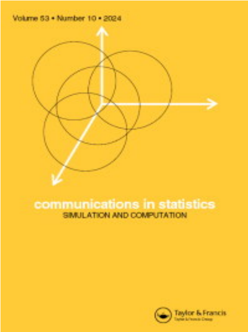

Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/lssp20

Parametric testing for normality against bimodal and unimodal alternatives using higher moments

Geoffrey Poitras

To cite this article: Geoffrey Poitras (2024) Parametric testing for normality against bimodal and unimodal alternatives using higher moments, Communications in Statistics - Simulation and Computation, 53:10, 4771-4789, DOI: 10.1080/03610918.2022.2155313

To link to this article: https://doi.org/10.1080/03610918.2022.2155313

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
2024, VOL. 53, NO. 10, 4771–4789
https://doi.org/10.1080/03610918.2022.2155313

Parametric testing for normality against bimodal and unimodal alternatives using higher moments

Geoffrey Poitras

Beedie School of Business, Simon Fraser University, Vancouver, BC, Canada

ABSTRACT
This study examines population and small sample properties of the standardized fifth and sixth moments – the "higher moments" – for assessing univariate normality against bimodal and selected unimodal alternatives. Population parameters and distributions for selected bimodal mixtures are calculated and contrasted with those for the normal distribution. Using Gram-Charlier series expansion methods, an omnibus goodness of fit test incorporating the higher moments is specified and Monte Carlo simulation used to compare test power with parametric tests based on the standardized third and fourth sample moments: the asymptotic and size corrected versions of the Jarque-Bera score test and the omnibus D'Agostino K^2 test. The studentized range and directional tests using the third through sixth moments are also considered. The results demonstrate that incorporating the fifth and sixth moments can provide enhanced parametric normality test power for bimodal normal mixture alternatives but not for various unimodal alternatives.

ARTICLE HISTORY
Received 6 May 2022
Accepted 27 August 2022

KEYWORDS
Bimodal distribution; Monte Carlo simulation; Normality test; Omnibus test; Score test

1. Introduction

Following a literature review on previous parametric normality testing procedures used to assess the presence of bimodality, standardized population moments for skewness ($\sqrt{\beta_1}$), kurtosis (β_2), and the "higher" fifth ($\sqrt{\beta_3}$) and sixth (β_4) standardized moments for a range of normal mixture bimodal density specifications are provided together with a selection of density plots. Bimodal mixtures of normal with non-central Student t densities are also examined. The associated implications for conventional goodness of fit testing are illustrated by examining the relevance of Hermite polynomials for modeling bimodal densities using Gram-Charlier (Type A) series expansions. This background motivates the investigation of omnibus and directional goodness of fit tests that use estimates of the standardized sample moments ($\sqrt{b_1}$, b_2, $\sqrt{b_3}$ and b_4) to test the null hypothesis of normality against bimodal and selected unimodal alternatives. Results are provided for a Monte Carlo study that compares the power of omnibus and directional tests over a range of bimodal and selected unimodal distributional alternatives. The paper concludes with some speculations about directions for further research.

2. Literature review

The vast literature on testing goodness-of-fit for the null hypothesis of normality against unimodal alternatives provides various useful studies comparing the performance of the sizable

CONTACT Geoffrey Poitras poitras@sfu.ca Beedie School of Business, Simon Fraser University, Vancouver, BC V5A lS6, Canada.

number of methods that have been proposed, e.g., Shapiro, Wilk, and Chen (1968), Bontemps and Meddahi (2005), Romão, Delgado, and Costa (2010), Wijekularathna, Manage, and Scariano (2022), Uhm and Yi (2021). When faced with the possibility that the underlying data generating process is bimodal, the common method used to identify distributional shape is non-parametric kernel density estimation, e.g., Einbeck and Taylor (2013); Schmitt and Westerhoff (2017). Popularized by Silverman (1986) and Hall and York (2001), this sophisticated evolution of histogram analysis is confronted with the problem of appropriate bandwidth specification, a free parameter problem that can substantively impact the power, size and practical complexity of testing for fitted densities. Alternatively, when there is a theoretical or structural basis for the assumption of bimodality, parametric and semi-parametric approaches proceed by fitting either a two (or more) component mixture of distributions, e.g., Aitkin (2011); Vollmer et al. (2013); McLachlan and Peel (2019), or a type of density that has sufficient flexibility to incorporate bimodality for specific parameter configurations, such as the: "odd log normal logistic" (Duarte et al. 2018); flexible generalized skew normal (Venegas et al. 2018; Andrade and Rathie 2016); quartic exponential (Matz 1978; Poitras and Heaney 2015); or, bimodal power normal (Bolfarine, Martínez-Flórez, and Salinas 2018).

Historically, prior to the advent of computing power sufficient to handle complex density specifications, the presence of bimodality was addressed by employing transformations to unimodality, e.g., Baker (1930). Aided by work on the sampling distributions of $\sqrt{b_1}$ and b_2, e.g., Fisher (1930), it was generally held there was little information beyond that revealed by goodness of fit tests employing the standardized sample moments of skewness and kurtosis ($\sqrt{b_1}$ and b_2). As Geary (1947, p.68) observed: "For large samples, $\sqrt{b_1}$ and b_2 are the most efficient tests of skewness and kurtosis, respectively, among large fields of alternative universes". Similarly, Cramér (1946, p.229) observed about parametric testing for normality: "In practice it is usually not advisable to go beyond the third and fourth moments." Subsequently, a variety of more statistically sophisticated omnibus and directional parametric tests for normality based on $\sqrt{b_1}$ and b_2 have been developed including asymptotic and robust Jarque-Bera Lagrange Multiplier (score) tests, D'Agostino's K^2 and tests involving modified measures of skewness and kurtosis, e.g., Romão, Delgado, and Costa (2010:548-54)

Despite substantial theoretical and simulation evidence about the properties of normality tests based on $\sqrt{b_1}$ and b_2, evidence on the properties of the "higher standardized sample moments" ($\sqrt{b_3}$ and b_4) as directional tests or as components of omnibus tests is scant. Included in the few available finite sample simulation and empirical studies examining higher moments as directional tests, Lau, Lau, and Wingender (1990) use the rate of increase in the ratio of the fourth and sixth moments as sample size increases to find that a sample of stock returns does not follow a stable distribution as the moment ratio does not increase as rapidly as expected. A detailed simulation study by Thode, Smith, and Finch (1983) examining the performance of twenty-seven goodness of fit tests for normality in detecting scale contaminated normal samples reports commonly used tests such as the Kolmogorov-Smirnov, Wilk-Shapiro and Anderson-Darling performed poorly compared to directional moment-type tests including those based on the standardized fifth and sixth moments. In contrast to selective use of higher moments as directional tests, higher moments appear in more recent work on series solutions e.g., Rayner, Thas, and Best (2009); Thas (2010, ch.4), Aldo and Yu (2018, ch.4), that motivate omnibus smooth tests with $k > 2$ incorporating higher moments as components along with $k \leq 2$ moments, i.e., $\sqrt{b_1}$ and b_2. Higher moments also appear in calculations for the positive definite and unimodal regions for Gram-Charlier series (Kwon 2019; Lin and Zhang 2020).

Though the use of orthogonal series solutions to model distributional shape stretch back to 19[th] century contributions by Laplace, Thiele and Gram (Hald 2000) – the development of the omnibus "smooth test" by Neyman (1937) provided a seminal series solution method that produces a "smooth alternative of order k to the null hypothesis" by embedding the null hypothesis within a set of distributions that are "close" to the null. As properties and variants of smooth tests were gradually developed correspondences with widely used omnibus goodness of fit tests derived with alternative methods were identified. In addition to the Pearson chi-squared, D'Agostino K^2 and Jarque-Bera LM tests, Koziol (1987) identifies three other instances of univariate normality tests that can be interpreted as smooth tests with a predetermined k. An obvious complication with smooth tests based on orthogonal polynomials is the selection of the free parameter, the degree k of the polynomial. Extending the series solution specification, a smooth test can allow for inclusion of the higher moments $\sqrt{b_3}$ $(k=3)$ and b_4 $(k=4)$ in the polynomial series. However, this imposes parameter restrictions if it is necessary that the associated densities are positive definite, e.g., Kwon (2019).

3. Normality, bimodality and higher moments

At least since Pearson (1894) introduced a moment-based method for determining the "dissection of abnormal frequency curves", a mixture of two normal distributions has been used to model bimodality. Absent theoretical or structural information about the component distributions, this approach presents difficulty determining the mixing parameter (γ) and, for a mixture of normals, the first and second moments of the sub-population distributions (μ_1, $\sigma_{1;}$ μ_2, σ_2). More precisely, consider the conventional specification of the density $f[\cdot]$ for the mixture of two normal distributions:

$$f[x;\mu_1,\mu_2,\sigma_1,\sigma_2,\gamma] = (1-\gamma)\ \Phi[x;\mu_1,\sigma_1] + \gamma\ \Phi[x;\mu_2,\sigma_2]\ \text{ where } 0 \le \gamma \le 1$$

and

$$\Phi[x;\mu,\sigma] = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2o^2}\right]$$

To determine whether the mixture is bimodal, or not, requires unbundling a complicated relationship between the mixing parameter and the first and second moments of the sub-populations where, if possible, the mixture density satisfies conditions for a local minimum ($f[x]' = 0$; $f[x]'' > 0$) at the anti-mode. In general, it does not follow that mixing two normal distributions with distinct modes is sufficient to ensure bimodality of the combined distribution.

