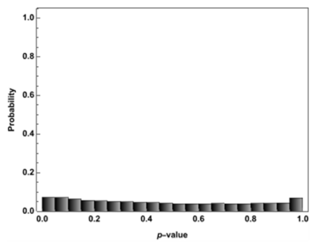
Test	$n = 20$	$n = 50$	$n = 100$
KS			
CvM			
AD			
V			
U^2			
BHEP			

Figure 5.3: Power comparison of EDF and BHEP goodness-of-fit tests for the generalized secant hyperbolic pi distribution

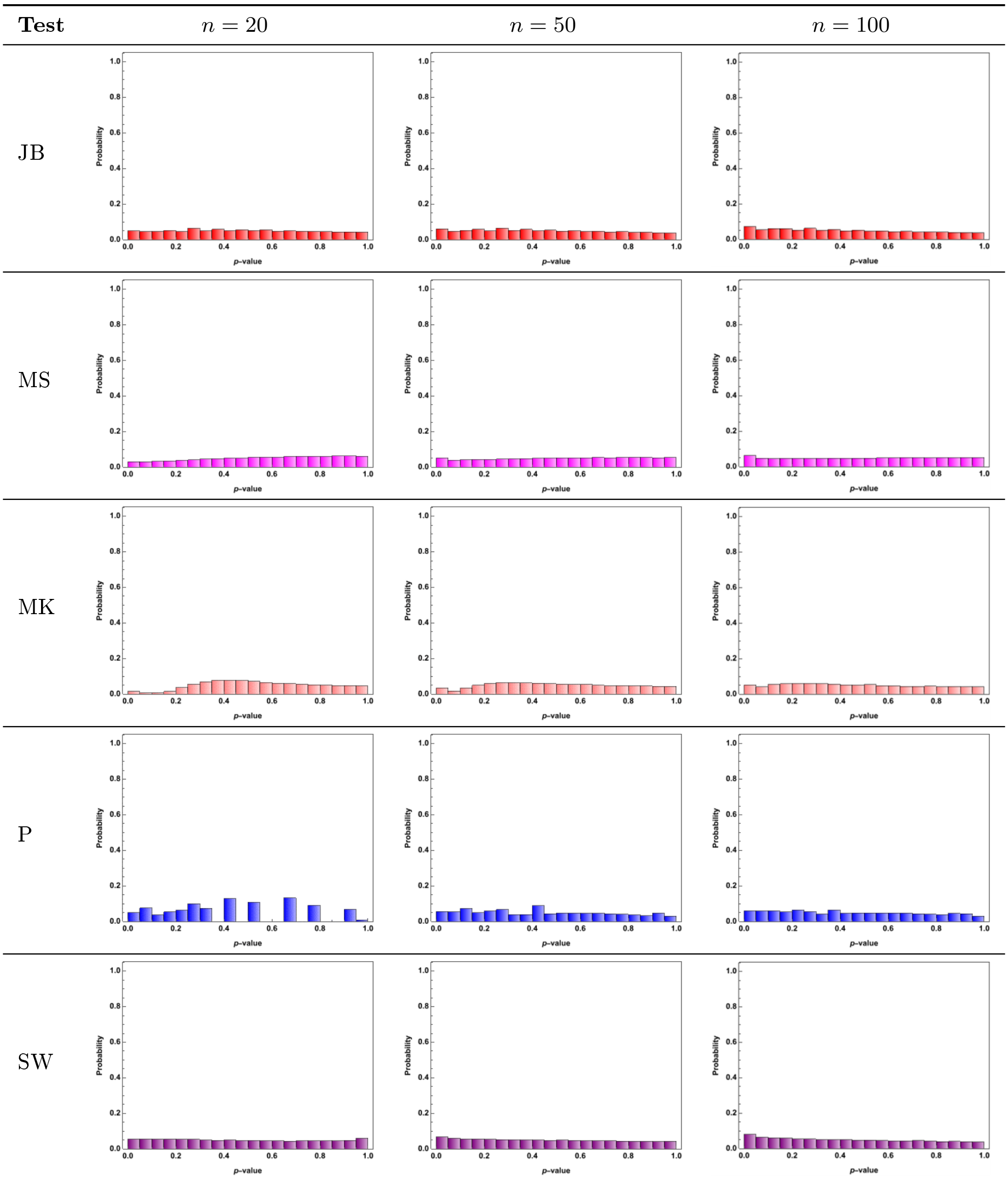


Figure 5.4: Power comparison of moment-based, Pearson χ^2 and SW goodness-of-fit tests for the generalized secant hyperbolic pi distribution

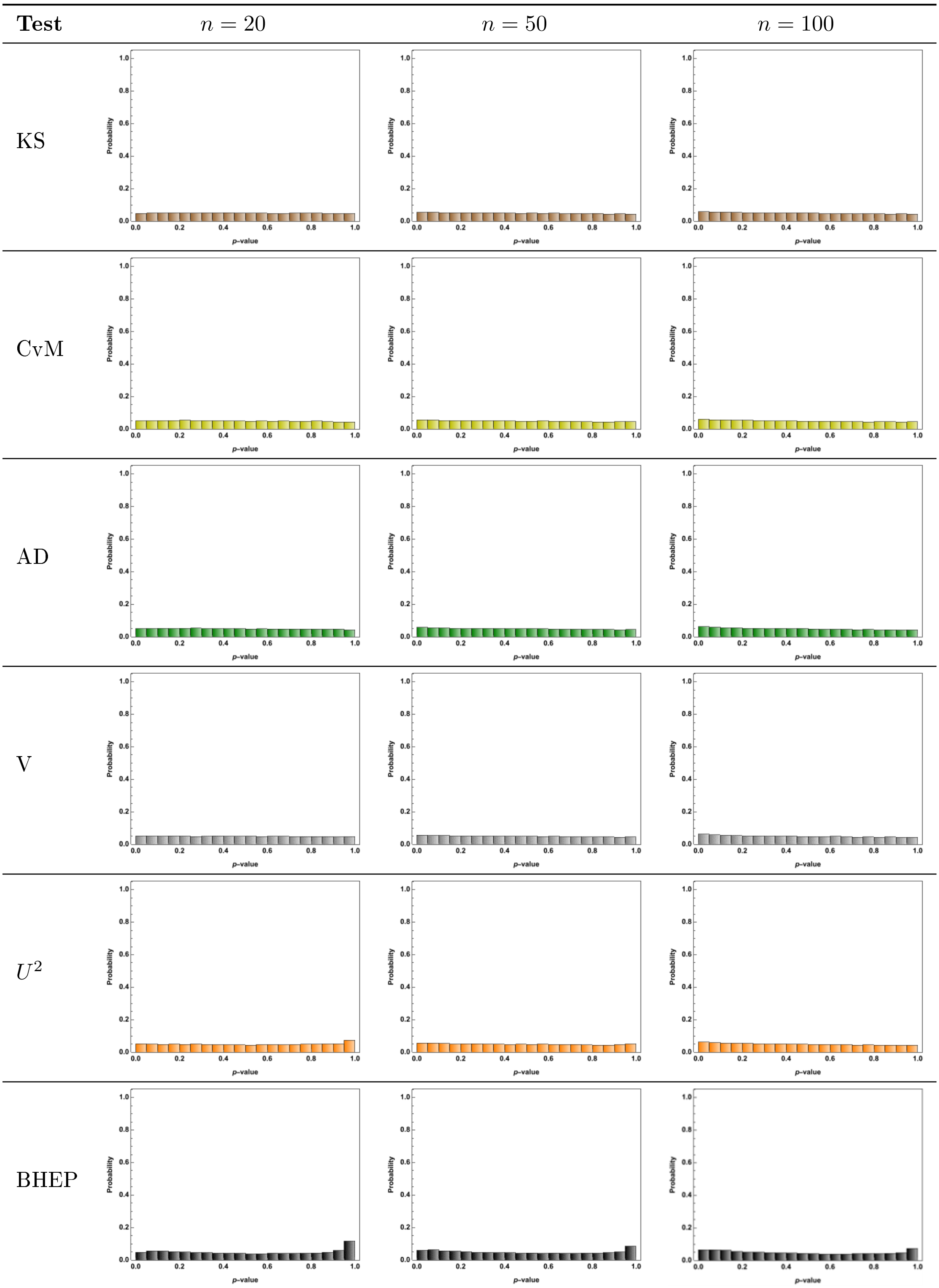


Figure 5.5: Power comparison of EDF and BHEP goodness-of-fit tests for the quantile-based flattened logistic distribution

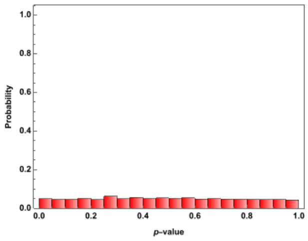
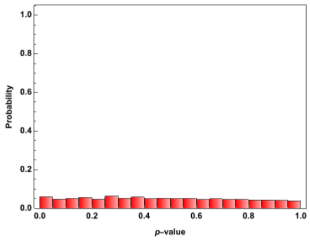
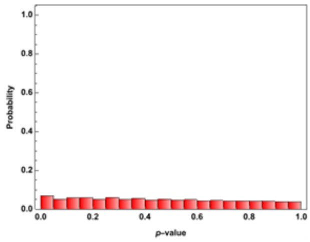
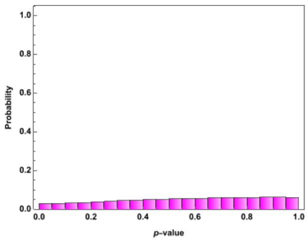
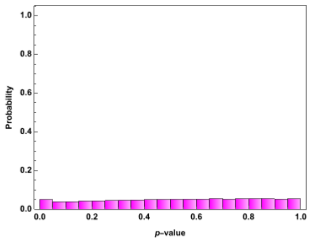
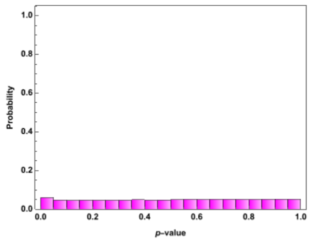
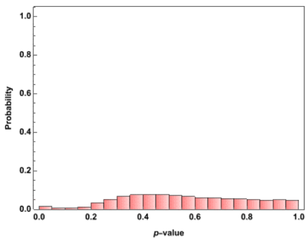
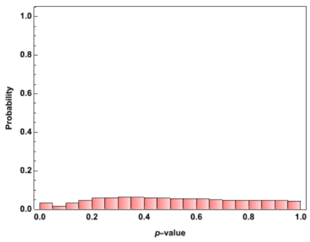
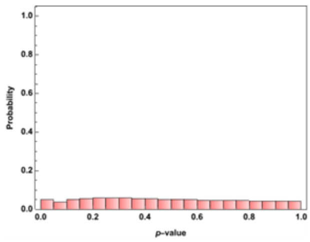
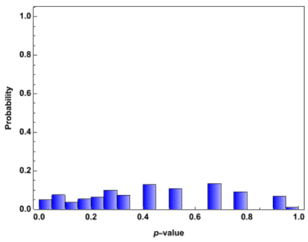
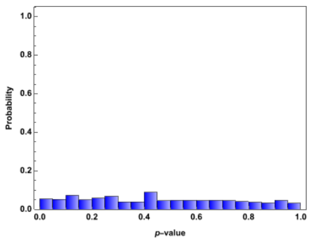
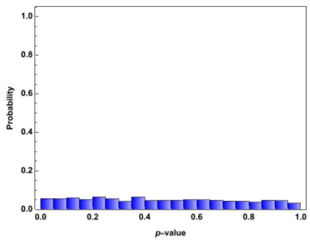
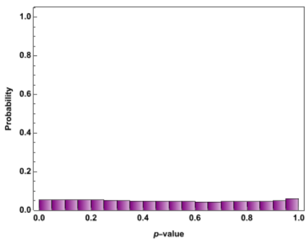
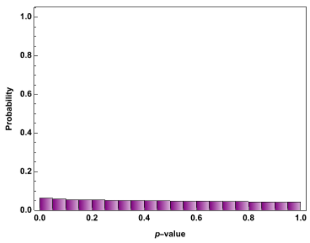
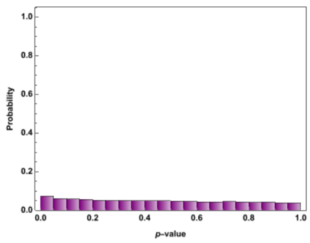
Test	$n = 20$	$n = 50$	$n = 100$
JB			
MS			
MK			
P			
SW			

Figure 5.6: Power comparison of moment-based, Pearson χ^2 and SW goodness-of-fit tests for the quantile-based flattened logistic distribution

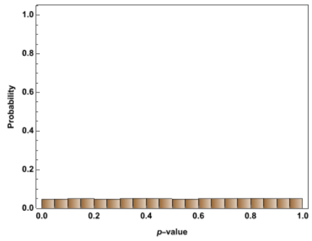
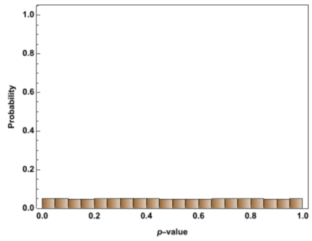
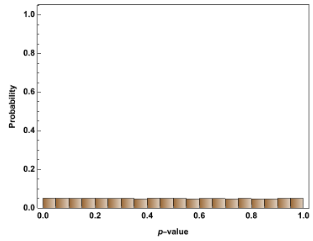
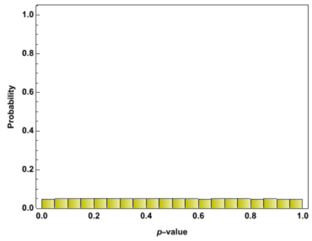
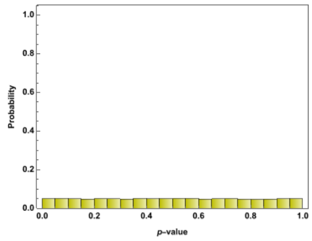
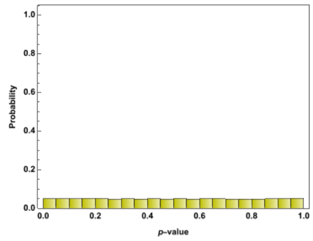
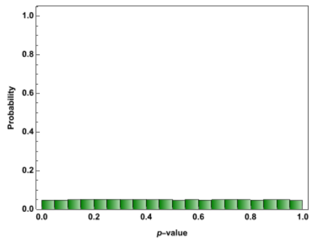
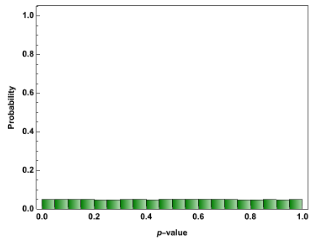
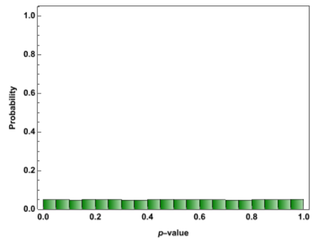
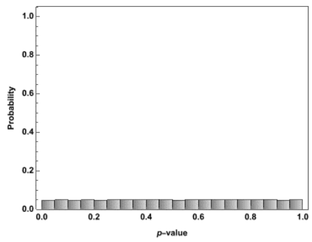
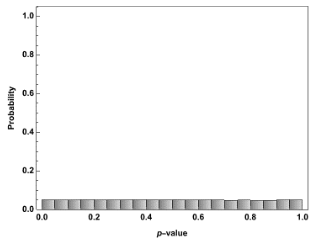
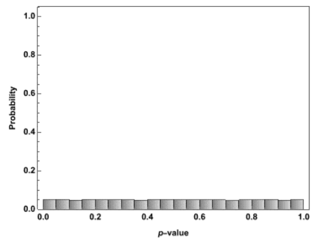
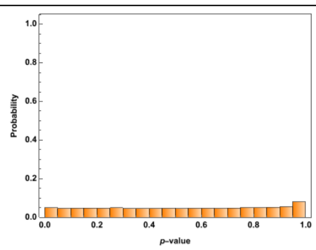
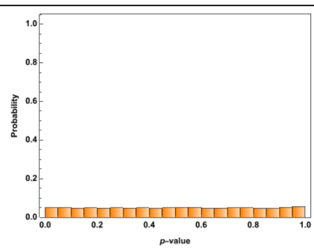
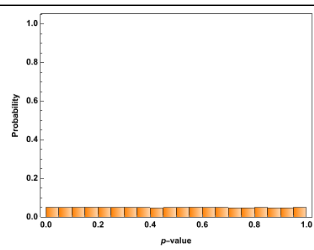
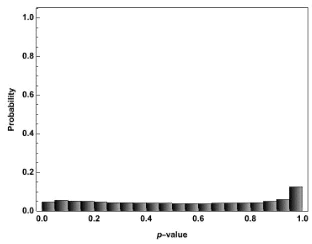
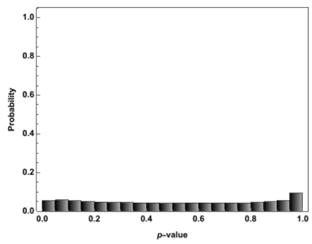
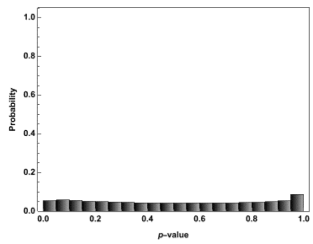
Test	$n = 20$	$n = 50$	$n = 100$
KS			
CvM			
AD			
V			
U^2			
BHEP			

Figure 5.7: Power comparison of EDF and BHEP goodness-of-fit tests for the Tukey lambda distribution with $\lambda = 0.1349$