Using a mixture of two normal distributions is attractive for depicting bimodality due to the variety of distributional shapes that can be achieved. Available sufficient conditions for determining bimodality of normal mixtures are summarized in the following:

Proposition 1: Sufficient Conditions for Bimodal Normal Mixtures

Evaluating $f\,[x;\,\cdot]\,' = 0$; $f\,[x;\,\cdot]\,'' > 0$ for a mixture of normal densities, $f\,[x;\,\gamma,\,\mu_1,\,\mu_2,\,\sigma_1,\,\sigma_2]$, it follows that, in general, for some, but necessarily not all, $\gamma \in (0,1)$, a sufficient condition for bimodality is (Eisenberger 1964);

$$(\mu_2-\mu_1)^2 > \frac{8\sigma_1^2\ \sigma_2^2}{\sigma_1^2+\sigma_2^2}\ (1a) \qquad or \qquad \frac{(\mu_2-\mu_1)^2\ (\sigma_1^2+\sigma_2^2)}{8\sigma_1^2\ \sigma_2^2} > 1 \qquad (1b)$$

This sufficient condition can be sharpened to: $|\mu_2-\mu_1| > 2\min[\sigma_1,\sigma_2]$ (Behboodian 1970). In the special case where the mixtures have $\sigma_1 = \sigma_2 = \sigma$:

$$|\mu_2-\mu_1| > 2\sigma\sqrt{1+\frac{|\ \text{Log}[\gamma] - \text{Log}[1-\gamma]\ |}{2}} \qquad (1c)$$

A corollary of Proposition 1 using a mixture of normals with different means (modes), $\gamma = .5$ and $\sigma_1 = \sigma_2 = \sigma$ is that $|\mu_2-\mu_1| > 2\ \sigma$ is a necessary condition for bimodality, e.g., Schilling, Watkins, and Watkins (2002). Only limited results are available where the mixing parameter (γ)

and $(\mu_1, \sigma_1; \mu_2, \sigma_2)$ are unrestricted and the first and second derivative conditions for bimodality are satisfied (Kemperman 1991). Consequently, practical attempts to specify goodness of fit tests for normal mixtures that incorporate γ have required some other parameter restriction, e.g., McLachlan and Peel (2000); Vollmer (2009).

In contrast to situations where there is sufficient information regarding component densities, e.g., where there is known group structure in the data such as male and female heights mixing to form the distribution for human height, parameters required for using conditions such as those in Proposition 1 to determine bimodality of a mixture of distributions are often unavailable. In such situations, is it possible to use the higher moments to determine whether a distribution is bimodal, or not? To this end, Table 1 provides calculations of the first six standardized population moments for selected γ and $(\mu_1, \sigma_1; \mu_2, \sigma_2)$ normal mixtures, supplemented by plots of a sub-set of these densities in Figures 1a and 1b. The first three cases in Table 1 satisfy the 1 b) condition for bimodality while the bottom case is for a normal mixture that is not bimodal for any $\gamma \in (0,1)$. These results permit some heuristic observations: $\sqrt{\beta_3}$ and β_4 do not appear to add substantially to directional goodness of fit information about bimodality provided by $\sqrt{\beta_1}$ and β_2 alone; and it is not possible to determine goodness of fit to bimodality based on a consideration of the relative sizes of σ_1 and σ_2 together with the separation of modes μ_2 and μ_1 without also considering the mixing condition γ, even if a sufficient condition from Proposition 1 is satisfied. In addition, the anecdotal claim that bimodality of a normal mixture is associated with kurtosis < 3 appears to be based on an implicit assumption of γ values closer to .5 than .1 (.9).

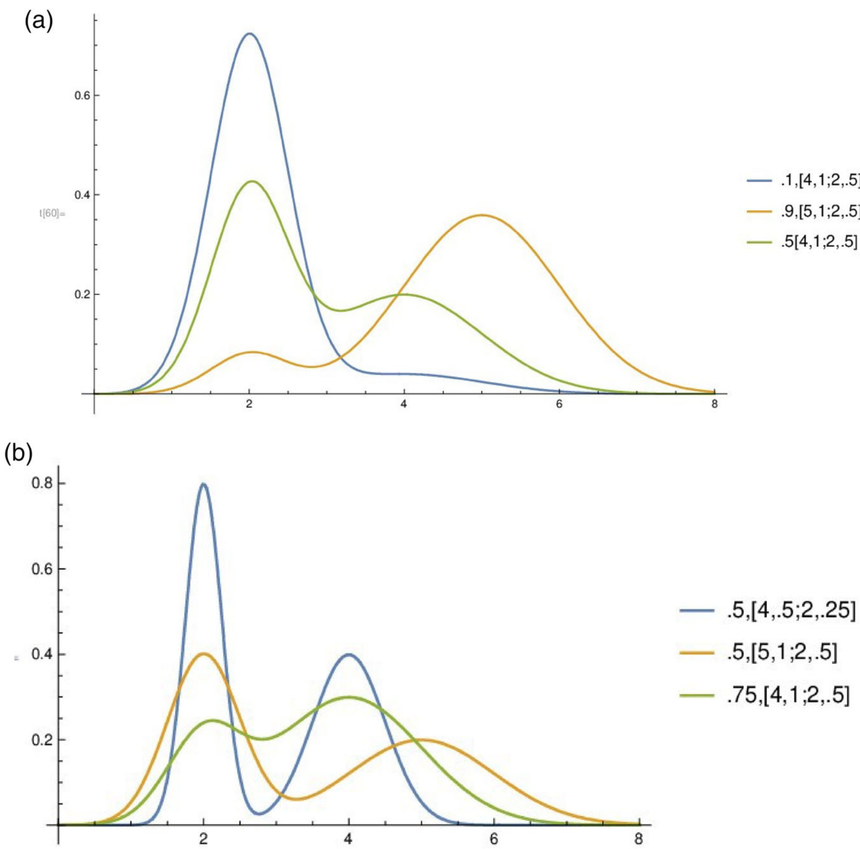

Figure 1. *Legend refers to Table 1 γ [μ_1, σ_1 ; μ_2, σ_2]. In Figure 1a, the $\gamma = .1$ distribution has $\sqrt{\beta_1} = 1.73$ and $\beta_2 = 7.497$; for $\gamma = .9$ (-0.585, 3.196); and $\gamma = .5$ (0.543, 2.402). In Figure 1b, $\gamma = .5$ ($\mu_1 = 4$) (0.226, 1.524); $\gamma = .5$ ($\mu_1 = 5$) (0.346, 1.826); $\gamma = .75$ (0.048, 2.285).

Table 1. Mean, SD and standardized third through sixth moments for different parametric specifications of two normal mixtures.*

		$\mu_1 = 4$	$\mu_2 = 2$	$\sigma_1 = 1$	$\sigma_2 = 0.5$	
Standardized moment (2.5)	$\gamma=$	**.1**	**.25**	**.5**	**.75**	**.9**
μ		**2.2**	2.5	**3**	*3.5*	3.8
σ		**0.828**	1.090	**1.275**	*1.250*	1.134
$\sqrt{\beta_1}$		**1.730**	1.231	**0.543**	*0.048*	−0.117
β_2		**7.497**	4.155	**2.402**	*2.285*	2.708
$\sqrt{\beta_3}$		**26.353**	9.990	**3.203**	*0.730*	−0.397
β_4		**110.589**	30.925	**9.764**	*7.769*	11.035
		$\mu_1 = 4$	$\mu_2 = 2$	$\sigma_1 = 1$	$\sigma_2 = 0.25$	
Standardized moment (10)	$\gamma=$	**.1**	**.25**	**.5**	**.75**	**.9**
μ		2.2	2.5	3	3.5	3.8
σ		0.664	0.927	*1.075*	0.976	0.769
$\sqrt{\beta_1}$		2.311	1.206	*0.226*	−0.579	−1.044
β_2		8.115	3.090	*1.524*	2.145	3.995
$\sqrt{\beta_3}$		27.069	6.216	*0.958*	−2.073	−7.729
β_4		96.419	14.694	*3.207*	5.803	23.031
		$\mu_1 = \mathbf{5}$	$\mu_2 = 2$	$\sigma_1 = \mathbf{1}$	$\sigma_2 = .5$	
Standardized moment (5.6)	$\gamma=$	**.1**	**.25**	**.5**	**.75**	**.9**
μ		2.3	2.75	*3.5*	4.25	**4.7**
σ		1.065	1.458	*1.696*	1.581	**1.317**
$\sqrt{\beta_1}$		2.110	1.228	*0.346*	−0.320	**-0.585**
β_2		7.985	3.490	*1.826*	2.139	**3.196**
$\sqrt{\beta_3}$		27.518	7.571	*1.656*	−1.041	**-3.841**
β_4		104.799	19.971	*5.061*	6.017	**14.935**
		$\mu_1 = \mathbf{4}$	$\mu_2 = 2$	$\sigma_1 = 1$	$\sigma_2 = 1$	
Standardized moment (1)	$\gamma=$	**.1**	**.25**	**.5**	**.75**	**.9**
μ		2.2	2.5	3	3.5	3.8
σ		1.166	1.323	1.414	1.323	1.166
$\sqrt{\beta_1}$		0.363	0.324	0.0	−0.324	−0.363
β_2		3.358	2.878	2.500	2.878	3.358
$\sqrt{\beta_3}$		3.356	2.314	0.0	−2.314	−3.356
β_4		20.024	13.303	9.500	13.303	20.024

*Number in brackets is value from solving condition (1) required for bimodality. The condition is not satisfied for the bottom set of distributions which are =1 when solving for the lhs of the inequality in (1b). Values for standard normal are: $\sigma = 1$; $\sqrt{\beta_1} = \sqrt{\beta_3} = 0$; $\beta_2 = 3$; $\beta_4 = 15$. Results in bold (italics) correspond to densities in Figure 1a (1b).