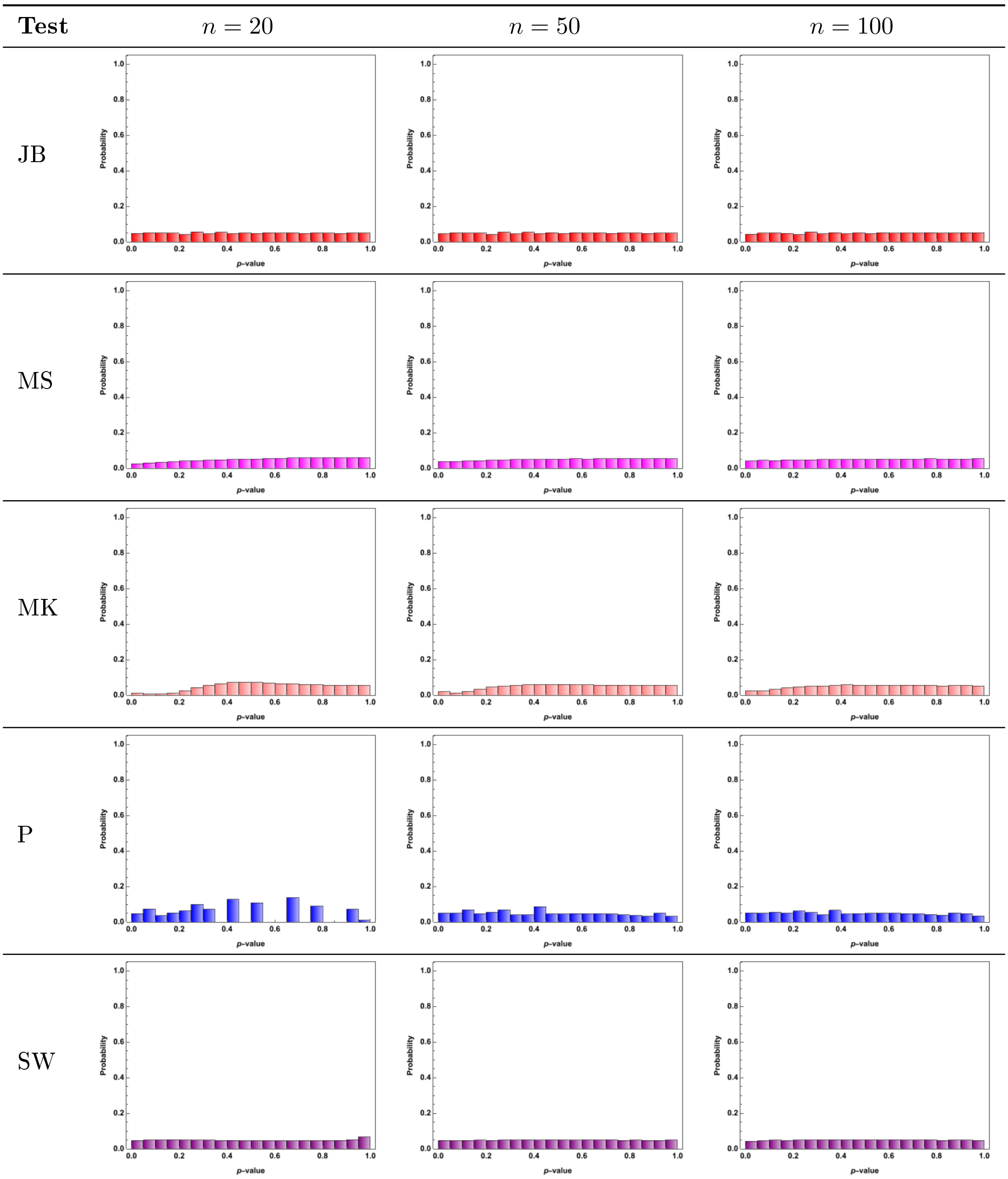


Figure 5.8: Power comparison of moment-based, Pearson χ^2 and SW goodness-of-fit tests for the Tukey lambda distribution with $\lambda = 0.1349$

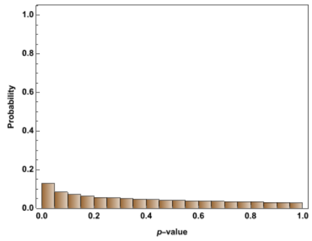
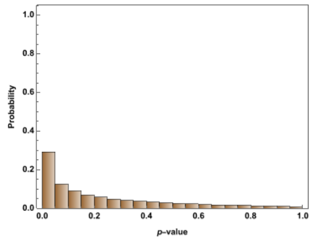
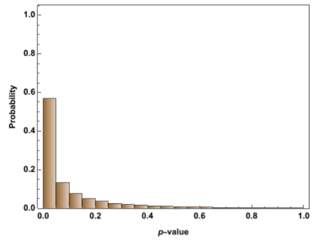
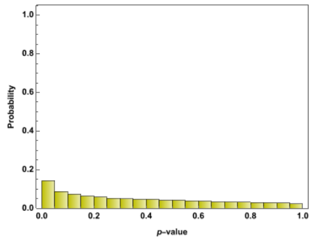
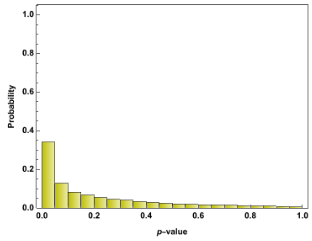
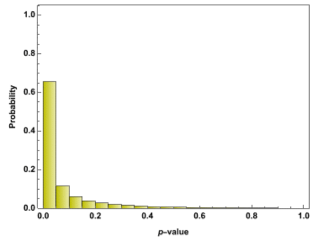
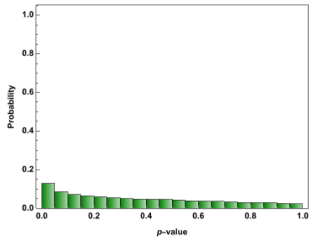
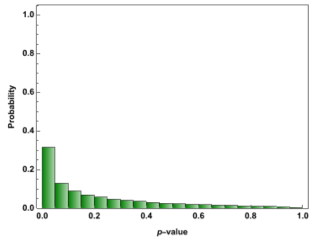
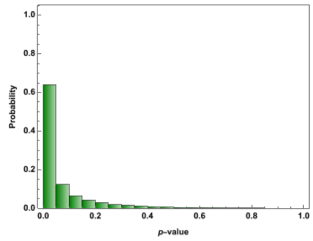
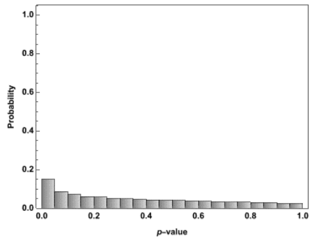
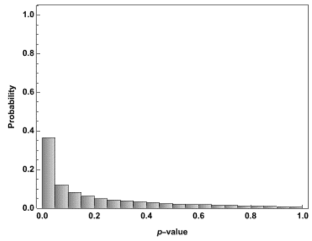
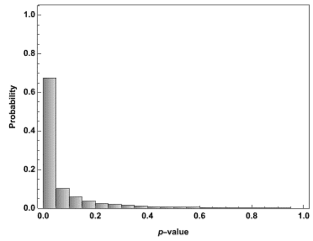
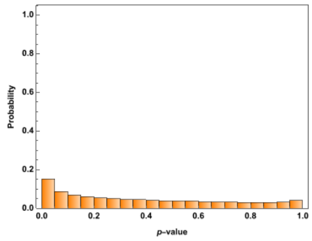
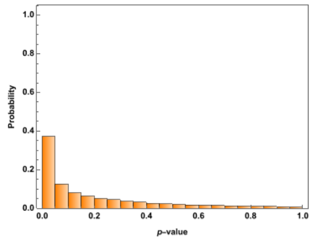
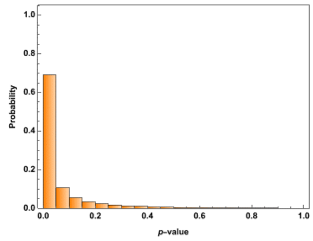
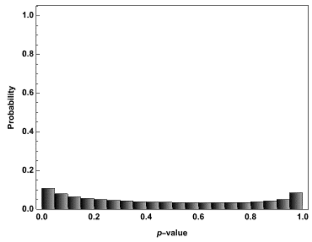
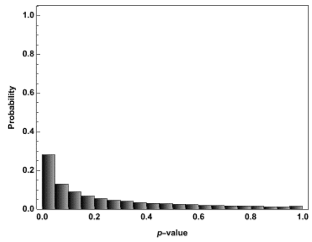
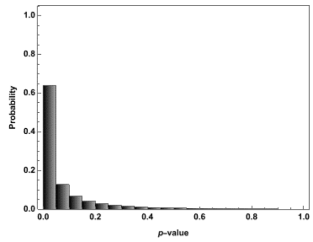
Test	$n = 20$	$n = 50$	$n = 100$
KS			
CvM			
AD			
V			
U^2			
BHEP			

Figure 5.9: Power comparison of EDF and BHEP goodness-of-fit tests for the Tukey lambda distribution with $\lambda = 5.2029$

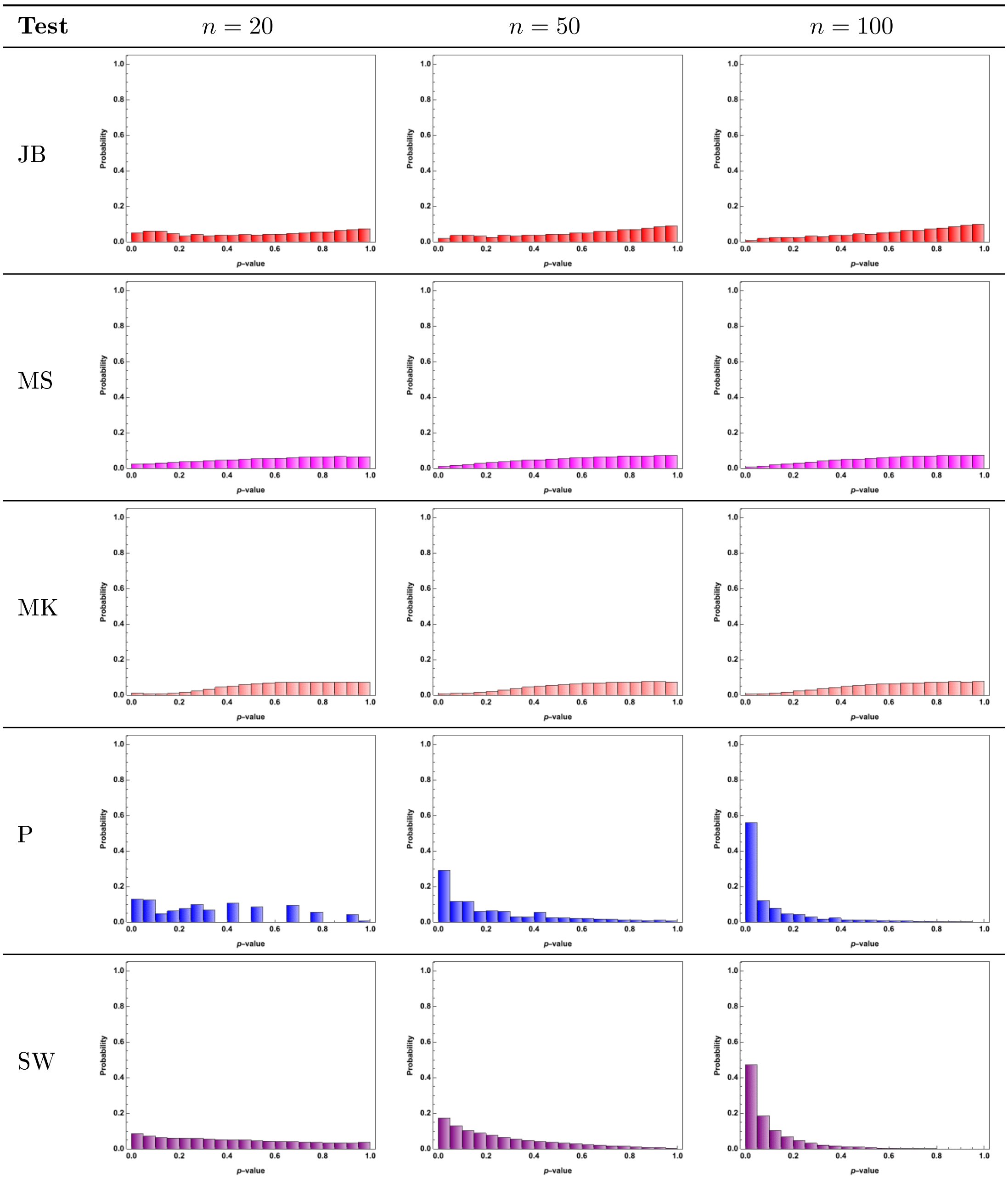


Figure 5.10: Power comparison of moment-based, Pearson χ^2 and SW goodness-of-fit tests for the Tukey lambda distribution with $\lambda = 5.2029$

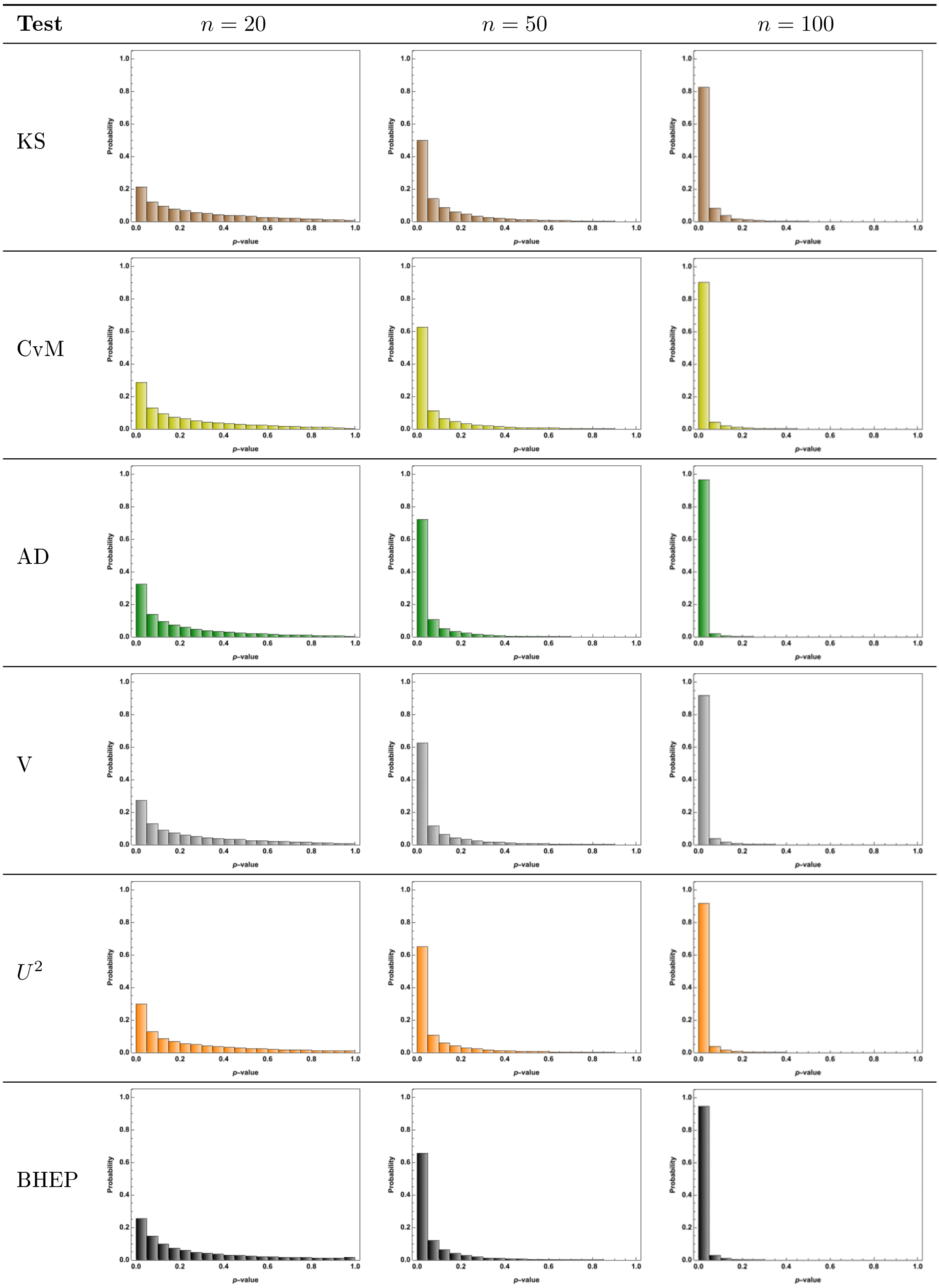


Figure 5.11: Power comparison of EDF and BHEP goodness-of-fit tests for the generalized Tukey lambda distribution

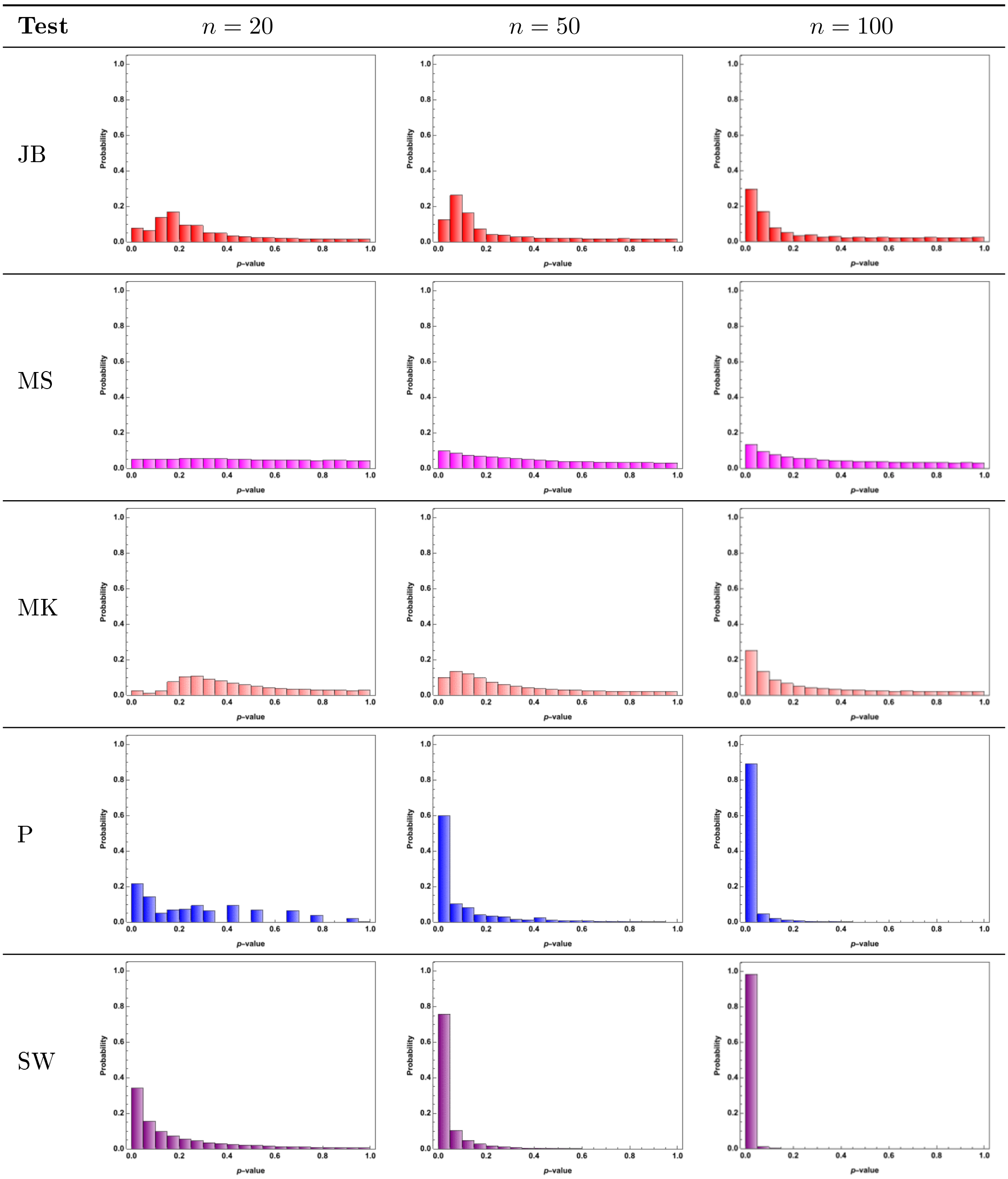


Figure 5.12: Power comparison of moment-based, Pearson χ^2 and SW goodness-of-fit tests for the generalized Tukey lambda distribution

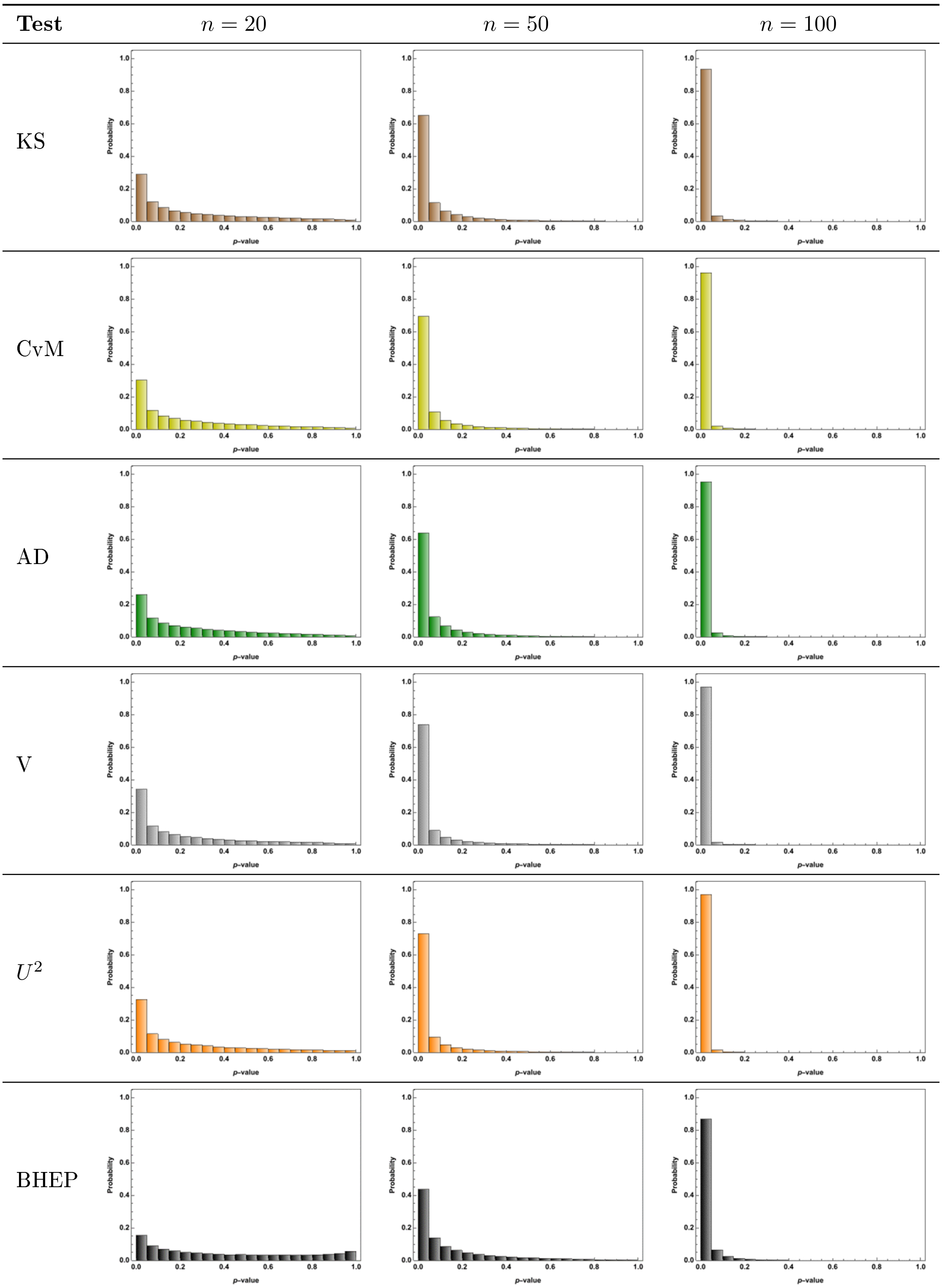


Figure 5.13: Power comparison of EDF and BHEP goodness-of-fit tests for the Schmeiser-Deutsch distribution

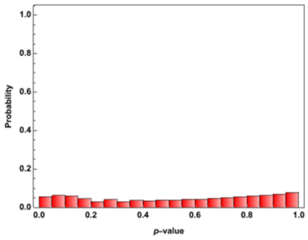
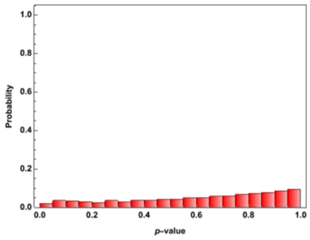
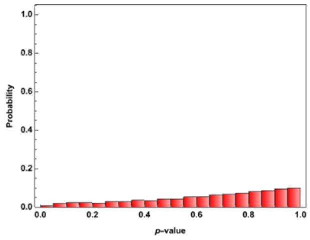
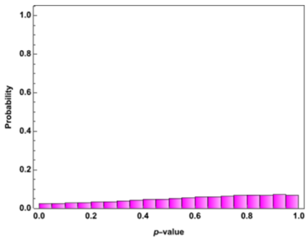
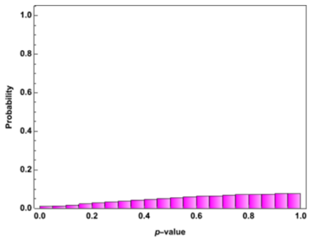
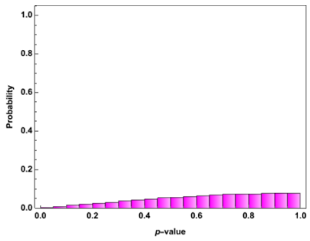
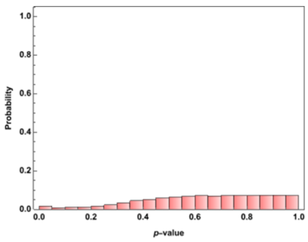
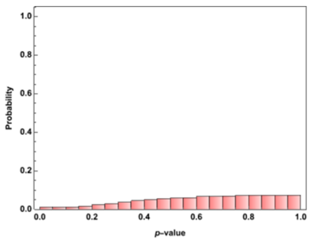
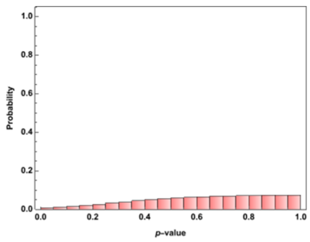
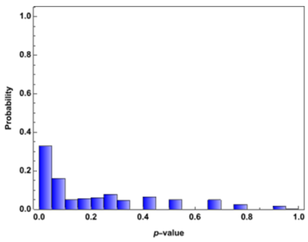
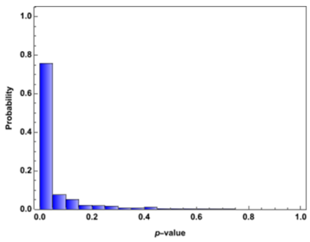
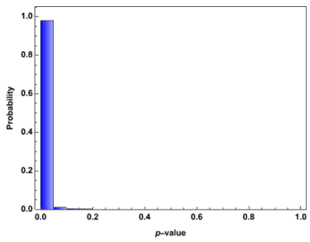
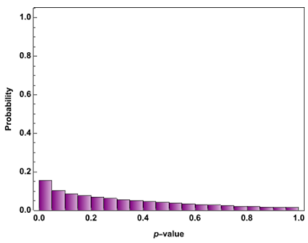
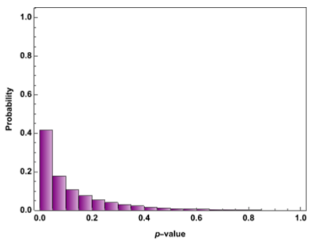
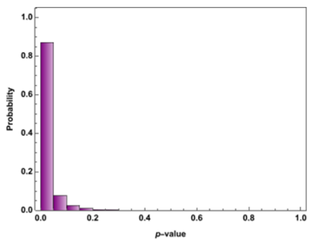
Test	$n = 20$	$n = 50$	$n = 100$
JB			
MS			
MK			
P			
SW			

Figure 5.14: Power comparison of moment-based, Pearson χ^2 and SW goodness-of-fit tests for the Schmeiser-Deutsch distribution

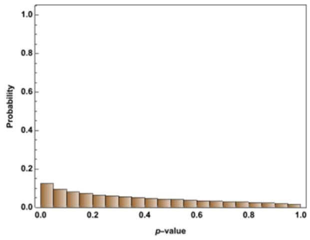
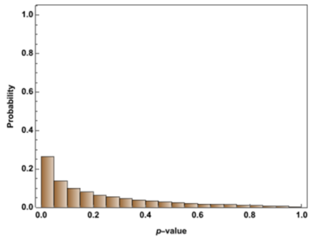
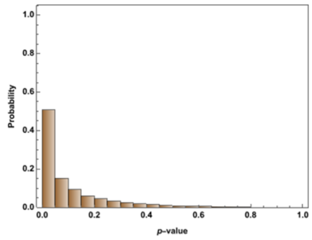
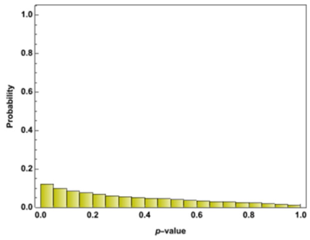
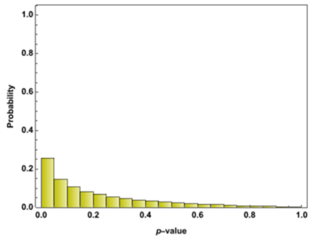
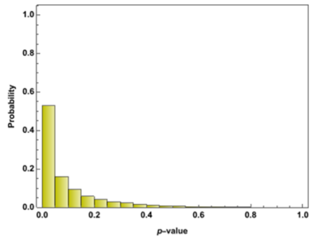
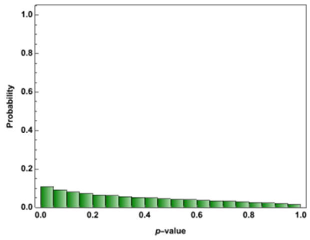
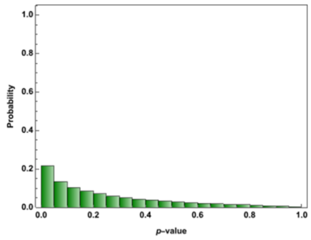
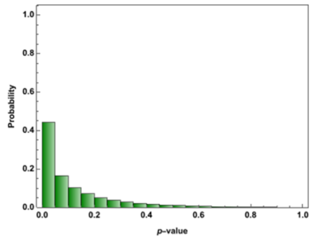
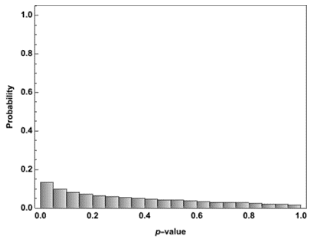
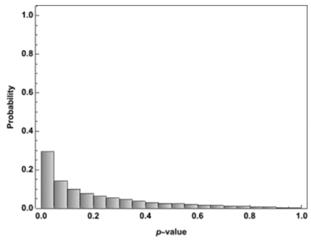
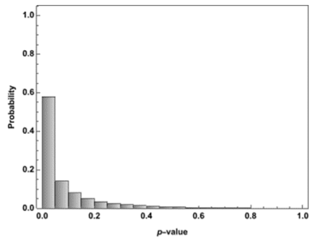
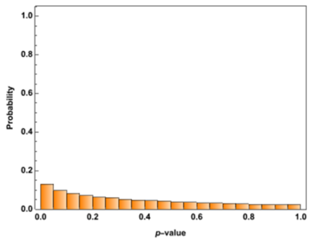
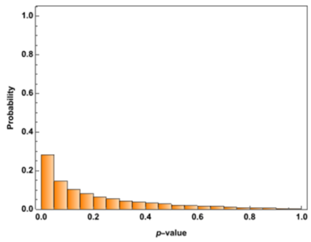
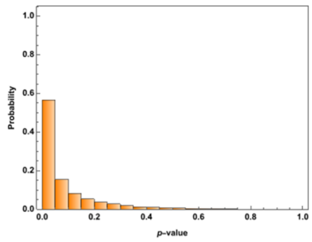
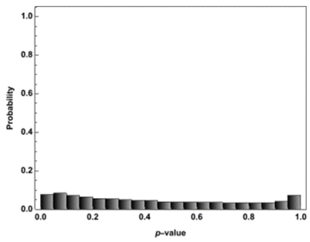
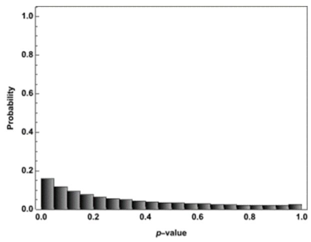
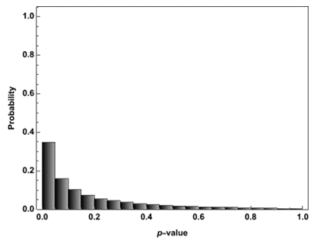
Test	$n = 20$	$n = 50$	$n = 100$
KS			
CvM			
AD			
V			
U^2			
BHEP			

Figure 5.15: Power comparison of EDF and BHEP goodness-of-fit tests for the two-tailed gamma distribution

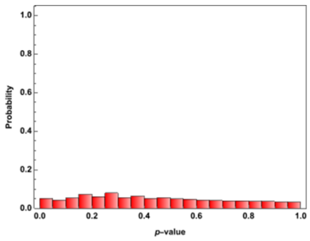
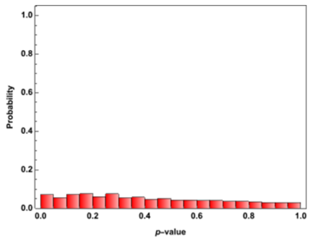
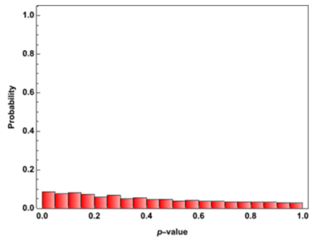
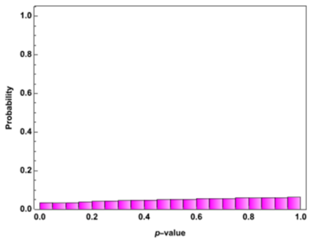
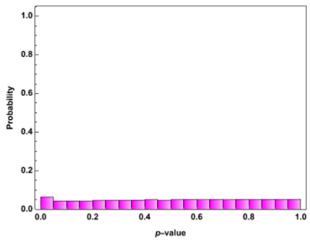
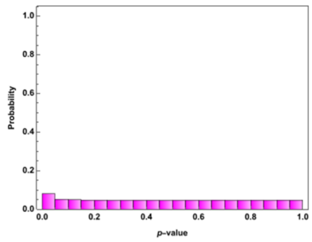
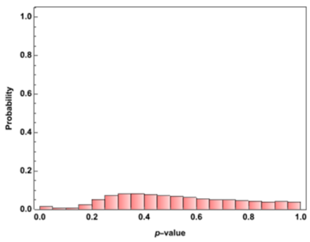
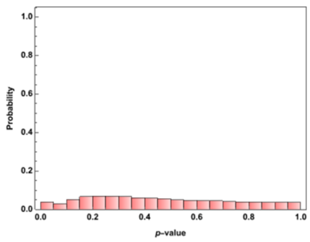
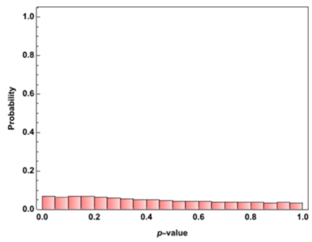
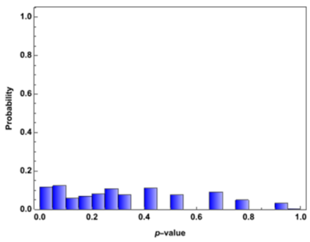
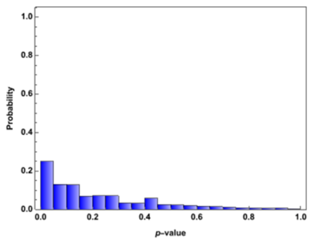
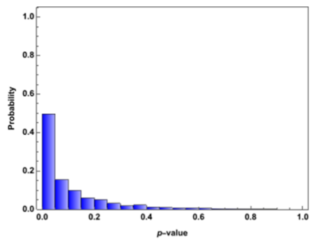
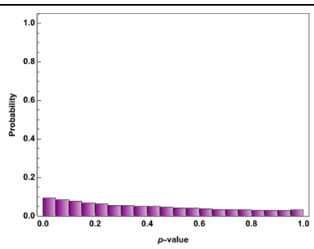
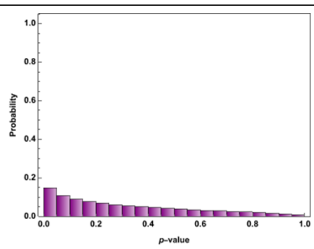
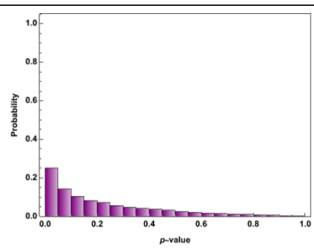
Test	$n = 20$	$n = 50$	$n = 100$
JB			
MS			
MK			
P			
SW			