Figures 1a and 1b provide some visual information about specific Table 1 results. Figure 1a compares the normal mixture distribution with one of the largest population "higher moments" ($\gamma = .1$; $\mu_1 = 4$; $\mu_2 = 2$; $\sigma_1 = 1$; $\sigma_2 = .5$) with a distribution using the same parameters except $\gamma = .5$ and a distribution ($\gamma = .9$; $\mu_1 = 5$; $\mu_2 = 2$; $\sigma_1 = 1$; $\sigma_2 = .5$) with skewness (absolute value) comparable to the $\gamma = .5$ case but with (β_2, β_4) almost identical to normal distribution values ($\beta_2 = 3$; $\beta_4 = 15$). The distributions in Figure 1b were selected for having the lowest values of (β_2, β_4) of the first three distributions listed in Table 1. These results reveal the complicated relationship between the degree of bimodality and the moments associated with mixtures of normal distributions. Figures 1a) and 1b) provide visual evidence that, without incorporating information from the mixing parameter, the sufficient conditions in Proposition 1 are problematic in practice. Similarly, because skewness is controlled by the mixing parameter and the difference in means, a mixing parameter of $\gamma = .5$ with equal volatilities appears necessary to produce zero skewness.

This cursory examination of population higher moments of a mixture of normals raises two questions. Does the use of the higher moments in determining goodness of fit for a bimodal alternative improve where one or both mixture densities is a non-normal distribution? Would the higher moments improve the power of omnibus goodness of fit tests that employ a combination of moments? Answers to the first question serve to situate this study within the vast literature on mixture distributions. In contrast to the semi-parametric use of mixtures to model unknown distributional shapes or construct flexible Bayesian priors, the parametric goodness of fit problem involves using distributional parameters to specify a series approximation to the mixture distribution. The implications of mixing the normal with a non-normal distribution is illustrated in Table 2 and Figure 2 where a normal ($\gamma \in \{.1, .25, .5, .75, .9\}$; $\mu = 4$, $\sigma = 1$) is mixed with a non-central t distribution (centrality parameter $\delta = 8$; degrees of freedom $\nu \in \{5, 7\}$). The contrast with Table 1 results is striking: substantively higher moment (> 2) estimates; and, marked reduction in the visual achievement of bimodality. Whether such decisive differences in population parameters compared to the normal mixtures translates to similar differences in finite sample goodness of fit testing for such types of bimodal distributions is examined in the Monte Carlo simulations.

4. Series expansions and goodness of fit tests

The core goodness of fit problem is concerned with testing a simple or composite null hypothesis that a sample is drawn from a given population or some alternative distribution(s). Determining the effectiveness of the higher moments in goodness of fit testing for univariate normality against bimodal or unimodal alternatives can be motivated by classical series solutions for a continuous location-scale density function, $G[\xi]$, with population mean μ and standard deviation σ. Various accessible sources for the classical results include Johnson, Kotz, and Balakrishnan (1994, p.25-33), Cramér (1946, sec.17.6), Kendall and Stuart (1963, p.155-60). Hall (1992, p.79-81) and Hald (2007, ch.15) are useful summaries of the historical development. It is well known that, though the Gram-Charlier (Type A) and Edgeworth expansions are formally identical, truncation at a finite term in these series produces different results with the remainder for the Gram-Charlier approximation not tending regularly to zero. However, for purposes of specifying goodness of fit tests, the Gram-Charlier expansion has substantive advantages.

It is well known that, because the terms in the Gram-Charlier expansion occur in sequence determined by the order of $n[x]$ derivatives, successive terms in this expansion do not necessarily decrease in order of magnitude, e.g., Cramér (1946, p.225-7). Consequently, the remainder term of a truncated Gram-Charlier series will include terms of the same or lower order of T than terms included in the approximation, possibly resulting in a deteriorating fit when an additional term is included in the approximation, e.g., Kendall and Stuart (1963, p.160). By reordering terms of the Gram-Charlier series in powers of $\sqrt{T}$, the Edgeworth expansion provides a truncated series approximation to $G[\xi]$ with a remainder that converges regularly to zero. However, this analytical advantage comes with increased complexity. For purposes of determining moment restrictions to

Table 2. Mean, SD and standardized third through sixth moments for different parametric specifications of normal and non-central t mixtures.*

		$\mu_1 = 4$	$\sigma_1 = 1$	$\delta = 8$	$\upsilon_2 = 5$	
Standardized moment	**γ=**	**.9**	**.75**	**.5**	**.25**	**.1**
μ		4.55	5.38	6.76	8.137	8.96
σ		5.42	10.90	17.0	19.3	18.85
$\sqrt{\beta_1}$		4.72	3.44	2.59	2.41	2.62
β_2		63.7	35.44	24.92	25.24	29.29
$\sqrt{\beta_3}$		N/A	N/A	N/A	N/A	N/A
β_4		N/A	N/A	N/A	N/A	N/A
		$\mu_1 = 4$	$\sigma_1 = 1$	$\delta = 8$	$\upsilon_2 = 7$	
Standardized moment	**γ=**	**.9**	**.75**	**.5**	**.25**	**.1**
μ		4.50	5.25	6.50	7.76	8.51
σ		4.14	7.92	11.70	12.36	11.24
$\sqrt{\beta_1}$		3.18	2.31	1.56	1.32	1.47
β_2		22.48	12.38	8.12	8.01	9.53
$\sqrt{\beta_3}$		229.1	95.1	54.28	55.0	72.66
β_4		4751.3	1535.1	813.5	902.69	1331.73

*υ is the degrees of freedom and δ is the non-centrality parameter for the non-central t component of the mixture. Moments of the non-central t distribution require the degrees of freedom to be higher than the moment power required.

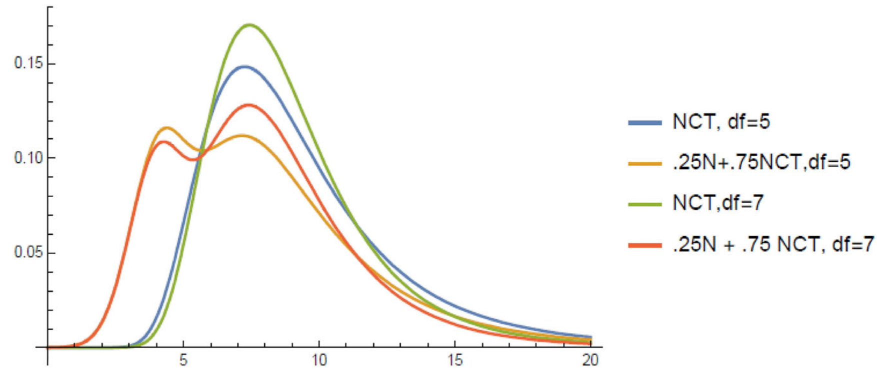

Figure 2. Non-central Student t with $\mu = 8$ and degrees of freedom $\upsilon\ \varepsilon\{5, 7\}$ and Mixture with Normal Distribution [4,1].

ensure a truncated series approximation is unimodal and positive-definite "the boundary of the positive definite and unimodal regions cannot be expressed analytically for the Edgeworth series in general" (Kwon 2019, 10), while such conditions can be derived for the Gram-Charlier (Type A) series. In addition, terms associated with specific powers of $\sqrt{T}$ include complicated combinations of Hermite polynomials and moments of different order undermining goodness of fit test specification.

Extending the classical series solution provided by the Gram-Charlier Type A expansion to goodness of fit testing for a bimodal mixture is complicated by the two additional scale and location parameters and the mixing parameter. However, in many practical situations, the relevant goodness of fit problem for testing the null hypothesis of normality is associated with samples from populations where it is not known if the alternative density is bimodal or unimodal. The series solution to addressing this goodness of fit problem proceeds by approximating the general density, making progressive adjustments to the standard normal density $n[x]$ where $x = (\xi - \mu)\ /\sigma$. This leads to the following well known proposition:

Proposition 2: Gram-Charlier (A-4) Series Approximation

For an *iid* random variable x expressed in standard measure ($\mu_x = 0$, ${\sigma_x}^2 = 1$), the Gram-Charlier (A-4) series approximation applicable to densities $g[x]$ "sufficiently close" to the normal is:

$$g[x] = n[x]\left\{1 - \frac{1}{3!}\frac{\mu_3}{\sigma^3}H_3 + \frac{1}{4!}\left(\frac{\mu_4}{\sigma^4} - 3\right)H_4 - \frac{1}{5!}\left(\frac{\mu_5}{\sigma^5} - 10\frac{\mu_3}{\sigma^3}\right)H_5 + \frac{1}{6!}\left(\frac{\mu_6}{\sigma^6} - 15\frac{\mu_4}{\sigma^4} + 30 + \left(\frac{\mu_3}{\sigma^3}\right)^2\right)H_6\right\}$$

where μ_j is the jth central population moment of $g[x]$ and $(-1)^r H_r$ is the Hermite polynomial determined from the rth derivative of the standard normal density function:

$$(-1)H_1 n[x] = \frac{d}{dx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} = \frac{d}{dx}n[x] = -x\ n[x]$$

$$H_2\ n[x] = \frac{d^2}{dx^2}n[x] = \frac{d}{dx}H_1\ n[x] = \left[x^2 - 1\right]n[x];\ (-1)^3 H_3\ n[x] = \left(3x - x^3\right)n[x]$$

$$H_4 = x^4 - 6x^2 + 3;\ H_5 = x^5 - 10x^3 + 15x;\ H_6 = x^6 - 15x^4 + 45x^2 - 15$$

Proposition 2 can be used to motivate goodness of fit score tests for a normal null hypothesis based on application of these Hermite polynomials to form an orthogonal basis with respect to the standard normal density (Kendall and Stuart 1963, p.155; Hall 1992, p.44). In turn, moment restrictions on $\{\sqrt{\beta_1}, \beta_2\}$ provided by Barton and Dennis (1952), Draper and Tierney (1972) and Spiring (2011) needed for the Gram-Charlier (A-2) series approximation:

$$g^*[x] = n[x]\left\{1 - \frac{1}{3!}\frac{\mu_3}{\sigma^3} H_3 + \frac{1}{4!}\left(\frac{\mu_4}{\sigma^4} - 3\right) H_4\right\}$$

to be positive definite and unimodal are expanded by Kwon (2019) and Lin and Zhang (2020) to encompass the $\{\sqrt{\beta_1}, \beta_2, \sqrt{\beta_3}, \beta_4\}$ regions of the Gram-Charlier (A-4) series.