Figure 5.16: Power comparison of moment-based, Pearson χ^2 and SW goodness-of-fit tests for the two-tailed gamma distribution

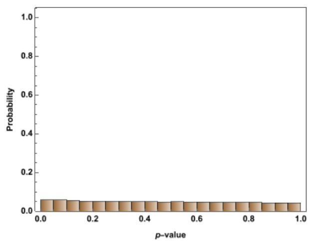
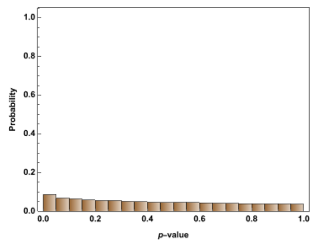
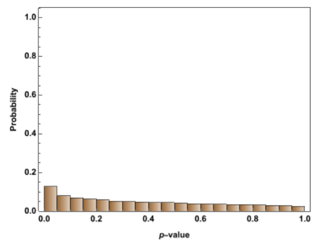
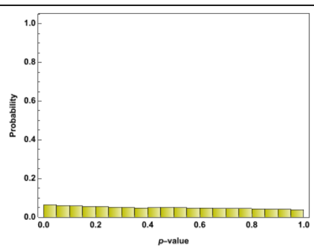
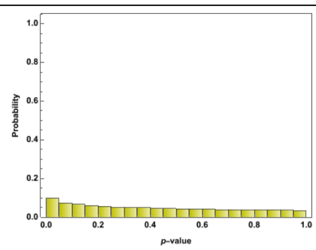
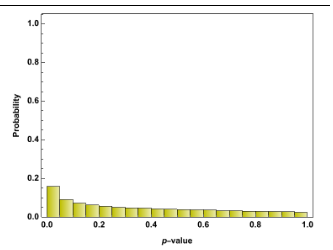
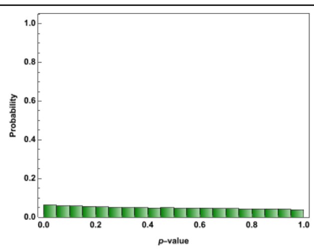
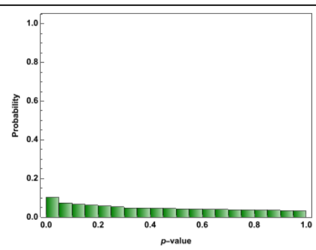
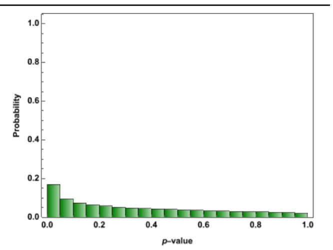
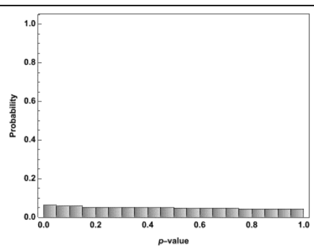
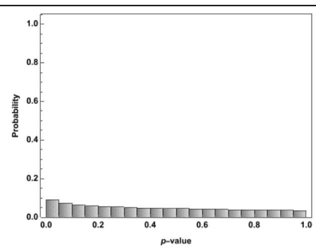
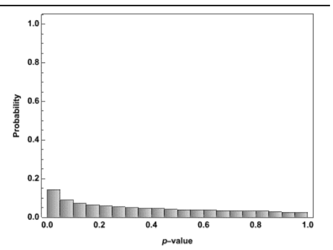
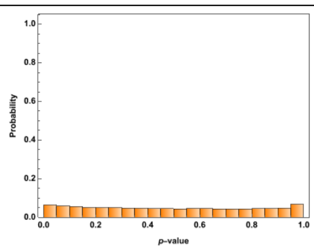
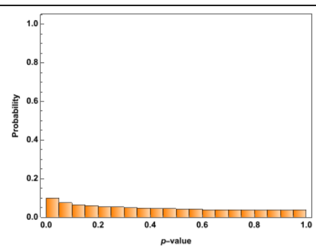
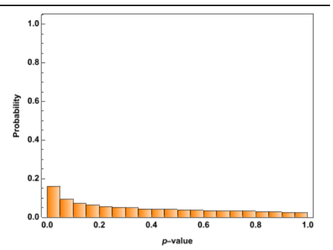
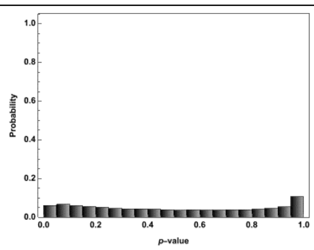
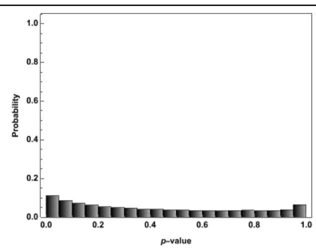
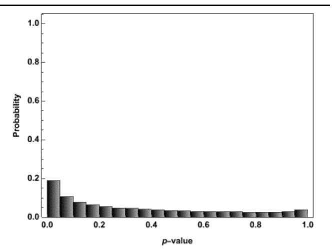
Test	$n = 20$	$n = 50$	$n = 100$
KS			
CvM			
AD			
V			
U^2			
BHEP			

Figure 5.17: Power comparison of EDF and BHEP goodness-of-fit tests for the Burr type III distribution

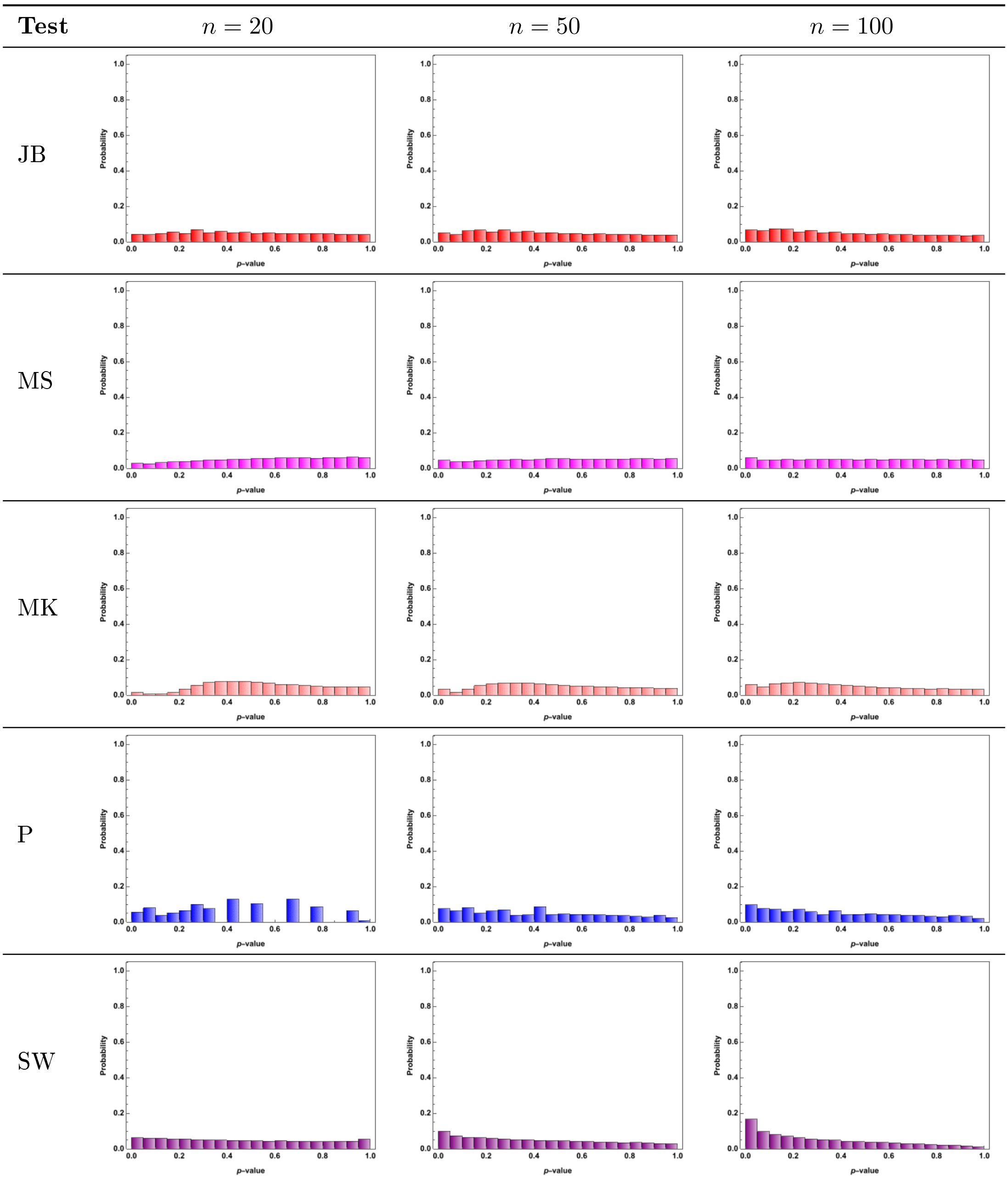


Figure 5.18: Power comparison of moment-based, Pearson χ^2 and SW goodness-of-fit tests for the Burr type III distribution

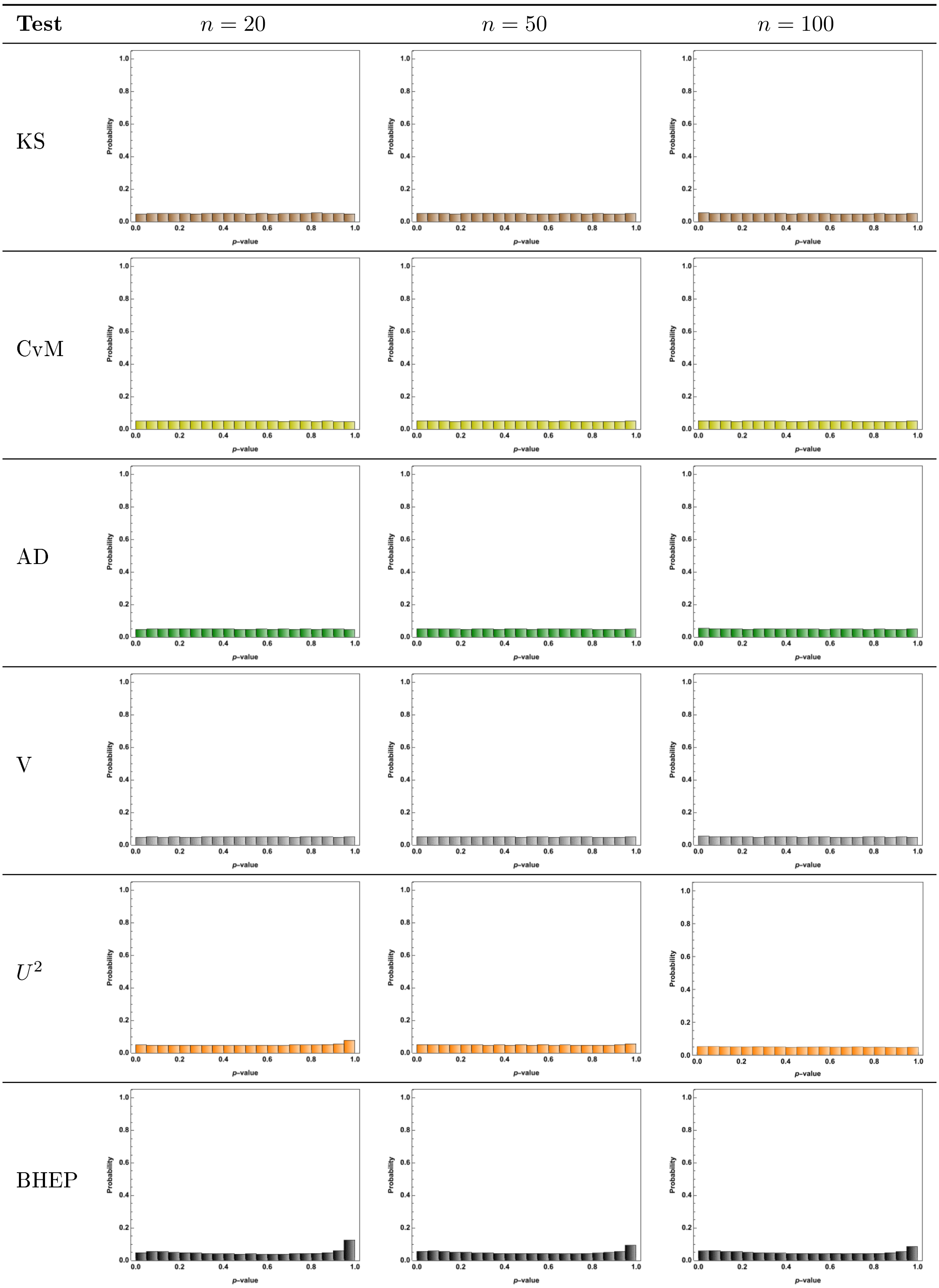


Figure 5.19: Power comparison of EDF and BHEP goodness-of-fit tests for the Burr type XII distribution

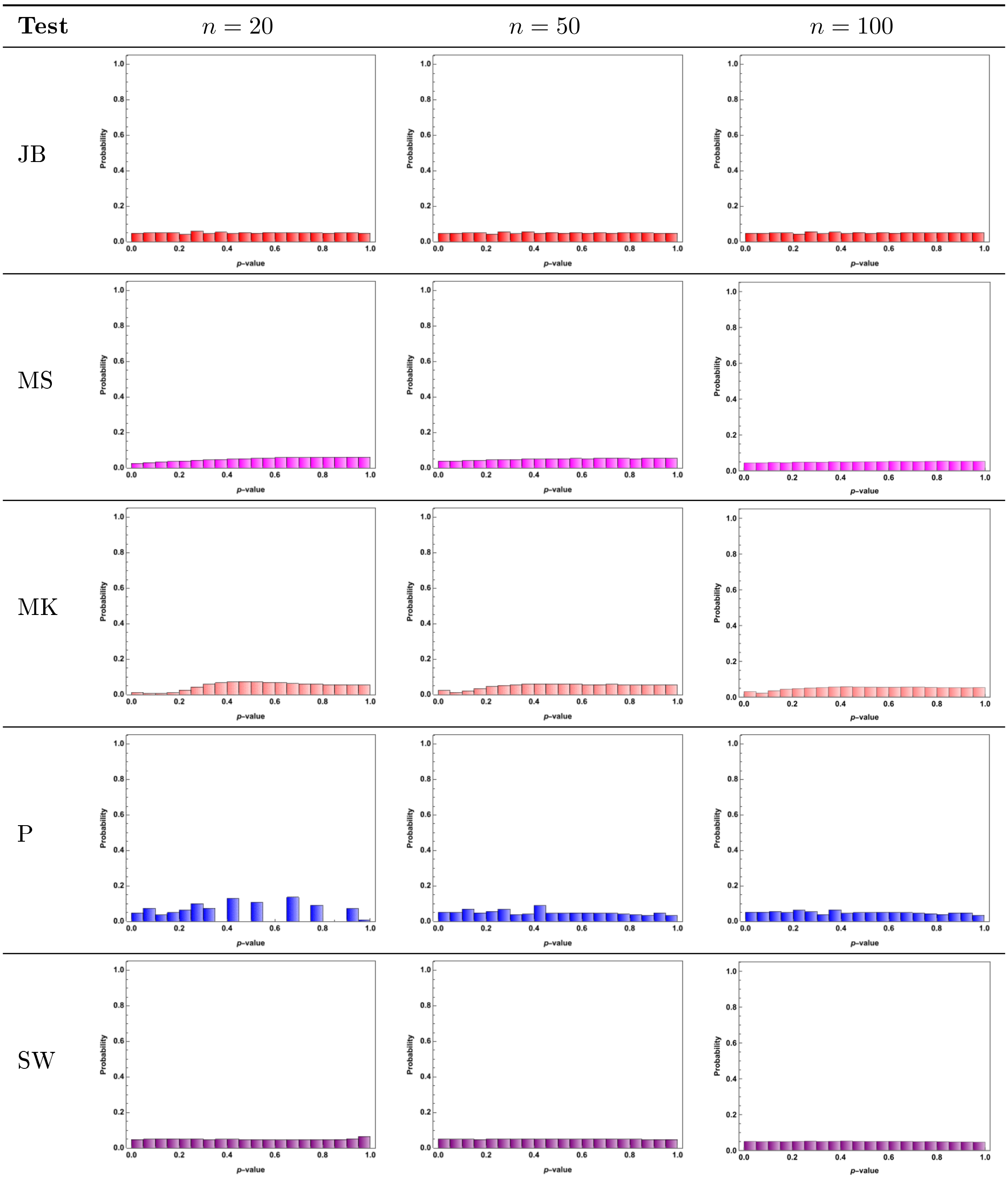


Figure 5.20: Power comparison of moment-based, Pearson χ^2 and SW goodness-of-fit tests for the Burr type XII distribution

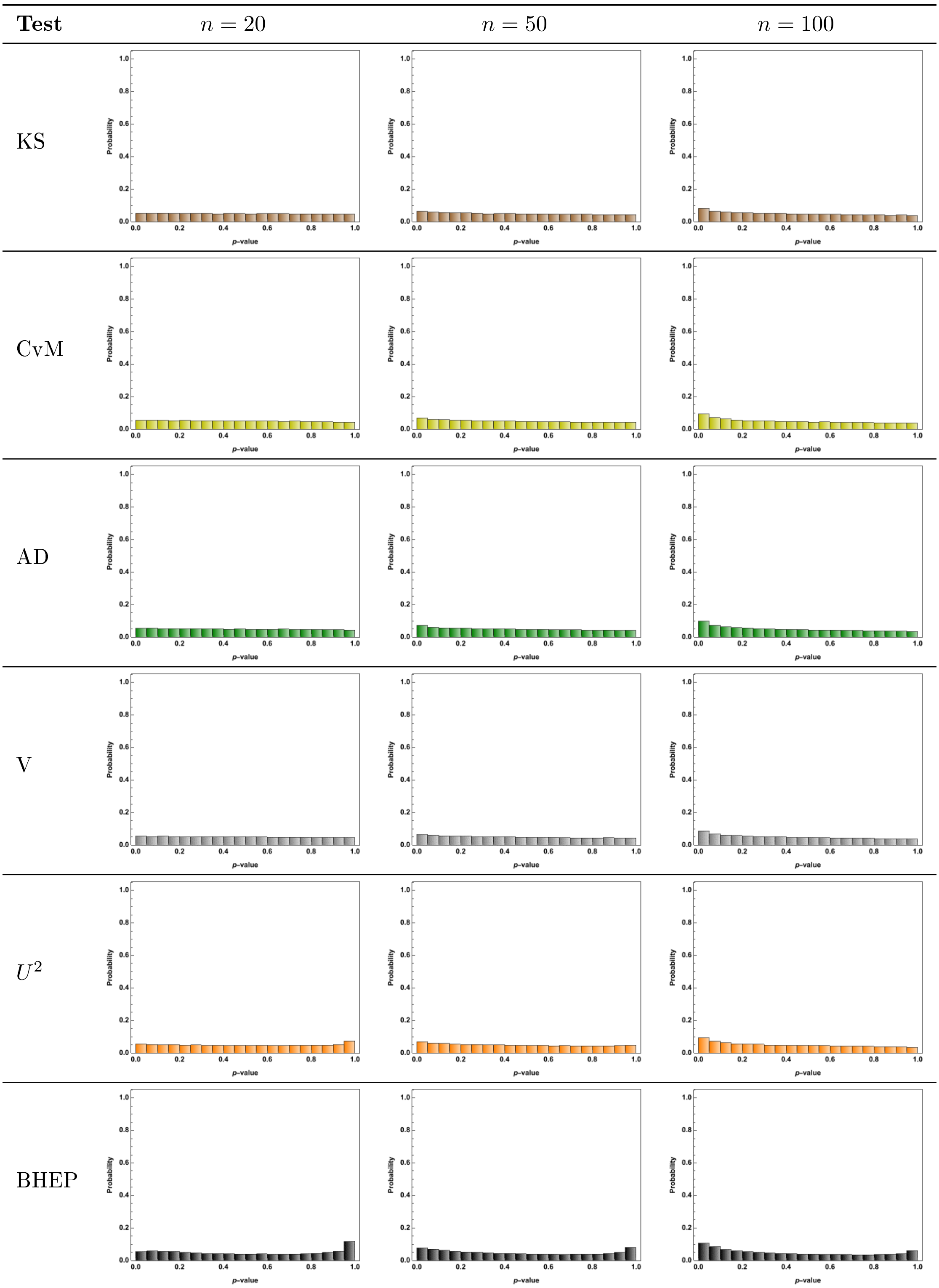


Figure 5.21: Power comparison of EDF and BHEP goodness-of-fit tests for the Davies distribution

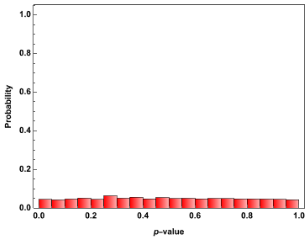
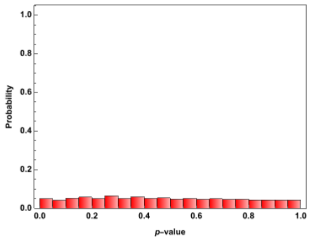
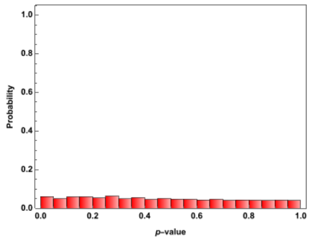
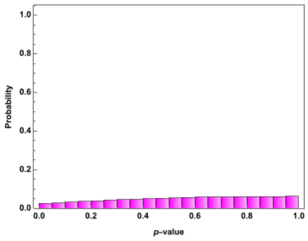
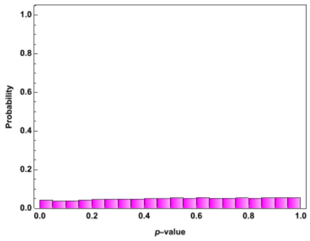
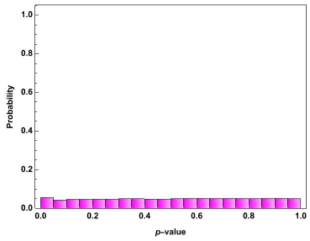
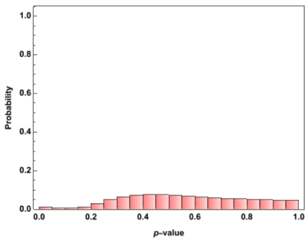
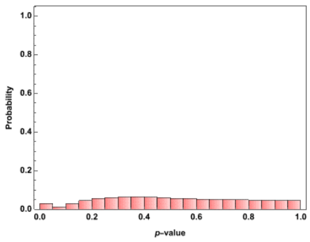
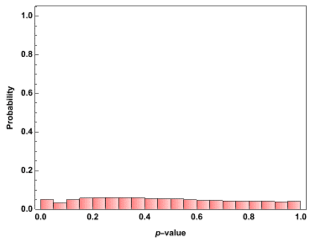
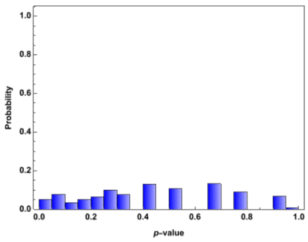
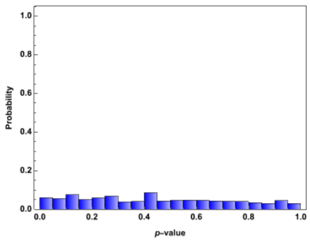
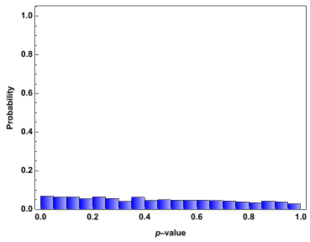
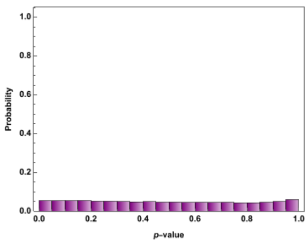
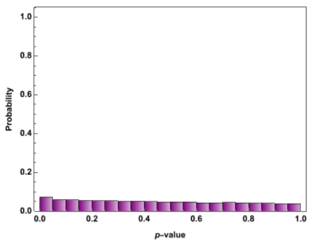
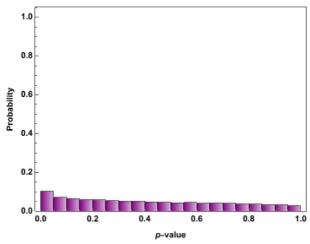
Test	$n = 20$	$n = 50$	$n = 100$
JB			
MS			
MK			
P			
SW			

Figure 5.22: Power comparison of moment-based, Pearson χ^2 and SW goodness-of-fit tests for the Davies distribution

Chapter 6

Simulation study for the b value in the Baringhaus-Henze-Epps-Pulley test

The Baringhaus-Henze-Epps-Pulley (BHEP) test is a multivariate goodness-of-fit test with $d \geq 1$ the number of dimensions. A special case of the BHEP test is where $d = 1$, the univariate case, considered in this study. The BHEP test uses a nonparametric kernel density estimator with bandwidth parameter

$$h = \frac{1}{b\sqrt{2}},$$

where b is the smoothing parameter in the calculation of the test statistic (Henze and Zirkler, 1990). This normality test therefore depends on the choice of this smoothing parameter which is investigated in this chapter.

In a study by Tenreiro (2009) on the choice of the smoothing parameter for the BHEP test, he stated that for the univariate case, the test with the bandwidth $h = 0.49$ or $h = 0.71$ is an omnibus normality test, that is for $b = 1.443$ and $b = 0.996$. I.e., the test is able to detect departures from normality for a wide range of alternative distributions. If $h = 0.92$ ($b = 0.769$) the BHEP test performs well for leptokurtic distributions, while $h = 0.35$ ($b = 2.020$) is preferred for platykurtic distributions. Tenreiro (2009) also mentioned that it has become standard procedure to simply use a bandwidth of $h = 0.71$ for the univariate case of the test. The conclusions of Tenreiro (2009) confirmed the results of Henze and Wagner (1997) who also found that the BHEP test is more powerful with a larger b value for platykurtic distributions and a smaller b value for leptokurtic distributions. Their study also indicated that under a certain range of smoothing parameter values, the power of the BHEP test does not increase with an increase in the sample size.

Wolfram Mathematica has a default setting for the BHEP test that uses the value of b calculated by the expression in Equation 3.4 as suggested by Henze and Zirkler (1990). Wolfram Mathematica allows the user to customise the value of the smoothing parameter b. Therefore a simulation study was conducted considering a range of b values for the various distributions to investigate the effect that the smoothing parameter has on the test.

6.1 Power for different b values

The focus of the simulation study for the power of the BHEP test with different values of b is on $n = 50$ using a 5% significance level. For $n = 50$ and $d = 1$, the default value for the smoothing parameter from Equation 3.4 used by Wolfram Mathematica is $b = 1.45979$. To show the effect of the b value on the test's power, we considered b equal to 0.25, 0.5, 0.75, $\cdots$, 2. These values were chosen based on the study by Tenreiro (2009). Table 6.1 summarizes the power comparison for the different b values chosen for the BHEP test.

Table 6.1: Power comparison for different b values for the BHEP test with $n = 50$ and a 5% significance level

Distributions	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**1.75**	**2**
Normal	6.054	5.874	5.670	5.600	5.484	5.516	5.472	5.398
Generalized secant hyperbolic pi	7.038	6.406	5.952	5.910	6.252	6.410	6.454	6.420
Quantile-based flattened logistic	6.854	6.330	5.868	5.788	5.988	6.104	6.086	5.988
Tukey lambda with ($\lambda = 0.13491$)	5.874	5.790	5.666	5.592	5.570	5.526	5.496	5.414
Tukey lambda ($\lambda = 5.2029$)	2.550	4.248	8.190	14.642	22.254	29.222	34.682	38.778
Generalized Tukey lambda	14.762	22.012	38.208	51.494	60.826	66.562	69.724	71.546
Schmeiser-Deutsch	2.150	4.212	9.932	20.476	33.152	45.478	55.384	62.490
Two-tailed gamma	8.296	7.582	7.528	9.874	13.032	16.412	19.882	22.764
Burr Type III	6.454	7.148	8.520	9.986	10.870	11.214	11.270	11.128
Burr Type XII	6.034	5.944	5.820	5.710	5.676	5.662	5.566	5.502
Davies	6.298	6.546	6.868	7.342	7.722	7.810	7.742	7.606

6.2 Empirical type I errors for different b values

Figure 6.6 represents the probability histograms of the p-values for the normal distribution for various values of b for the BHEP test. Figure 6.1 is a visualization of the empirical type 1 errors from Table 6.1. For all values of b the BHEP test is liberal, but more so for smaller values of b. Because there is very little change in the empirical type I errors for $b \geq 1.25$, these values of the smoothing parameter are appropriate under the null hypothesis of normality.

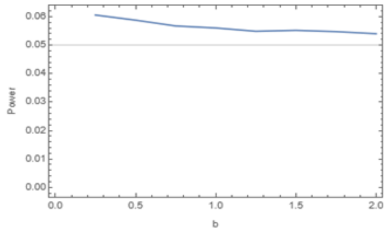


Figure 6.1: Plot of the empirical type I errors for the BHEP test for different b values for the normal distribution with a 5% reference line

6.3 Power against Group 1 distributions for different b values

Recall that the distributions in Group 1 all provide excellent approximations to the normal distribution, in particular the Tukey lambda distribution with $\lambda = 0.13491$. Therefore the power values against these distributions are similar to the empirical type I errors for the standard normal distribution for all values of b as can be seen in Table 6.1 and by comparing Figures 6.1 with 6.2. Also see the probability histograms in Figures 6.7 to 6.9.

For all three distributions in Group 1 the power is always more than 5% for all values of b and in the case of the generalized secant hyperbolic pi and the quantile-based flattened logistic distributions more than the empirical type I errors. For the Tukey lambda distribution with $\lambda = 0.13491$ the power is lower than the empirical type I errors for $0.25 \leq b \leq 1$ and higher than the empirical type I errors for $1.25 \leq b \leq 2$.

(a) Generalized secant hyperbolic pi (b) Quantile-based flattened logistic (c) Tukey lambda with $\lambda = 0.13491$

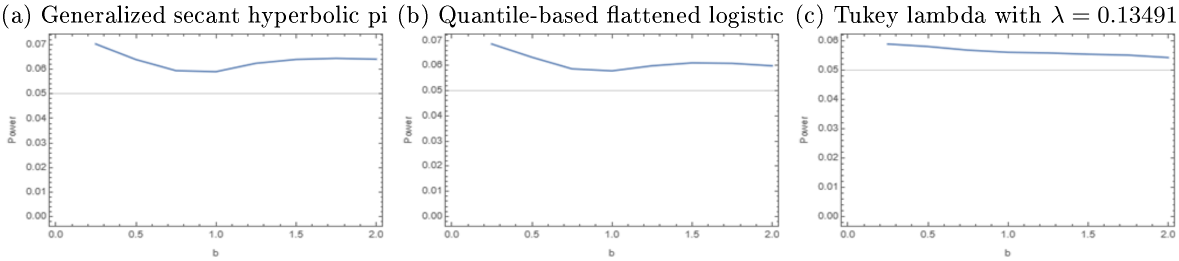


Figure 6.2: Plot of the power of the BHEP test for different b values for the Group 1 distributions with a 5% reference line

6.4 Power against Group 2 distributions for different b values

The distributions in Group 2 are all truncated with bounded support. In contrast to the distributions in Group 1 the power of the BHEP test against the alternatives in Group 2 is monotonically increasing for an increase in the value of the smoothing parameter - see Table 6.1 and Figure 6.3. Hence, the power is highest for large values of b. Figures 6.10 to 6.12 confirm that larger values of b should be used when the alternative distribution is truncated with bounded support. It follows that the choice of the value of b for mesokurtic distributions with bounded support is the same as for platykurtic distributions with bounded support as concluded by Henze and Wagner (1997) and Tenreiro (2009).

From Table 6.1 and Figure 6.3 (a) and (c) we note that for the Tukey lambda distribution with $\lambda = 5.2029$ and for the Schmeiser-Deutsch distribution the power of the BHEP test is less than 5% when $b \leq 0.5$, that is when b is small. Furthermore, Figure 6.10 (a) and (b) and Figure 6.12 (a) and (b) show that the probability histograms of the p-values exhibit an increasing pattern instead of a decreasing pattern. This is phenomenon is not observed for the generalized Tukey lambda distribution, presumably because unlike the other distributions in Group 2 this distribution is asymmetric.

(a) Tukey lambda with $\lambda = 5.2029$ (b) Generalized Tukey lambda (c) Schmeiser-Deutsch

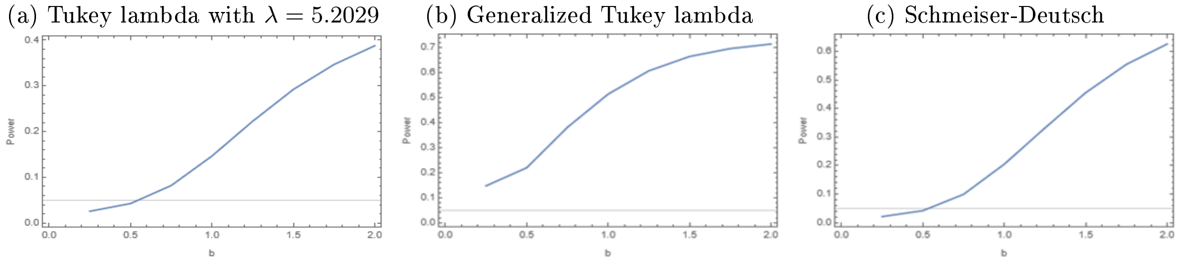


Figure 6.3: Plot of the power of the BHEP test for different b values for the Group 2 distributions with a 5% reference line

6.5 Power against Group 3 distributions for different b values

Figure 6.4 and also Table 6.1 shows that the power of the BHEP test against the two-tailed gamma distribution decreases when b increases from 0.25 to 0.75 and then increases when b increases from 0.75 until 2. Thus, as can also be seen from the probability histograms in Figure 6.13, larger values of b is preferred for the two-tailed gamma distribution.

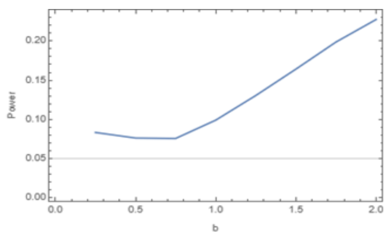


Figure 6.4: Plot of the power of the BHEP test for different b values for the Group 3 distributions with a 5% reference line

6.6 Power against Group 4 distributions for different b values

The probability histograms in Figures 6.14, 6.15 and 6.16 for the distributions in Group 4 look very similar for the respective values of b. However, although all the distributions in Group 4 are asymmetric with half-infinite support, Figure 6.5 and Table 6.1 indicate that the behaviour of the BHEP test in terms of the power for the Burr type III and the Davies distributions is equivalent. The power values for the Burr type XII distribution are more similar to the power values of the Group 1 alternatives, specifically the Tukey lambda distribution with $\lambda = 0.13491$ - compare Figure 6.2 (c) for the Tukey lambda distribution with $\lambda = 0.13491$ with Figure 6.5 (b) for the Burr type XII distribution. Both these distributions are excellent approximations to the normal distribution - see again the probability density curves in Figure 4.4 and Figure 4.9.