To achieve some intuition into the rationale for using a Gram-Charlier (A-4) versus (A-2) approximation to test goodness of fit for a bimodal alternative, consider the Hermite polynomial elements involved in fitting the $g[x]$ density:

$$\frac{1}{3!}H_3\ n[x] = H3^*;\ \frac{1}{4!}H_4\ n[x] = H4^*;\ \frac{1}{5!}H_5\ n[x] = H5^*;\ \frac{1}{6!}H_6\ n[x] = H6^*$$

In Proposition 2, each of these weighting functions is multiplied by a term that contains standardized sample moments. Figure 3 illustrates a $\{\sqrt{\beta_1}, \beta_2\} = (.4, 5.6)$ combination where the $g^*[x]$ Gram-Charlier (A-2) series density has been shown to be positive definite and not unimodal. The relevant region of such positive definite, not unimodal, $\{\sqrt{\beta_1}, \beta_2\}$ combinations feature a lowest possible $\beta_2 = 3.4$ that requires $\sqrt{\beta_1}$ =.6; for $\beta_2 > 5.4$ a wider range of $\sqrt{\beta_1} \in [0,1]$ is possible, e.g.,

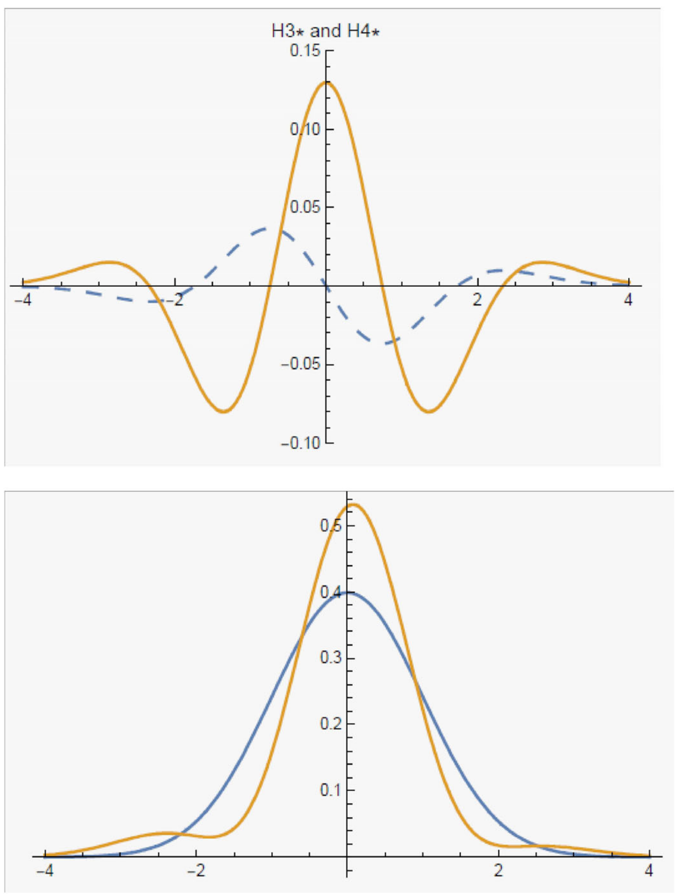

Figure 3. Example of positive definite, non-unimodal Gram-Charlier type (A-2).
$\{\sqrt{\beta_1}, \beta_2\} = (.4, 5.6)$; H3 and H4* with plot of density versus N[0,1].

Draper and Tierney (1972). Within these values, Figure 3 is representative of the trimodal distributions with "bumps" in the two tails of the density generated by the positive definite, non-unimodal Gram-Charlier (A-2) series densities. Comparison of the population parameters in Table 1 and associated density shapes in Figures 1a and 1b reveals that, due to absence of a central anti-mode, such Gram-Charlier (A-2) series are problematic for fitting bimodal normal mixtures.

Figure 4 reconfigures the Gram-Charlier (A-2) series with $\{\sqrt{\beta_1}, \beta_2\}$ values more consistent with values for the bimodal densities reported in Table 1, also providing a Gram-Charlier (A-4) series that includes $\{\sqrt{\beta_3}, \beta_4\}$ values consistent with Table 1. The presence of a small area with negative values in the tails indicates "defective densities" (d-density) that are not fully positive definite. Despite having negative regions, orthogonality of the Hermite polynomials in $g[x]$ of Proposition 2 allows such d-densities to satisfy the essential condition, e.g., Chateau and Dufresne (2017, eq.15):

$$\int_{-\infty}^{+\infty} g[x]\ dx = \int_{-\infty}^{+\infty} n[x]\ dx + \int_{-\infty}^{+\infty} \left(g[x] - n[x]\right)\ dx = 1$$

While in some instances positivity is an essential requirement for the series approximation, e.g., Schlögl (2013); Chateau and Dufresne (2017), the implications for testing the null hypothesis that a sample is drawn from a normal density are unclear. Consequently, because the Gram-Charlier (A-2) d-density shows no evidence of bimodality, a goodness of fit test based on a Gram-Charlier (A-4) d-density approximation could provide higher power when the alternative hypothesis of bimodality is true.

The importance of incorporating higher moments to test goodness of fit for a bimodal alternative is usefully illustrated in Figure 4 by the Gram-Charlier (A-4) d-density; including $\{\sqrt{\beta_3}, \beta_4\}$ in the series solution generates bimodality which is not present in the Gram-Charlier (A-2) solution. However, Figure 5 illustrates two cases that provide serious caution about the promising Gram-Charlier (A-4) d-density in Figure 4. Two cases of Gram-Charlier (A-2) and (A-4) d-densities are

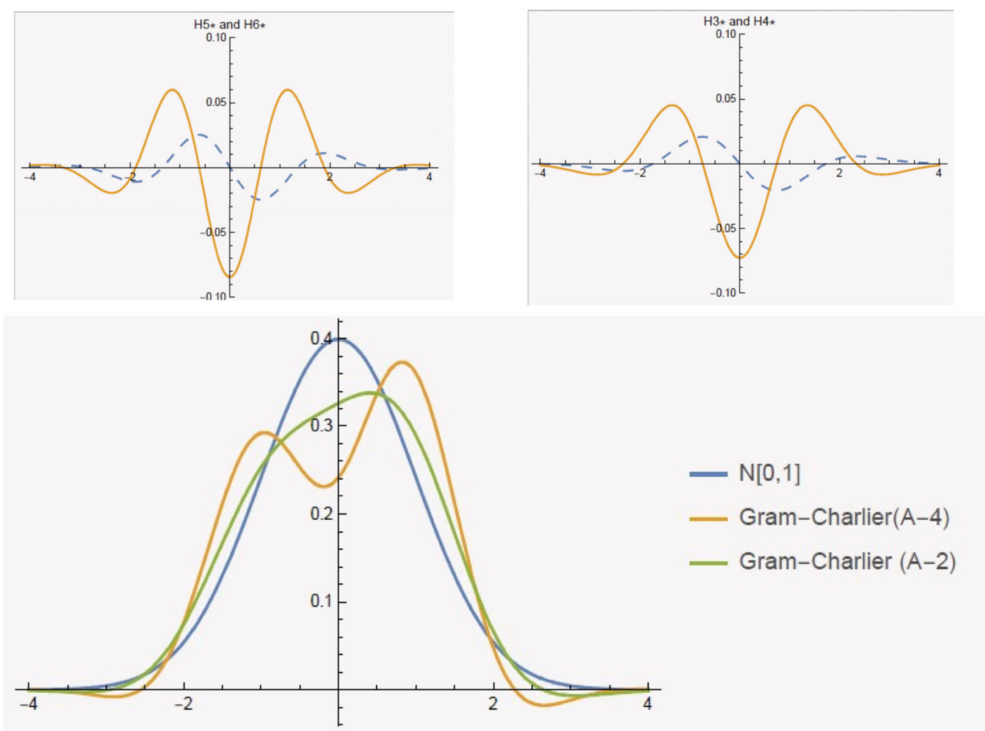

Figure 4. Gram-Charlier (A-2) and (A-4) series solutions for mixture of normals: $\gamma = .5$, N[$\mu_1 = 4$, $\mu_2 = 2$, $\sigma_1 = 0.5$, $\sigma_2 = 0.25$].

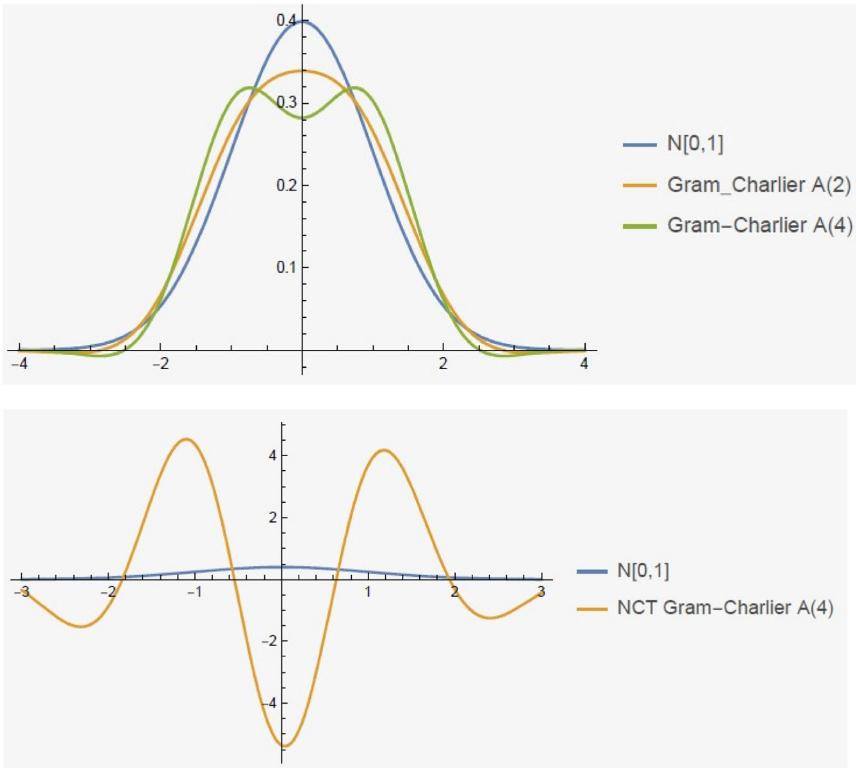

Figure 5. Gram-Charlier series solutions for uniform (A2) and (A4) and normal-non-central t (A4) distributions.

considered using parameters for two different densities: the unimodal uniform; and a normal-+non-central t bimodal mixture. For the uniform density, while the Gram-Charlier (A-2) d-density has a similar shape to the normal, the Gram-Charlier (A-4) d-density gives the misleading impression of bimodality. The normal + non-central t mixture demonstrates the debilitating implications for the non-positive region of the d-density for this type of mixture. This suggests that such bimodal alternatives violate the essential assumption in Proposition 2 that convergence of the series approximation is applicable to densities "sufficiently close" to the normal.