(a) Burr type III (b) Burr type XII (c) Davies

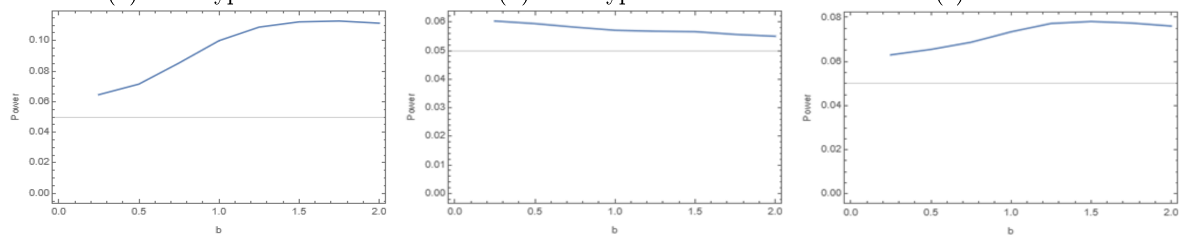


Figure 6.5: Plot of the power of the BHEP test for different b values for the Group 4 distributions with a 5% reference line

6.7 Appendix

In this appendix the distributions of the simulated p-values for the BHEP test with different values of b and for various mesokurtic distributions are plotted using probability histograms.

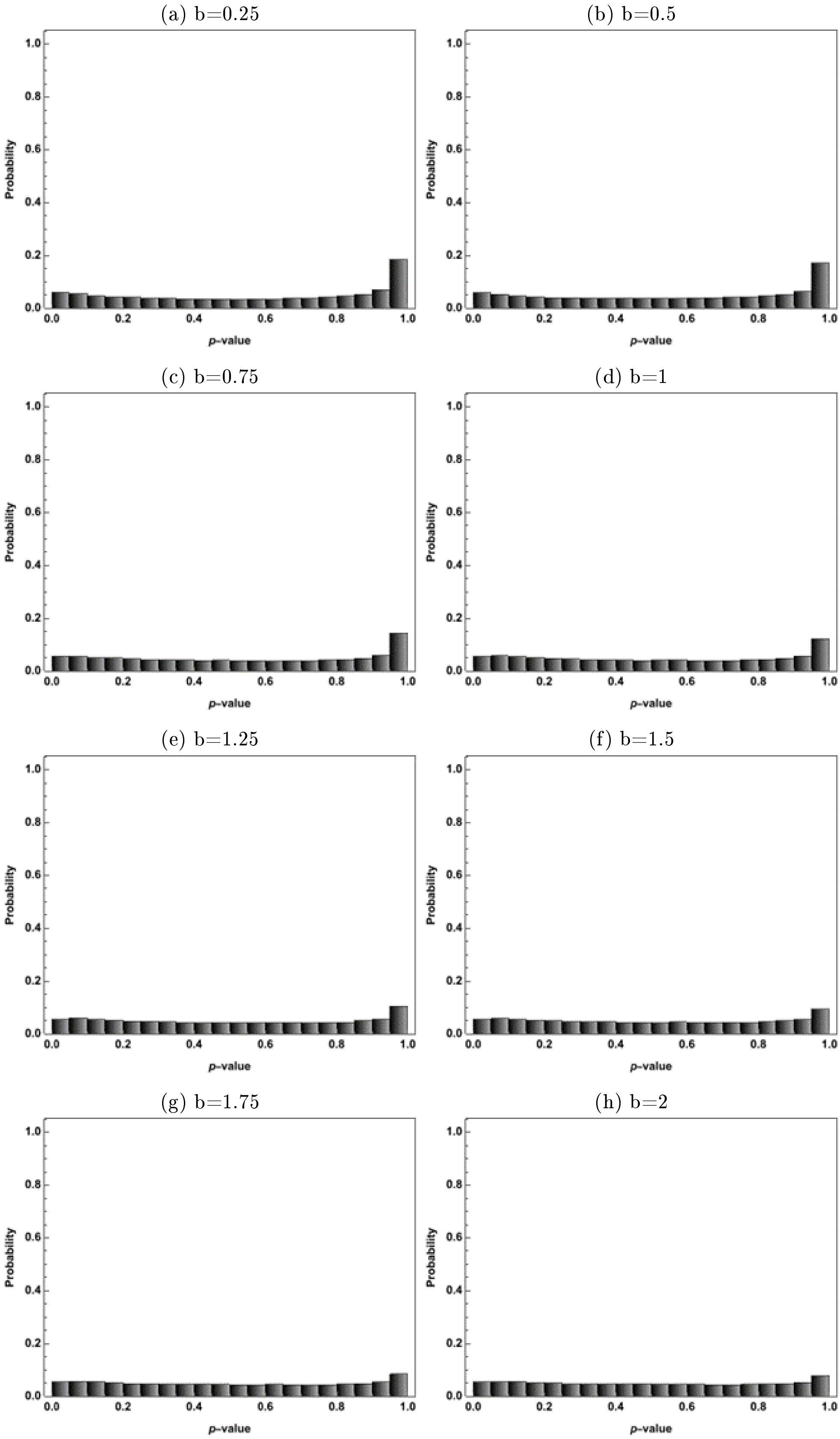


Figure 6.6: Distribution of the p-values for the BHEP test b values for the normal distribution

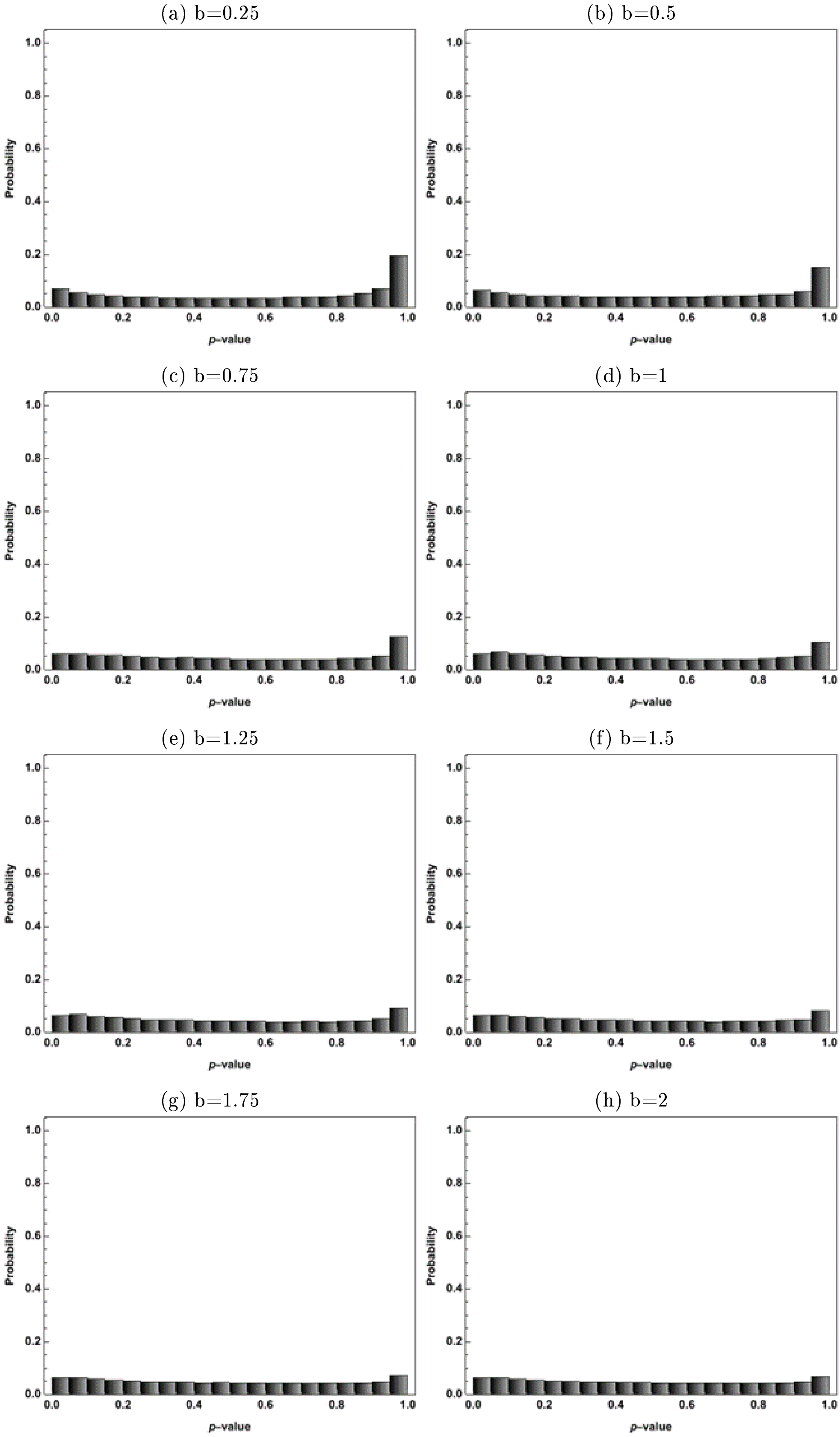


Figure 6.7: Distribution of the p-values for the BHEP test b values for the generalized secant hyperbolic pi distribution

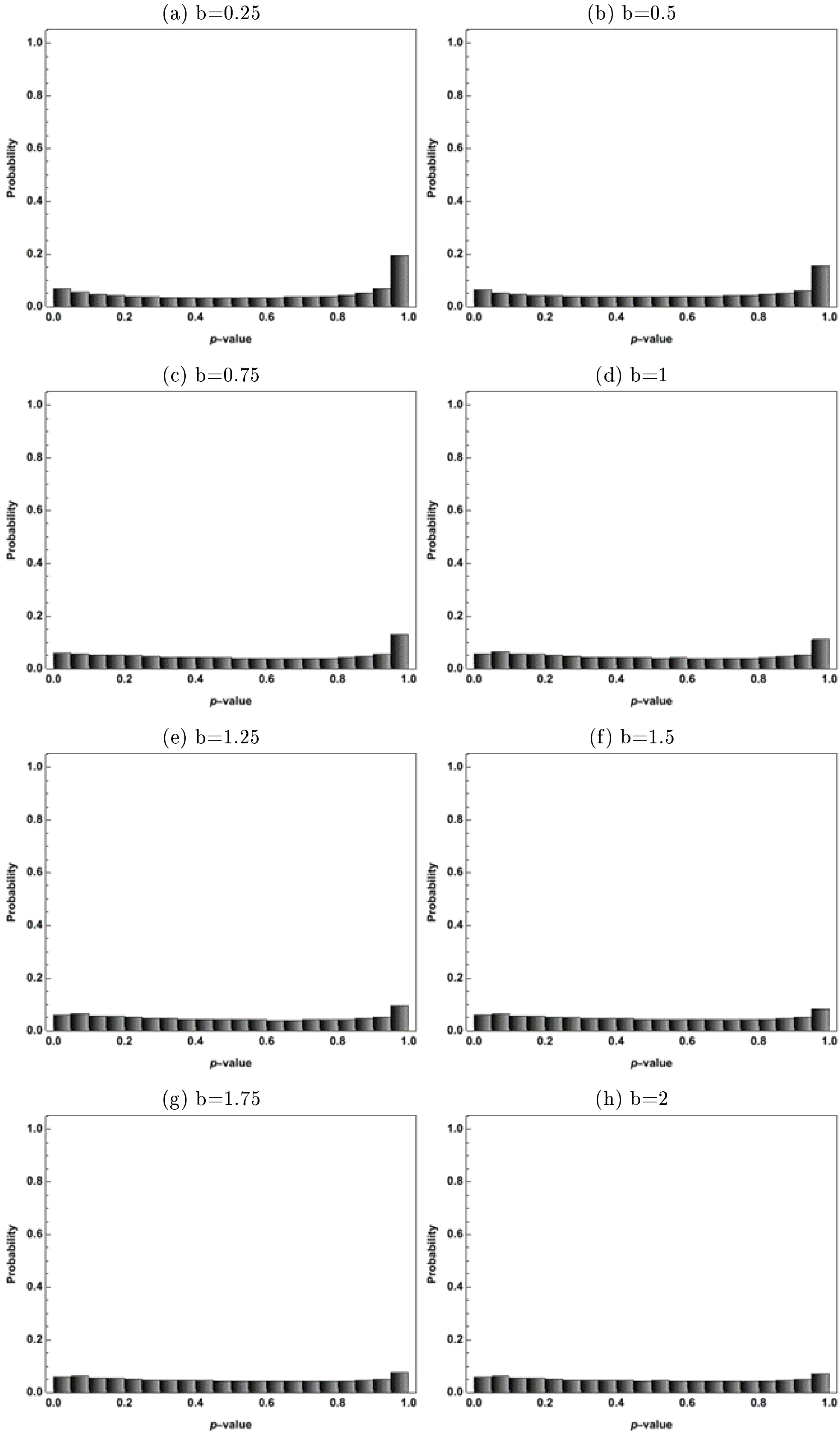


Figure 6.8: Distribution of the p-values for the BHEP test b values for the quantile-based flattened logistic distribution

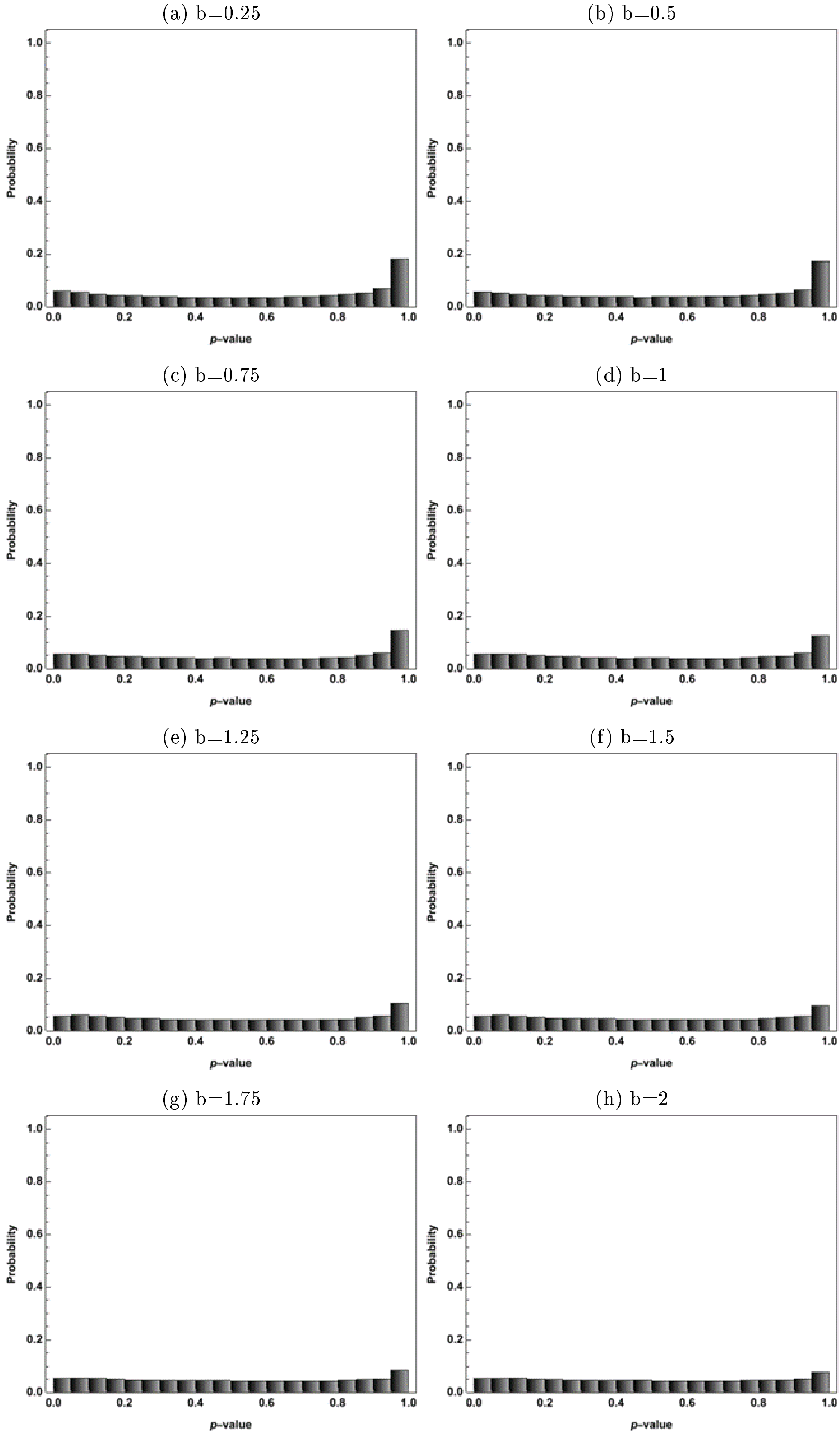


Figure 6.9: Distribution of the p-values for the BHEP test b values for the Tukey lambda ($\lambda = 0.13491$) distribution

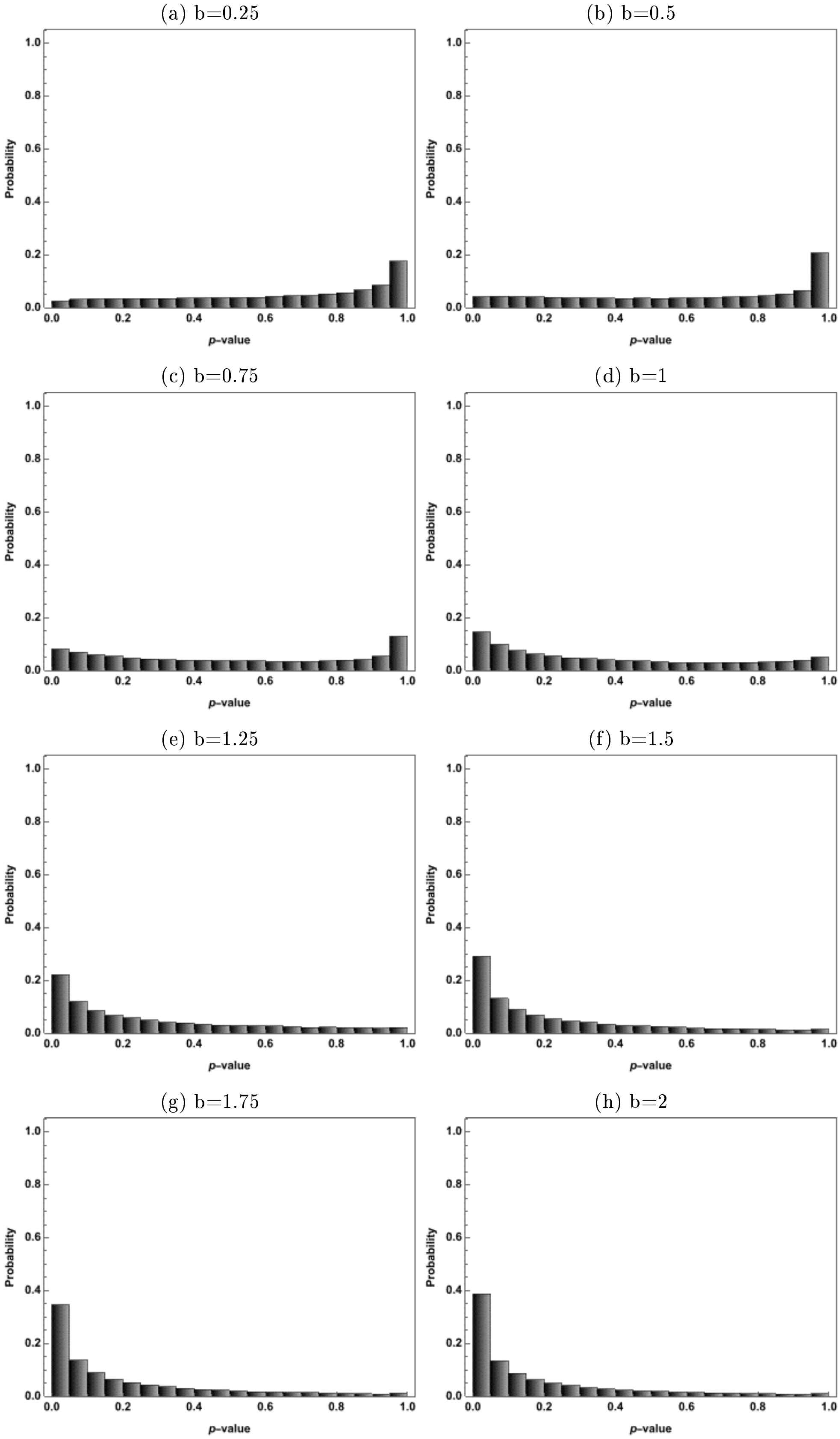


Figure 6.10: Distribution of the p-values for the BHEP test b values for the Tukey lambda ($\lambda = 5.2029$) distribution

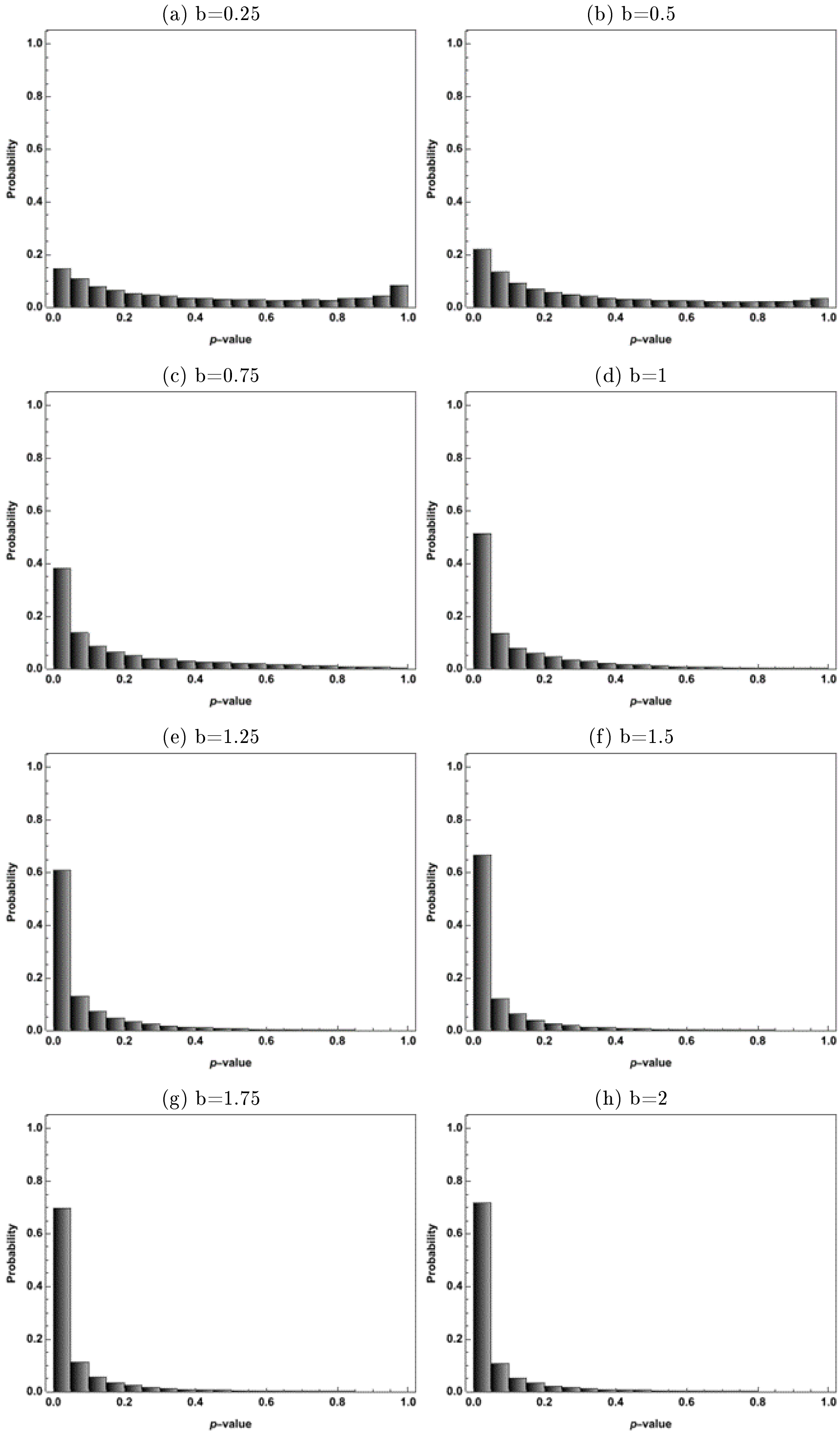


Figure 6.11: Distribution of the p-values for the BHEP test b values for the generalized Tukey distribution

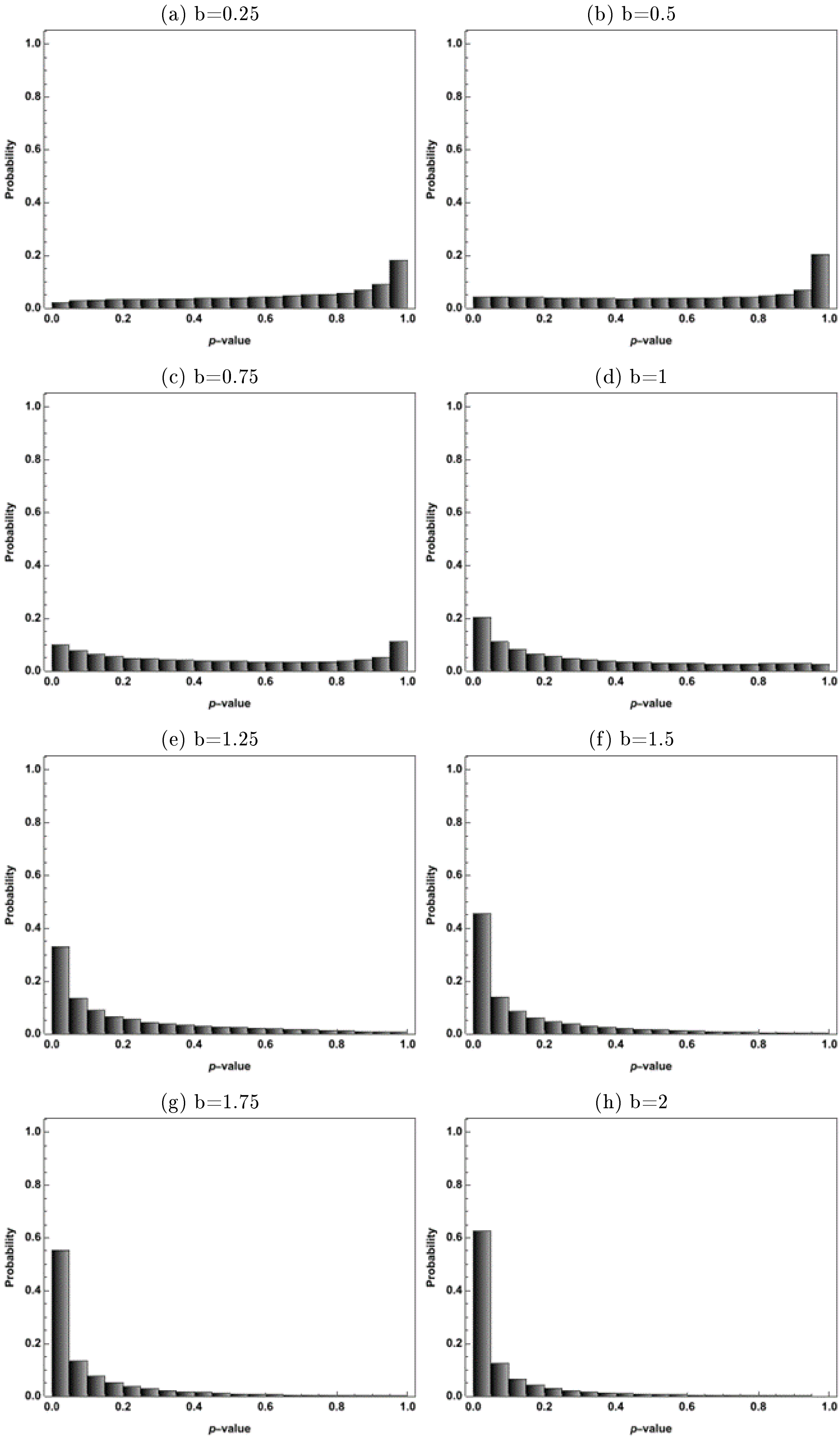


Figure 6.12: Distribution of the p-values for the BHEP test b values for the Schmeiser-Deutsch distribution

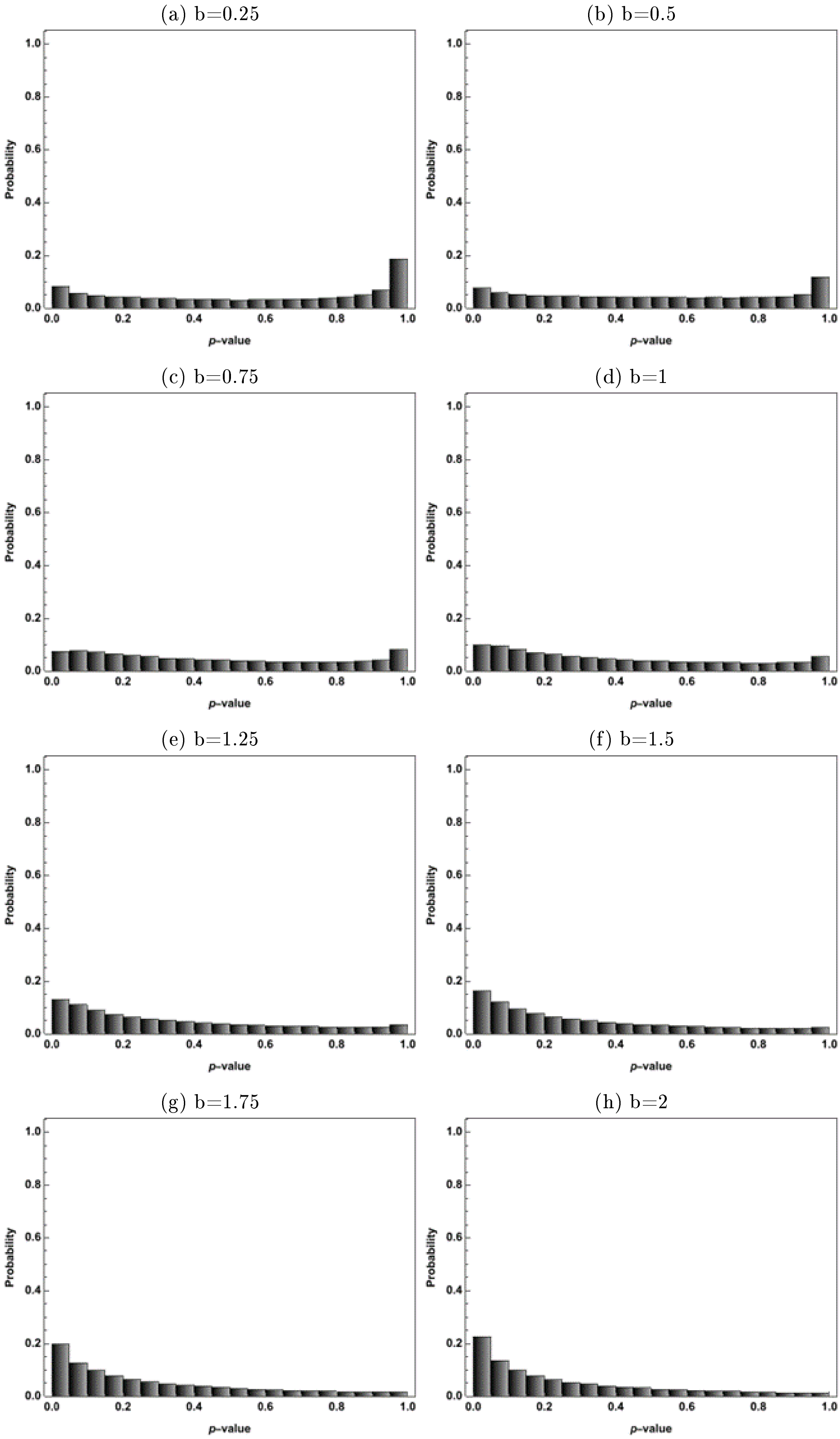


Figure 6.13: Distribution of the p-values for the BHEP test b values for the two-tailed gamma distribution

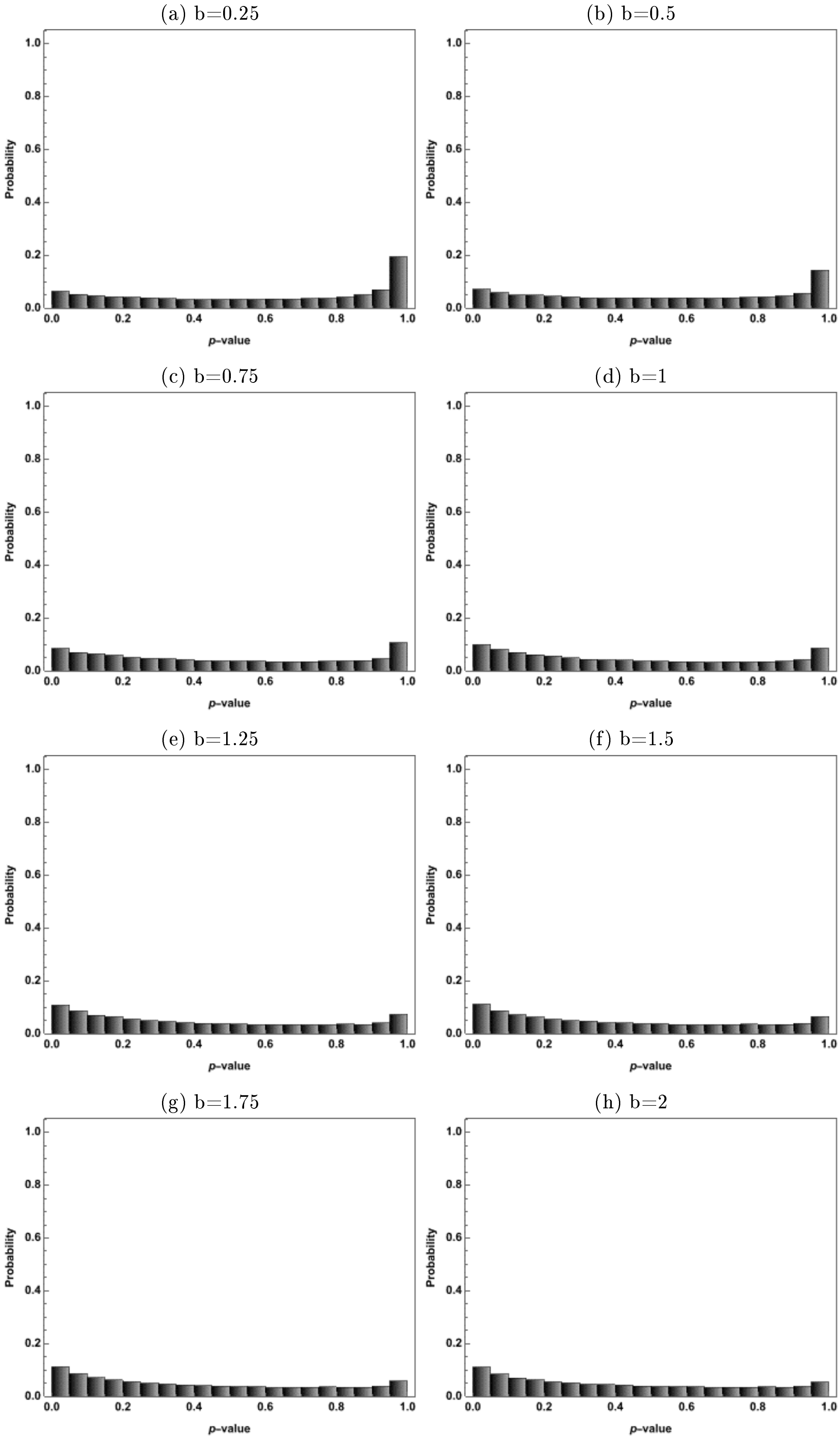


Figure 6.14: Distribution of the p-values for the BHEP test b values for the Burr type III distribution