5. Parametric normality tests with higher moments

Various potential approaches are available for constructing directional and omnibus parametric normality tests using standardized sample moments. While methods involving transformations to normality and Monte Carlo simulation of the finite sample distributions are available for tests based on $\sqrt{b_1}$ and b_2, there is an absence of results for tests involving the higher moments leaving normality test specification to be based on asymptotic theory and simulated critical values. In general, while asymptotic normality of the standardized sample moments under the null hypothesis could facilitate specification of omnibus and directional chi-squared tests, D'Agostino and Pearson (1973) and Bowman and Shenton (1975) demonstrate for $\sqrt{b_1}$ and b_2 that finite sample moments are not independent variables, fleshing out a result known since the late 19th century. In addition, it is well known convergence to the asymptotic values becomes increasingly slower as the degree of the moment increases, e.g., b_2 converges more slowly than $\sqrt{b_1}$. Consequently, determining the theoretical solutions for the power and size of such goodness of fit tests involving higher moments in finite samples is also, almost certainly, problematic. With these qualifications in mind, a plausible approach to goodness of test specification is an appropriately specified score test.

Consistent with most practical applications, the null hypothesis of normality is specified in composite form where variates are standardized by estimating μ and σ for each sample using the maximum likelihood estimates and transforming the observations to standard measure. While an omnibus test combining the information in $\sqrt{b_1}$ and b_2 and, possibly, both the higher moments is indicated for a general non-normal alternative, increased power can be obtained by narrowing the class of alternatives and using a directional test targeted at specific distributional characteristics. Where a specific alternative distribution is indicated, the use of an appropriately specified likelihood ratio (LR) test makes it possible that a uniformly most powerful, location and scale invariant (UMP) test is available. However, due to the complexity of determining a solution, the number of such tests is small and, in cases where such tests are available, e.g., the studentized range is the UMP test for normality vs. the uniform distribution (Uthoff 1970), the power of the test is often poor against alternative distributions other than the specific alternative. In cases where the alternative distribution is a specific form of bimodality, a directional test using information from one of the higher moments could be effective. However, absent appropriately adjusted critical values to ensure such directional tests are properly sized, a correctly sized omnibus test using all four standardized sample moments is more practical for a general bimodal alternative than tests using some subset of the four components.

Turning to the construction of omnibus tests that combine information from $\sqrt{b_1}$ and b_2, various approaches have been suggested, e.g., Thode (2002, p.54-8), D'Agostino and Stephens (1986, ch.7 and 9), Dufour et al. (1998). Within the class of score tests, the simplest construction is found with the Jarque-Bera (*JB*) test (Jarque and Bera 1980, 1987; Urzúa 1996; Poitras 2006) where the asymptotic normal values for skewness and kurtosis of the normal distribution $\{\mu_3 = 0, (\mu_4/\sigma^4) = 3\}$ and the asymptotic variance of the standardized sample statistics, $\text{var}[\sqrt{b_1}] = (6/T)$ and $\text{var}[b_2] = (24/T)$, are used to construct a $\chi^2(2)$ test that is specified as:

$$JB = T\left[\frac{(\sqrt{b_1})^2}{6} - \frac{(b_2 - 3)^2}{24}\right] \sim \chi^2(2)$$

where T is the sample size, $\sqrt{b_1} = m_3/(m_2)^{3/2}$, $b_2 = m_4/(m_2)^2$ and the central moments are defined as $m_j = [\sum (x_j - m_1)^{\,j}]/T$ and m_1 is the sample mean. Jarque and Bera (1987) demonstrate that, when the alternative distributions are from the Pearson family – effectively distributions fully defined by $\{\sqrt{b_1}, b_2\}$ – then the *JB* test corresponds to the score (Lagrange multiplier) test for normality. Several subsequent studies, e.g., Gel and Gastwirth (2008); Bontemps and Meddahi (2005), have provided small sample improvements for the power and size of the asymptotic *JB* critical values.

Precise derivation of the score test relies on the analytical assessment of the score statistic (S) evaluated at the restricted estimates:

$$S = \left(\frac{\partial \ ln \ L}{\partial\theta}\right) E\left(-\frac{\partial^2 \ ln \ L}{\partial\theta \ \partial\theta'}\right)^2 \left(\frac{\partial \ ln \ L}{\partial\theta}\right)$$

where L is the likelihood function and θ contains the parameters of interest, determined by the form of the expansion used to specify L. The score test will be asymptotically distributed as $\chi^2(p)$ where p is the number of parameter restrictions being tested. This leads to the following:

Proposition 3: Specification of the score test

Let $\theta = \{(\mu_3/\sigma^3), (\mu_4/\sigma^4), (\mu_5/\sigma^5), (\mu_6/\sigma^6)\}$ and define the likelihood function L using the truncated Gram-Charlier (A-4) series:

$$L = \prod_{t=1}^{T} f[x_t] = \prod_{t=1}^{T} n[x_t]\left\{1 - \frac{1}{3!}\frac{\mu_3}{\sigma^3} H_3 + \frac{1}{4!}\left(\frac{\mu_4}{\sigma^4} - 3\right)H_4 - \frac{1}{5!}\left(\frac{\mu_5}{\sigma^5} - 10\frac{\mu_3}{\sigma^3}\right)H_5\right.$$

$$+\ \frac{1}{6!}\left(\frac{\mu_6}{\sigma^6}\ -\ 15\frac{\mu_4}{\sigma^4}+\ 30+\left(\frac{\mu_3}{\sigma^3}\right)^2\right)H_6\Bigg\}$$

To test for the null hypothesis of normality $\{(\mu_3/\sigma^3) = (\mu_5/\sigma^5) = 0, (\mu_4/\sigma^4) = 3, (\mu_6/\sigma^6) = 15\}$, by evaluating the derivatives of the log likelihood – where the maximum likelihood estimators of $\sqrt{b_1}$, b_2, $\sqrt{b_3}$ and b_4, are used to estimate the relevant parameters (Klar 2000) and the orthogonality of the Hermite polynomials allows the cross partial derivatives in the information matrix to be set to zero – leads to the following omnibus goodness of fit test statistic:

$$HM4 = T\left[\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)^2}{24} + \frac{(\sqrt{b_3} - 10\sqrt{b_1})^2}{120} + \frac{(b_4 - 15b_2 + 30)^2}{720}\right] \sim \chi^2(4)$$

A general solution for the information matrix applicable to use of Hermite polynomials is provided by Cox and Hinkley (1974, p.111). Kiefer and Salmon (1983) provide solutions for the relevant derivatives that maximize the log likelihood and Bontemps and Meddahi (2005) provide further theoretical and Monte Carlo details.

The score test of Proposition 3 presents various analytical requirements and implementation complications. Given the values for the standardized fifth and sixth central moments for the normal distribution $\{(\mu_5/\sigma^5) = 0, (\mu_6/\sigma^6) = 15\}$, the asymptotic variances of $\sqrt{b_3}$ and b_4 can be obtained from results for k statistics, e.g., Kendall and Stuart (1963, p.292-3), as $\text{var}[\sqrt{b_3}] = (120/T)$ and $\text{var}[b_4] = (720/T)$. Being constructed as an omnibus test for normality against a general alternative, HM_4 inherits the local power shortcomings associated with omnibus tests that have been documented in numerous studies. Using the terminology of Escanciano (2009) facilitates the practical question: does a useful subset of bimodal alternatives lie within the "principal space" where the *HM4* test has substantial local power? This question can be clarified by augmenting investigation of *HM4* with "simple" directional test statistics based on the higher moments that can be compared with results for parametric omnibus tests – *JB*, D'Agostino's K^2 – and directional tests – $\sqrt{b_1}$ and b_2. These parametric directional tests are *M5, M6* and *HM2*:

$$M5 = T\left[\frac{(\sqrt{b_3} - 10b_2)^2}{120}\right] \sim \chi^2(1) \quad M6 = T\left[\frac{(b_4 - 15b_2 + 30)^2}{720}\right] \sim \chi^2(1)$$

$$HM2 = \text{T}\left[\frac{(\sqrt{b_3} - 10\sqrt{b_1})^2}{120} + \frac{(b_4 - 15b_2 + 30)^2}{720}\right] \sim \chi^2(2)$$

where $\sqrt{b_3} = m_5/(m_2)^{5/2}$, $b_4 = m_6/(m_2)^3$. Being based on statistics with slow rates of convergence to asymptotic normal values, it is expected that the critical values for the sampling distributions for *M5, M6, HM2* and *HM4* will differ from the chi-squared distribution in finite samples. In other words, such tests will be incorrectly sized for testing the null hypothesis of normality. As has been done for the *JB* test it is possible, for large enough sample sizes, to simulate correctly sized critical values under the assumption that the null hypothesis of normality is true. In the following, simulation of critical values produces two alternative test versions: *JB** and *HM4**, where (*) indicates tests that are correctly sized under the assumption that the null hypothesis is true. Whether size corrected critical values account for complications associated with convergence of the statistics to asymptotic values and finite sample dependence of the standardized sample moments is unclear. Consequently, the results in the Monte Carlo simulations need to be interpreted with some caution.