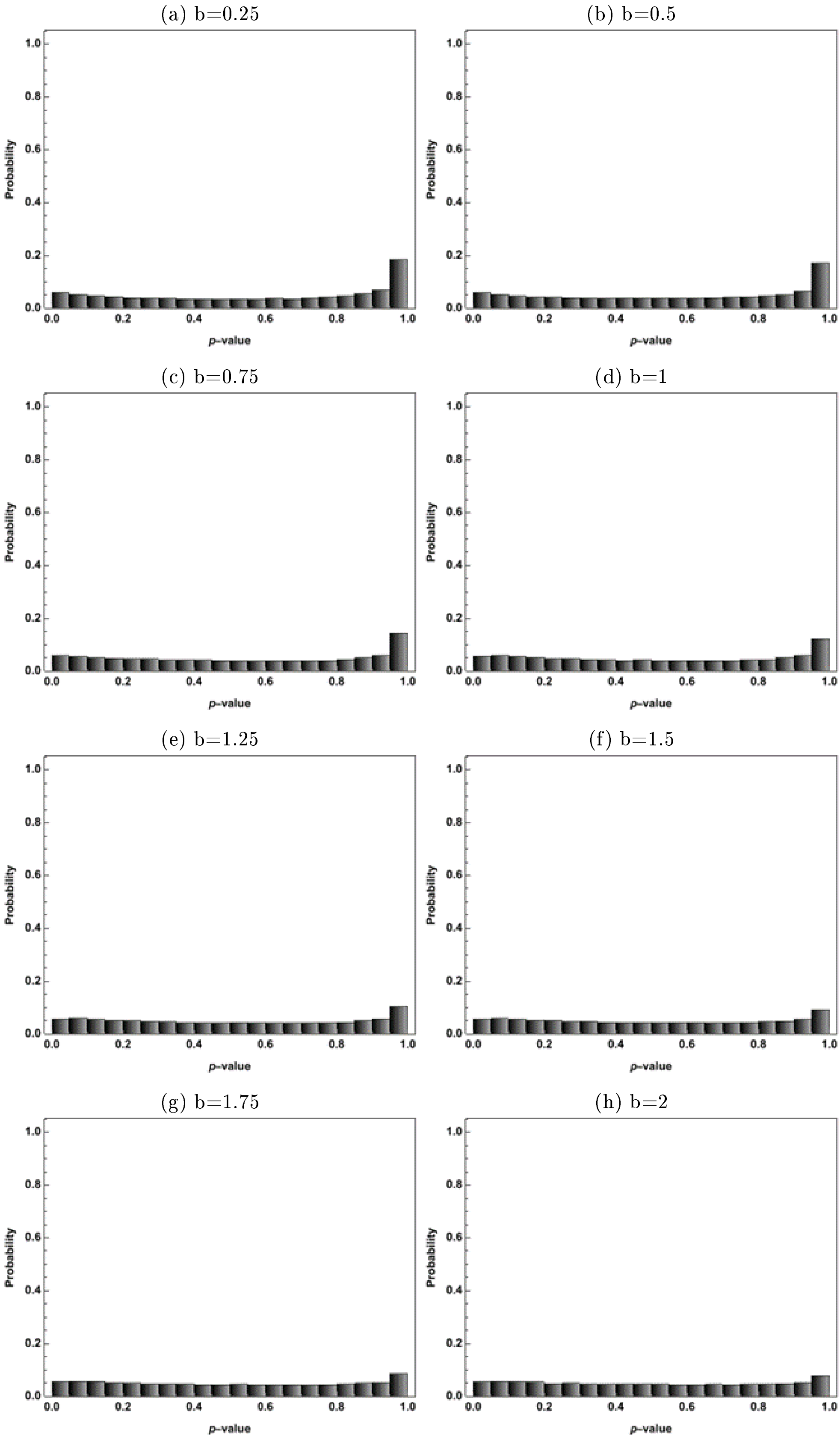


Figure 6.15: Distribution of the p-values for the BHEP test b values for the Burr type XII distribution

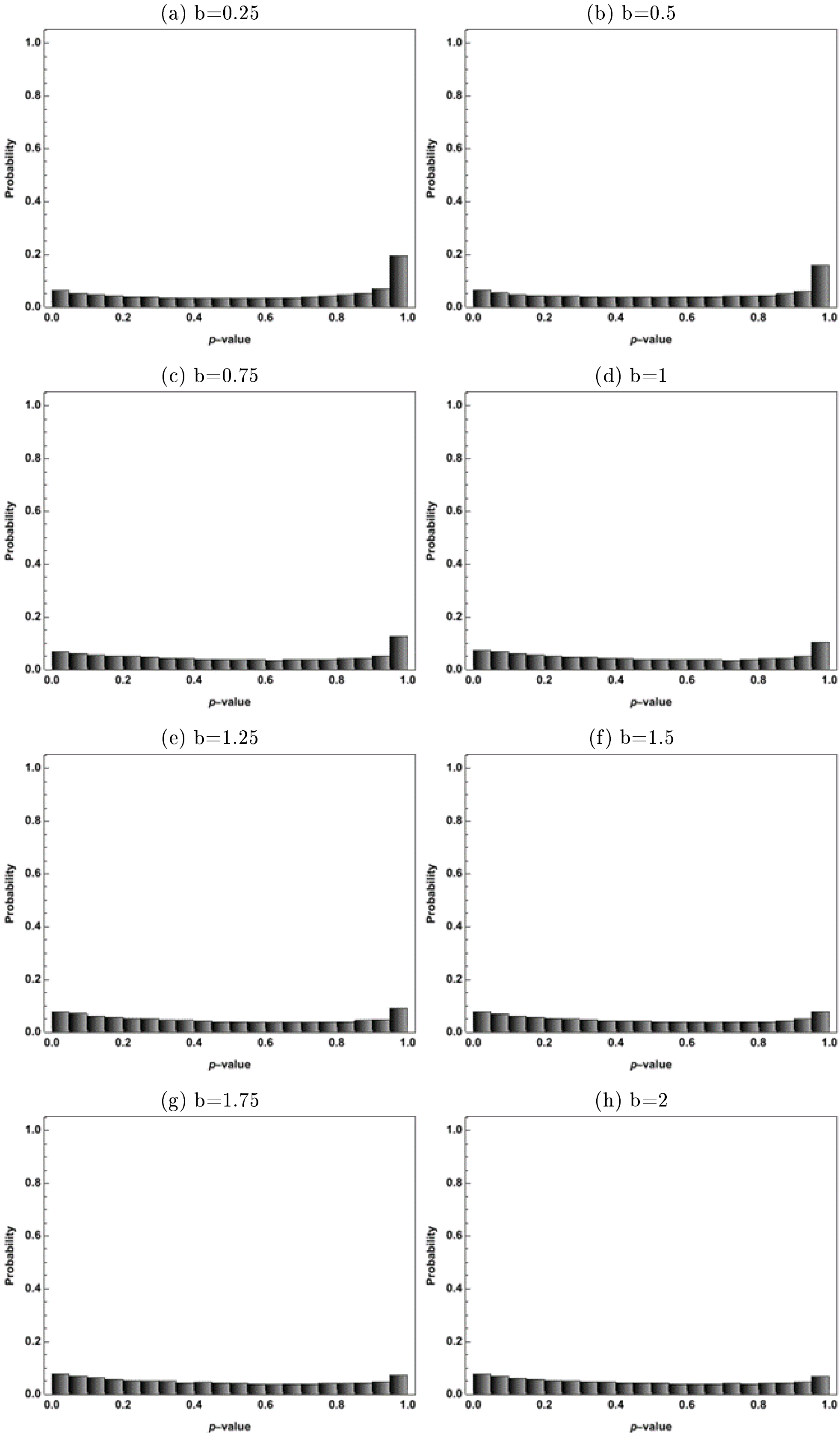


Figure 6.16: Distribution of the p-values for the BHEP test b values for the Davies distribution

Chapter 7

Conclusion

The topic of testing for normality was introduced using an example on 100-km ultra-marathon race times where the data was found to be approximately mesokurtic, that is the coefficients of skewness and kurtosis are $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$, respectively. In this example, the goodness-of-fit tests did not all reject normality even though the data is clearly bimodal. In light of this, an investigation into goodness-of-fit testing was conducted.

In order to fully understand normality testing, a discussion of each goodness-of-fit test considered in this mini-dissertation was done in Chapter 3. This was followed by a study of the distributions included in this mini-dissertation in Chapter 4. Even though all the distributions included were mesokurtic distributions, not all the distributions have the same distributional shape and properties as the standard normal distribution.

The simulation study in Chapter 5 concluded that most of the goodness-of-fit tests do not reject normality when a distribution is mesokurtic and has very similar distributional shape and properties to the standard normal distribution. For symmetric, truncated distributions (Tukey lambda distribution with $\lambda = 5.2029$ and Schmeiser-Deutsch distribution) tests based on empirical distribution or characteristic functions perform the best. The Shapiro-Wilk test has the highest power for the asymmetric, truncated generalized Tukey lambda distribution considered. For asymmetric distributions on half-infinite support the Baringhaus-Henze-Epps-Pulley test performs the best. Tests based on the empirical distribution function have the highest power for the two-tailed gamma distribution. Not surprisingly, moment-based tests performed by far the worst for all mesokurtic distributions considered in this study. These tests should be used for detecting mesokurtosis and not as goodness-of-fit tests for normality.

Finally, a simulation study on the effect of b in the BHEP test was conducted considering different b values. This test does not perform well with a small smoothing parameter value and a larger b value is preferred when considering mesokurtic distributions.

Bibliography

Anastasiou, A., Karagrigoriou, A., and Katsileros, A. (2020). Comparative evaluation of goodness-of-fit tests for normal distribution using simulation and empirical data. *Biometrical Letters*, 57(2):237–251.

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2):193–212.

Anderson, T. W. and Darling, D. A. (1954). A test of goodness-of-fit. *Journal of the American Statistical Association*, 49(268):765–769.

Baringhaus, L. and Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika*, 35(1):339–348.

Baten, W. (1934). The probability law for the sum of n independent variables, each subject to the law $(1/(2h))$ sech $(\pi x/(2h))$. *Bulletin of the American Mathematical Society*, 40(4):284–290.

Burr, I. W. (1942). Cumulative frequency functions. *The Annals of Mathematical Statistics*, 13(2):215–232.

Cramér, H. (1928). On the composition of elementary errors. *Scandinavian Actuarial Journal*, 11:13–74.

D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-Fit Techniques.* Marcel Dekker, Inc.

De Moivre, A. (1733). Approximatio ad summam terminorum binomii. *Self published pamphlet*, 7 pages.

Domański, C. and Szczepocki, P. (2020). Comparison of selected tests for univariate normality based on measures of moments. *Statistics in Transition New Series*, 21(5):151–178.

Doornik, J. A. and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70(s1):927–939.

Epps, T. W. and Pulley, L. B. (1983). A test for normality based on the empirical characteristic function. *Biometrika*, 70(3):723–726.

Freimer, M., Kollia, G., Mudholkar, G. S., and Lin, C. T. (1988). A study of the generalized Tukey lambda family. *Communications in Statistics-Theory and Methods*, 17(10):3547–3567.

Gan, F. F. and Koehler, K. J. (1990). Goodness-of-fit tests based on P-P probability plots. *Technometrics*, 32(3):289–303.

Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium.* Friedrich Perthes & Johann Heinrich Besser, Hamburg.

Gel, Y. R. and Gastwirth, J. L. (2008). A robust modification of the Jarque-Bera test of normality. *Economics Letters*, 99(1):30–32.

Gilchrist, W. (2000). *Statistical Modelling with Quantile Functions.* Chapman and Hall/CRC.

Hankin, R. K. and Lee, A. (2006). A new family of non-negative distributions. *Australian & New Zealand Journal of Statistics*, 48(1):67–78.

Henze, N. and Wagner, T. (1997). A new approach to the BHEP tests for multivariate normality. *Journal of Multivariate Analysis*, 62(1):1–23.

Henze, N. and Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods*, 19(10):3595–3617.

Hildebrand, D. K. (1971). Kurtosis measures bimodality? *The American Statistician*, 25(1):42–43.

Jarque, C. M. and Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2):163–172.

Johnson, N. and Tadikamalla, P. (1992). Translated families of distributions, Chapter 8 in Handbook of the Logistic Distribution, Balakrishnan, N ed.

Joiner, B. L. and Rosenblatt, J. R. (1971). Some properties of the range in samples from Tukey's symmetric lambda distributions. *Journal of the American Statistical Association*, 66(334):394–399.

Karian, Z. A. and Dudewicz, E. J. (2010). *Handbook of Fitting Statistical Distributions with R.* CRC press.

King, R. A. R. (1999). *New distributional fitting methods applied to the generalised λ distribution.* PhD thesis, Queensland University of Technology.

Klein, I. and Fischer, M. (2007). A note on the kurtosis ordering of the generalized secant hyperbolic distribution. *Communications in Statistics-Theory and Methods*, 37(1):1–7.

Knechtle, B., Nikolaidis, P. T., and Valeri, F. (2018). Russians are the fastest 100-km ultra-marathoners in the world. *PLoS One*, 13(7):e0199701.

Kollo, T. (2008). Multivariate skewness and kurtosis measures with an application in ICA. *Journal of Multivariate Analysis*, 99(10):2328–2338.

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell' Intituto Italiano degli Attuari*, 4:83–91.

Kuiper, N. H. (1960). Tests concerning random points on a circle. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A*, 63:38–47.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318):399–402.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3):519–530.

Moore, D. S. (1986). Tests of chi-squared type. In *Goodness-of-Fit Techniques*, pages 63–96. Routledge.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175.

Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R., and Mykytka, E. F. (1979). A probability distribution and its uses in fitting data. *Technometrics*, 21(2):201–214.

Ramberg, J. S. and Schmeiser, B. W. (1972). An approximate method for generating symmetric random variables. *Communications of the ACM*, 15(11):987–990.

Ramberg, J. S. and Schmeiser, B. W. (1974). An approximate method for generating asymmetric random variables. *Communications of the ACM*, 17(2):78–82.

Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J., and Mykytka, E. F. (1980). Letters to the editor. *Technometrics*, 22(1):135–135.

Rodriguez, R. N. (1977). A guide to the Burr type XII distributions. *Biometrika*, 64(1):129–134.

Romão, X., Delgado, R., and Costa, A. (2010). An empirical power comparison of univariate goodness-of-fit tests for normality. *Journal of Statistical Computation and Simulation*, 80(5):545–591.

Royston, P. (1992). Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and Computing*, 2(3):117–119.

Schmeiser, B. W. and Deutsch, S. J. (1977). A versatile four parameter family of probability distributions suitable for simulation. *AIIE Transactions*, 9(2):176–182.

Schorr, B. (1974). On the choice of the class intervals in the application of the chi-square test. *Mathematische Operationsforschung und Statistik*, 5(4-5):357–377.

Seier, E. (2002). Comparison of tests for univariate normality. *InterStat Statistical Journal*, 1:1–17.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3 & 4):591–611.

Sharma, D. and Chakrabarty, T. K. (2019). The quantile-based flattened logistic distribution: Some properties and applications. *Communications in Statistics: Theory and Methods*, 48(14):3643–3662.

Smirnov, N. (1939). Sur les écarts de la courbe de distribution empirique:. *Recueil Mathématique*, 6(1):3–26.

Stephens, M. A. (1974). EDF statistics for goodness-of-fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737.

Tadikamalla, P. R. (1980). A look at the Burr and related distributions. *International Statistical Review/Revue Internationale de Statistique*, 48(3):337–344.

Talacko, J. (1956). Perks' distributions and their role in the theory of Wiener's stochastic variables. *Trabajos de Estadistica*, 7(2):159–174.

Tenreiro, C. (2009). On the choice of the smoothing parameter for the BHEP goodness-of-fit test. *Computational Statistics & Data Analysis*, 53(4):1038–1053.

Thadewald, T. and Büning, H. (2007). Jarque-Bera test and its competitors for testing normality- A power comparison. *Journal of Applied Statistics*, 34(1):87–105.

Thas, O. (2010). *Comparing Distributions*, volume 233. Springer.

Tukey, J. W. (1960). The practical relationship between the common transformations of percentages or fractions and of amounts. *Technical Report 36, Statistical Techniques Research Group, Princeton University*, 6:211–219.

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33(1):1–67.

Urzúa, C. M. (1996). On the correct use of omnibus tests for normality. *Economics Letters*, 53(3):247–251.

Uyanto, S. S. (2022). An extensive comparisons of 50 univariate goodness-of-fit tests for normality. *Austrian Journal of Statistics*, 51(3):45–97.

van Staden, P. J. et al. (2013). *Modeling of generalized families of probability distribution in the quantile statistical universe.* PhD thesis, University of Pretoria.

van Staden, P. J. and Loots, M. T. (2009). Teaching the concept of kurtosis in introductory statistics courses using Mathematica: Searching for platypuses and kangaroos beneath the cloth of Table Mountain. *Proceedings of the Seventh Southern Right Delta Conference on the Teaching and Learning of Undergraduate Mathematics and Statistics*, page 244.

Vaughan, D. C. (2002). The generalized secant hyperbolic distribution and its properties. *Communications in Statistics: Theory and Methods*, 31(2):219–238.

von Mises, R. (1931). *Vorlesungen aus dem Gebiete der angewandten Mathematik. 1. Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und Theoretischen Physik.* Franz Deuticke, Vienna, Austria.

Watson, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika*, 48(1 & 2):109–114.

Yap, B. W. and Sim, C. H. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12):2141–2155.

Appendix (Wolfram Mathematica code)

The Wolfram Mathematica code for the Davies distribution can be found at https://figshare.com/s/7a57ae8242e49b47918a, the code for the other distributions is similar to the Davies distribution's code but is not included.