6. Directional and omnibus tests: Monte Carlo results

Testing goodness of fit for normality versus bimodal and selected unimodal alternatives is complicated by the vast number of possible alternative density specifications. Considering the divergence from positivity of the Gram-Charlier (Type A) series approximation using parameters from the normal-non-central t mixture illustrated in Figure 5, where does the search for a "principal space" with alternatives "sufficiently close" to the normal begin? To benchmark this search using bimodal density population parameters reported in Table 1, Table 3 tabulates calculations for goodness of fit tests – $HM4 = JB + HM2$ and $HM2 = M5 + M6$ – with $T = 50$ scaling the calculations. Recognizing that zero will result where normal distribution population parameters are used, as T increases $HM4$ and JB for each distribution considered will eventually reject the null hypothesis. Hence, Table 3 provides some indication about the convergence rate for the tests as sample size increases. Specifically, in Table 3 the largest JB and $HM4$ values occur for $\gamma = .1$ mixtures that have higher weight on the lower mean and variance component. Based on Figures 1a and 1b, this indicates that the largest calculated JB and $HM4$ values do not correspond to the $\gamma = .5$ [4, .5; 2, .25] density with the most visually significant degree of bimodality. The mixture with $\gamma \in .5$ and equal variance that, from Table 1, is neither normal nor bimodal exhibits such small values that exceedingly large samples would be required for goodness of fit testing of such alternative distributions to reject the null hypothesis.

Though a useful first step, evaluating goodness of fit tests using population parameters fails to address test power and size complications that arise in finite samples. To benchmark the Monte Carlo simulations, Table 4 reports the power of the directional and omnibus goodness of fit test statistics for three sample sizes, $T = \{25, 50{,}100\}$ when the null hypothesis of normality is true. Tables 5 and 6 report test power results for these sample sizes when the alternative hypothesis is true. Table 5 examines nine normal mixture distributions and Table 6 four unimodal distributions. A range of directional and omnibus parametric tests are reported: $\sqrt{b_1}$ and b_2 using $\chi^2(1)$ critical values and $\sqrt{b_1}^*$ and b_2^* evaluated using the size corrected critical values provided by Bowman and Shenton (1986); the higher moment directional tests $M5$ and $M6$ evaluated using the $\chi^2(1)$ value; the JB and higher moment $HM2$ tests evaluated using the $\chi^2(2)$ value and $HM4$

Table 3. Omnibus and directional normality tests evaluated using population parameters of normal mixture densities for T = 50.*

		$\mu_1 = 4$	$\mu_2 = 2$	$\sigma_1 = 1$	$\sigma_2 = 0.5$	
Tests	**γ=**	**.1**	**.25**	**.5**	**.75**	**.9**
JB		**67.1**	15.4	**3.20**	*1.08*	0.29
HM4		**156.2**	17.8	**6.24**	*1.96*	0.55
M5		**34.15**	2.24	**2.07**	*0.26*	0.25
M6		**54.97**	0.14	**0.97**	*0.85*	0.01
HM2		**89.12**	2.38	**3.04**	*0.87*	0.26
		$\mu_1 = 4$	$\mu_2 = 2$	$\sigma_1 = .5$	$\sigma_2 = 0.25$	
Tests	**γ=**	.1	.25	.5	.75	.9
JB		99.0	12.13	*4.96*	4.31	11.15
HM4		107.1	26.56	*13.1*	10.99	17.51
M5		6.53	14.23	*0.71*	5.76	3.06
M6		1.53	0.19	*7.43*	0.91	3.30
HM2		8.06	14.42	*8.14*	6.67	6.36
		$\mu_1 = 5$	$\mu_2 = 2$	$\sigma_1 = 1$	$\sigma_2 = .5$	
Tests	**γ=**	**.1**	**.25**	**.5**	**.75**	**.9**
JB		88.9	13.07	*3.86*	2.40	**2.93**
HM4		121.7	22.70	*9.31*	5.41	**5.24**
M5		17.16	9.24	*1.36*	1.94	**1.68**
M6		15.68	0.39	*4.09*	1.07	**0.63**
HM2		32.84	9.63	*5.45*	3.01	**2.31**
		$\mu_1 = 4$	$\mu_2 = 2$	$\sigma_1 = 1$	$\sigma_2 = 1$	
Tests	**γ=**	**.1**	**.25**	**.5**	**.75**	**.9**
JB		1.36	0.91	0.52	0.91	1.36
HM4		1.40	1.26	0.80	1.26	1.40
M5		0.03	0.36	0.0	0.36	0.03
M6		0.01	0.00	0.28	0.00	0.01
HM2		0.04	0.36	0.28	0.36	0.04

*T is a scaling variable. At 5% (10%) level, x^2 **(2)** = 5.99 (4.61) and x^2**(4)** = 9.488 (7.78). Results in bold (italics) correspond to densities in Figure la (lb).

Table 4. Estimated goodness of fit test powers ($\alpha = 10\%$) when the null hypothesis of normality is true: T = 25, 50, 100.*

	$\sqrt{b_1}$*	b_2*	SR	K^2	JB*	JB	HM2	HM4	$\sqrt{b_1}$	b_2	M5	M6	HM4*
T = 25	.093	.101	.106	.071	.100	.033	.006	.036	.059	.021	.013	.010	.100
T = 50	.092	.092	.104	.077	.100	.050	.016	.066	.064	.031	.027	.012	.100
T = 100	.087	.087	.100	.083	.100	.060	.016	.100	.080	.060	.034	.027	.100

Tabulated critical values for $\sqrt{b_1}$, b_2* taken from Bowman and Shenton (D'Agostino and Stephens 1986, p.300-1). Critical values for the studentized range (*SR*) provided by Pearson and Stephens (1964) with the maximum likelihood estimate (*MLE*) for the standard deviation used to determine *SR*. All other tests are calculated using the chi-squared critical values. The *MLE* for the standard deviation is used to standardize pseudo random numbers. The number of iterations at each sample size are 8000 for $T = 25$, and 4000 for $T = 50$ and 2000 for $T = 100$.

Simulated Size corrected *JB** and *HM4** test 10% critical values are estimated as: $T = 25$ *JB** 2.50 *HM4** 5.37; $T = 50$ *JB** 3.38 *HM4** 6.51; $T = 100$ *JB** 3.575 *HM4** 7.78

Table 5. Estimated goodness of fit test powers ($\alpha = 10\%$) when the alternative hypothesis of bimodality is true: $T = 25$, 50, 100.*

	$\sqrt{b_1}$*	b_2*	SR	K^2	JB*	JB	HM2	HM4	$\sqrt{b_1}$	b_2	M5	M6	HM4*
Normal Mixture ($\gamma = .5$, N[4,1], N[2,.5])													
T = 25	.269	.276	.301	.276	.309	.072	.060	.178	.186	.012	.189	.059	.508
T = 50	.471	.380	.470	.550	.636	.267	.318	.680	.412	.116	.404	.157	.788
T = 100	.795	.494	.612	.838	.972	.912	.772	.980	.780	.315	.710	.321	.980
Normal Mixture ($\gamma = .25$, N[4,1], N[2,.5])													
	$\sqrt{b_1}$*	b_2*	SR	K^2	JB*	JB	HM2	HM4	$\sqrt{b_1}$	b_2	M5	M6	HM4*
T = 25	.839	.342	.144	.570	.839	.581	.340	.675	.752	.230	.484	.304	.860
T = 50	.992	.439	.144	.804	.986	.954	.902	.978	.984	.392	.697	.381	.992
T = 100	.996	.606	.168	.998	1.00	1.00	.972	1.00	1.00	.604	.762	.496	1.00
Normal Mixture ($\gamma = .1$, N[4,1], N[2,.5])													
	$\sqrt{b_1}$*	b_2*	SR	K^2	JB*	JB	HM2	HM4	$\sqrt{b_1}$	b_2	M5	M6	HM4*
T = 25	.775	.626	.283	.695	.793	.707	.818	.884	.741	.518	.845	.689	.83 8
T = 50	.946	.883	.346	.923	.954	.934	.978	.992	.942	.866	.968	.944	.976
T = 100	.988	.990	.388	1.00	.992	.992	1.00	1.00	.980	.990	1.00	.996	1.00
Normal Mixture ($\gamma = .75$, N[4,1], N[2,.5])													
	$\sqrt{b_1}$*	b_2*	SR	K^2	JB*	JB	HM2	HM4	$\sqrt{b_1}$	b_2	M5	M6	HM4*
T = 25	.033	.233	.217	.120	.044	.006	.004	.030	.014	.002	.016	.033	.161
T = 50	.031	.394	.334	.277	.052	.006	.020	.174	.020	.079	.021	.106	.300
T = 100	.031	.666	.512	.555	.344	.120	.128	.508	.024	.409	.038	.271	.508
Normal Mixture ($\gamma = .5$, N[5,1], N[2,.5])													
	$\sqrt{b_1}$*	b_2*	SR	K^2	JB*	JB	HM2	HM4	$\sqrt{b_1}$	b_2	M5	M6	HM4*
T = 25	.131	.634	.678	.565	.399	.028	.194	.422	.092	.036	.173	.309	.789
T = 50	.204	.854	.880	.916	.858	.388	.732	.960	.090	.608	.292	.733	.982
T = 100	.390	.989	.964	.994	1.00	.992	.996	1.00	.390	.956	.488	.985	1.00
Normal Mixture ($\gamma = .9$, N[5,1], N[2,.5])													
	$\sqrt{b_1}$*	b_2*	SR	K^2	JB*	JB	HM2	HM4	$\sqrt{b_1}$	b_2	M5	M6	HM4*
T = 25	.329	.124	.113	.192	.320	.134	.051	.167	.228	.054	.119	.033	.323
T = 50	.551	.110	.104	.295	.464	.312	.176	.398	.460	.071	.268	.067	.488
T = 100	.830	.095	.124	.527	.784	.672	.504	.744	.828	.093	.532	.180	.744
Normal Mixture ($\gamma = .5$, N[4,5], N[2,.25])													
	$\sqrt{b_1}$*	b_2*	SR	K^2	JB*	JB	HM2	HM4	$\sqrt{b_1}$	b_2	M5	M6	HM4*
T = 25	.085	.871	.914	.863	.733	.017	.544	.826	.059	.199	.164	.673	.972
T = 50	.104	.995	.986	.995	.990	.828	.908	1.00	.096	.925	.211	.961	1.00
T = 100	.168	1.00	.996	***	1.00	1.00	1.00	1.00	.180	.998	.332	.999	1.00
Normal-Non-Central *t* Mixture ($\gamma = .75$, N[4,2], NCT[$\upsilon = 7$, $\delta = 8$])													
	$\sqrt{b_1}$*	b_2*	SR	K^2	JB*	JB	HM2	HM4	$\sqrt{b_1}$	b_2	M5	M6	HM4*
T = 25	.489	.321	.183	.379	.491	.349	.123	.348	.456	.231	.159	.091	.447
T = 50	.741	.492	.262	.599	.704	.620	.850	.982	.730	.452	.344	.261	.650
T = 100	.943	.712	.372	.848	.924	.896	.712	.892	.940	.706	.544	.467	.892
Normal Mixture (γ=.25, N[4,1], N[2,1])													
	$\sqrt{b_1}$*	b_2*	SR	K^2	JB*	JB	HM2	HM4	$\sqrt{b_1}$	b_2	M5	M6	HM4*
T = 25	.050	.156	.147	.076	.042	.007	.001	.011	.020	.004	.009	.012	.093
T = 50	.037	.228	.184	.153	.024	.012	.008	.058	.020	.030	.004	.040	.148
T = 100	.029	.350	.232	.256	.104	.012	.076	.188	.028	.158	.012	.084	.188

*NCT is the non-central Student t distribution with degrees of freedom v and centrality parameter δ. See Notes to Table 4 for further details on tests and critical values. *** indicates 24.5% of calculations involved negative square roots.

evaluated using $\chi^2(4)$. Two size corrected versions of the omnibus tests, *JB** and *HM4**, are also reported. Finally, two additional "powerful and informative" parametric goodness of fit tests also reported are: the D'Agostino K^2 test (D'Agostino, Belanger, and D'Agostino 1990; D'Agostino 1970) based on transformations of $\sqrt{b_1}$ and b_2 that aims to be correctly sized using $\chi^2(2)$; and the studentized range test (*SR*) that is UMP for the rectangular (uniform) distribution.

Due to differing methods of determining unbiased moments in small samples, all results in Tables 4-6 are "standardized" using asymptotically unbiased "raw" central moments not adjusted to be unbiased in finite samples. As the Bowman and Shenton (1986) simulated size corrected critical values for $\sqrt{b_1}^*$ and b_2^* use unbiased finite sample estimators, deviations from the 10% critical value in Table 4 reflect a combination of not using unbiased moments and sampling variation associated with simulated critical values. A similar comment applies to results for *SR* which also uses critical values that are correctly sized using the unbiased finite sample standard deviation. Being based on transformations of $\sqrt{b_1}$ and b_2 for a normal null hypothesis that result in correct sizing using χ^2 values (D'Agostino, Belanger, and D'Agostino 1990), the deviations from 10% in K^2 reported in Table 1 are due to sampling variation in this statistic and finite sample correlation between the two components of the statistic. In turn, size correction adjustments simulated for *JB** and *HM4** are reported. Correct sizing to achieve $\alpha = 10\%$ produced critical value estimates for *JB** and *HM4** that are smaller than the asymptotic χ^2 value. In turn, the central role given to estimates of test power and the use of simulated critical values requires attention be given to the implications of Monte Carlo error, e.g., Koehler, Brown, and Haneuse (2009).

Two key sources of Monte Carlo error arise from the properties of pseudo random number generators and the number of replications used to estimate test power. For Table 4, to control for possible sequence bias inherent in the generator, 20 blocks of 25,100 normally distributed variates were created (from Mathematica 11 using RandomVariate with a four to six digit seed changed for each block) and the first four moments of each block calculated. The most extreme eight blocks ($\mu = -.0026$; $\sigma = 1.007$; $\sqrt{b_1} = .0550$, $b_2 = .0040$) were discarded and the eight blocks with values closest to a normal population ($\mu = -.0005$; $\sigma = 1.008$; $\sqrt{b_1} = -.0129$, $b_2 = .0015$) were used for the Monte Carlo results reported in Tables 4-6. Variations in parametric normality test power due to Monte Carlo error when the null hypothesis of normality is true were estimated by comparing simulated critical values between the included (Sample 1) and discarded (Sample 2) blocks for *JB**. The upper bound on possible Monte Carlo error associated with the critical values for *JB** calculated for the eight blocks of Sample 2 and (Sample 1) are: $T = 25$, 2.56 (2.50); $T = 50$, 3.45 (3.38); $T = 100$, 4.15 (3.575).

Comparison of results for *JB** and *HM4** in Table 5 reveals that *JB** has lower power to reject the null for almost all normal mixture sample sizes but not for the normal-non-central t mixture. For "balanced" mixtures with $\gamma \in .5$, the power advantage of *HM4** over *JB** is substantial. Considering discussion in Sec. 4 and Figures 3-5, the implication is that inclusion of the higher moments $\sqrt{b_3}$ and b_4 in *HM4** provides a closer series approximation to the bimodal normal mixtures than for the approximation only provided by $\sqrt{b_1}$ and b_2 in *JB**. Sample size larger than $T = 100$ is usually required for *JB** to have power comparable to *HM4** for mixtures "sufficiently close" to the normal. For mixtures that are not "sufficiently close" as reflected in Figure 5 for the normal-non-central t, the inherent power advantage of *HM4** relative to *JB** is undermined by a slower rate of convergence. For the normal-normal mixtures where $\sqrt{b_1}^*$ has low power, *JB** has lower power than the K^2 and, especially, the *SR* omnibus tests for most sample sizes. However, for the asymmetric normal mixtures, *SR* has the lowest omnibus test power. For those normal mixtures where directional tests are most powerful, *HM4** has comparable power.

Results for four unimodal alternative distributions are considered in Table 6 – the uniform, lognormal, skew normal and skewed stable with characteristic exponent (stability index) of 1.6. The uniform is light tailed and is used to benchmark the power of the other goodness of fit tests against the studentized range which is the UMP test for this distribution. The positively skewed

Table 6. Estimated goodness of fit test powers ($\alpha = 10\%$) when the unimodal alternative hypothesis is true: $T = 25$, 50, 100.

	$\sqrt{b_1}$*	b_2*	*SR*	*M5*	*M6*	*JB*	*HM4*	*JB**	*HM2*	*HM4**	K^2
True H_1: **Skew Normal** [$\ell = 4$, $s = 1$, $\theta = 5$] [$\sqrt{\beta_1} = 0.85$ $\beta_2 = 3.7$ $\sqrt{\beta_3} = 8.5$ $\beta_4 = 30.5$]											
T = 25	.450	.212	.123	.129	.024	.240	.218	.410	.010	.406	.284
T = 50	.758	.267	.152	.292	.079	.522	.534	.686	.202	.676	.492
T = 100	.968	.336	.188	.476	.171	.928	.928	.980	.424	.928	.772
True H_1: **Stable** (Type 1) [$\tau = 1.6$, $\rho = 1$, $\ell = 5$, $s = 1$]											
T = 25	.674	.533	.333	.361	.280	.579	.572	.665	.339	.635	.598
T = 50	.902	.754	.524	.630	.556	.854	.826	.876	.830	.864	.861
T = 100	.995	.933	.736	.870	.828	.980	.976	.984	.976	.976	.982
True H_1: **Uniform** [0, 1][$\mu = .5$, $\sigma = 1/12$, $\beta_2 = 1.8$, $\beta_4 = 3.86$]											
T = 25	.016	.572	.697	.013	.140	.001	.103	.066	.019	.445	.377
T = 50	.014	.924	.978	.039	.576	.032	.736	.428	.310	.826	.865
T = 100	.016	.998	1.00	.038	.959	.904	1.00	.972	.928	1.00	.998
True H_1: **Lognormal** [$\mu_N = 1.4$, $\sigma_N = 0.7$] [$\sqrt{\beta_1} = 2.88$ $\beta_2 = 20.8$ $\sqrt{\beta_3} = 222.9$ $\beta_4 = 3823.7$]											
T = 25	.883	.578	.205	.451	.236	.725	.764	.884	.343	.873	.726
T = 50	.996	.814	.301	.661	.511	.978	.992	.998	.818	.996	.942
T = 100	1.00	.955	.432	.849	.785	1.00	1.00	1.00	.968	1.00	.996

*For the Skew Normal distribution parameters, ℓ = location, s = scale and θ = shape. For the Stable (Type 1) distribution parameters, τ = characteristic exponent (index of stability), ρ = skewness, ℓ = location, s = scale. See Notes to Table 4 for further details on tests and critical values. For the Lognormal μ_N and σ_N are the mean and variance derived from the normal distribution.

lognormal is included because of its importance in practical applications. In addition to location, scale and $\tau = 1.6$ stability index parameters, the fat-tailed stable distribution also has skewness parameter $\rho = 1$. Significantly, in contrast to the results for the mixture of normals, *HM4** has slightly lower power than *JB** for these three unimodal distributions. Only for the uniform distribution where *SR* is UMP is the power of *HM4** greater than *JB**. Comparing the omnibus *HM4** test to the directional tests $\sqrt{b_1}$* and b_2* for the skew normal, *HM4** power is comparable to $\sqrt{b_1}$* – the directional test with highest power. For the two skewed and fat tailed distributions – the lognormal and stable – the power of *HM4** exceeds the power of both $\sqrt{b_1}$* and b_2*.

7. Directions for future research

The promising results for the inclusion of higher moments in goodness of fit tests for normal mixtures reported in the Monte Carlo simulations, combined with the general absence of studies for such tests, provides considerable potential future research directions. Along the lines of Bowman and Shenton (1986) for the moments of the $\sqrt{b_1}$ and b_2 distributions, simulated estimates of the moments for *HM4* and the higher moments for a wide range of finite sample sizes, both smaller than 25 and larger than 100, would facilitate use in empirical studies of a wider range of correctly sized *HM4** critical values and allow the implementation of directional tests using components of *HM4**. In addition, determining correctly sized versions of *M5*, *M6* and *HM2* for power comparison with $\sqrt{b_1}$*, b_2*, *JB** and other goodness of fit tests such as the Shapiro-Wilk, Anderson-Darling, Kolmogorov-Smirnov and Cramer-von Mises across a range of true alternatives would clarify the results obtained for *HM4**. Another line for future research concerns the convergence of the Gram-Charlier series. As remainder term does not necessarily tend regularly to zero, it is likely that adding seventh and eighth moments to the series could improve the power of goodness of fit for certain alternative distributions. This could assist in determining what classes of multimodal distributions are "sufficiently close" to the normal such that *HM4** has higher power compared to *JB**. Comparison with tests based on the L-moments proposed by Hosking (1990) that have been shown to have desirable small sample properties might also be useful. Finally, given the lower power of *HM4** compared to *JB** as an omnibus test for the unimodal asymmetric distributions in Table 6 but not the symmetric uniform distribution, varying stable distributions over differing combinations of the stability index {τ: $0 \leq \tau \leq$

2} and skewness parameter {ρ: $-1 \leq \delta \leq 1$} would clarify whether JB^* is consistently superior to $HM4^*$ for unimodal distributions.

ORCID

Geoffrey Poitras http://orcid.org/0000-0003-0479-5130

References

Aitkin, M. 2011. How many components in a finite mixture. Chapter 13 in K. Mengerson, C. Robert and D. Titterington (ed.), *Mixtures: Estimation and Application*. New York, NY: John Wiley.

Aldo, M., and P. Yu. 2018. *A parametric approach to nonparametric statistics*. Cham: Springer International Publishing.

Andrade, B., and P. Rathie. 2016. Fitting asymmetric bimodal data with selected distributions. *Journal of Statistical Computation and Simulation* 86 (16):3205–24. doi:10.1080/00949655.2016.1157180.

Baker, G. 1930. Transformations of bimodal distributions. *The Annals of Mathematical Statistics* 1 (4):334–44. doi:10.1214/aoms/1177733063.

Barton, D., and K. Dennis. 1952. The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal. *Biometrika* 39 (3-4):425–7. doi:10.2307/2334037.

Behboodian, J. 1970. On the modes of a mixture of two normal distributions. *Technometrics* 12 (1):131–9. doi:10.1080/00401706.1970.10488640.

Bolfarine, H., G. Martínez-Flórez, and H. Salinas. 2018. Bimodal symmetric-asymmetric power-normal families. *Communications in Statistics - Theory and Methods* 47 (2):259–76. doi:10.1080/03610926.2013.765475.

Bontemps, C., and N. Meddahi. 2005. Testing normality: A GMM approach. *Journal of Econometrics* 124 (1):149–86. doi:10.1016/j.jeconom.2004.02.014.

Bowman, K. O., and L. R. Shenton. 1975. Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika* 62 (2):243–50. doi:10.2307/2335355.

Bowman, K. O., and L. R. Shenton. 1986. Moment ($\sqrt{\beta_1}$, β_2) Techniques. Chapter 7 in R. D'Agostino and M. Stephens, *Goodness-of-Fit Techniques*. New York: Marcel Dekker.

Chateau, J.-P., and D. Dufresne. 2017. Gram-Charlier processes and applications to option pricing. *Journal of Probability and Statistics* 2017:1–19. doi:10.1155/2017/8690491.

Cox, D., and D. Hinkley. 1974. *Theoretical statistics*. London, UK: Chapman and Hall.

Cramér, H. 1946. *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.

D'Agostino, R. 1970. Transformation to normality of the null distribution of g_1. *Biometrika* 57:679–81.

D'Agostino, R., A. Belanger, and R. D'Agostino. Jr 1990. A suggestion for using powerful and informative tests of normality. *American Statistician* 44:316–21.

D'Agostino, R., and E. S. Pearson. 1973. Tests for departures from normality. Empirical results for the distribution of $\sqrt{b_1}$ and b_2. *Biometrika* 60 (3):613–22. doi:10.2307/2335012.

D'Agostino, R., and M. Stephens. 1986. *Goodness-of-fit techniques*. New York, NY: Marcel Dekker.

Draper, N., and D. Tierney. 1972. Regions of positive and unimodal series expansion of the Edgeworth and Gram-Charlier expansion. *Biometrika* 59 (2):463–5. doi:10.1093/biomet/59.2.463.

Duarte, G., A. Braga, D. Miquelluti, and V. Ozaki. 2018. Modeling of soybean yield using symmetric, asymmetric and bimodal distributions: Implications for crop insurance. *Journal of Applied Statistics* 45 (11):1920–37. doi:10.1080/02664763.2017.1406902.

Dufour, J., A. Farhat, L. Gardiol, and L. Khalaf. 1998. Simulation-based finite sample normality tests in linear regressions. *The Econometrics Journal* 1 (1):C154–C173. doi:10.1111/1368-423X.11009.

Einbeck, J., and J. Taylor. 2013. A number-of-modes reference rule for density estimation under multimodality. *Statistica Neerlandica* 67 (1):54–66. doi:10.1111/j.1467-9574.2012.00531.x.

Eisenberger, I. 1964. Genesis of bimodal distributions. *Technometrics* 6 (4):357–63. doi:10.1080/00401706.1964.10490199.

Escanciano, J. C. 2009. On the lack of power of omnibus specification tests. *Econometric Theory* 25 (1):162–94. doi:10.1017/S0266466608090051.

Fisher, R. 1930. The moments of the distribution for normal samples of measures of departures from normality. *Proceedings of the Royal Society A* 130: 16–28.

Geary, R. 1947. Testing for normality. *Biometrika* 34 (Pt 3–4):209–42.

Gel, Y., and J. Gastwirth. 2008. A robust modification of the Jarque–Bera test of normality. *Economics Letters* 99 (1):30–2. doi:10.1016/j.econlet.2007.05.022.

Hald, A. 2007. *A history of parametric statistical inference from Bernoulli to Fisher, 1713–1935*. New York: Springer.

Hald, A. 2000. The early history of the cumulants and the Gram-Charlier series. *International Statistical Review* 68 (2):137–53. doi:10.1111/j.1751-5823.2000.tb00318.x.
Hall, P. 1992. *The bootstrap and Edgeworth expansion*. New York, NY: Springer-Verlag.
Hall, P., and M. York. 2001. On the calibration of Silverman's test for multimodality. *Statistica Sinica* 11:515–36.
Hosking, J. 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B (Methodological)* 52:105–24.
Jarque, C., and A. Bera. 1980. Efficient tests for normality, homoskedasticity and serial independence of regression residuals. *Economics Letters* 6 (3):255–9. doi:10.1016/0165-1765(80)90024-5.
Jarque, C., and A. Bera. 1987. A test for normality of observations and regression residuals. *International Statistical Review / Revue Internationale de Statistique* 55 (2):163–72. doi:10.2307/1403192.
Johnson, N., S. Kotz, and N. Balakrishnan. 1994. *Continuous univariate distributions*. Vol. I. New York: Wiley-Interscience.
Kemperman, J. 1991. Mixtures with a limited number of modal intervals. *Annals of Statistics* 19:2120–44.
Kendall, M., and A. Stuart. 1963. *The advanced theory of statistics*. Vol. I, 2nd ed. London: Charles Griffin.
Kiefer, N., and M. Salmon. 1983. Testing normality in econometric models. *Economics Letters* 11 (1–2):123–7. doi:10.1016/0165-1765(83)90172-6.
Klar, B. 2000. Diagnostic smooth tests of fit. *Metrika* 52:237–52.
Koehler, E., E. Brown, and S. Haneuse. 2009. On the assessment of Monte Carlo error in simulation-based statistical analyses. *The American Statistician* 63 (2):155–62. doi:10.1198/tast.2009.0030.
Koziol, A. 1987. An alternative formulation of Neyman's smooth goodness of fit tests under composite alternatives. *Metrika* 34 (1):17–24. doi:10.1007/BF02613126.
Kwon, O. K. 2019. Positive definite and unimodal regions for Gram-Charlier series. https://ssrn.com/abstract=2928665.
Lau, A., H. Lau, and J. Wingender. 1990. The distribution of stock returns: New evidence against the stable model. *Journal of Business & Economic Statistics* 8 (2):217–23. doi:10.2307/1391984.
Lin, W., and J. Zhang. 2020. The valid regions of Gram-Charlier densities. https://ssrn.com/abstract=3564891.
Matz, A. 1978. Maximum likelihood parameter estimation for the quartic exponential distribution. *Technometrics* 20 (4):475–84. doi:10.1080/00401706.1978.10489702.
McLachlan, G., and D. Peel. 2000. *Finite mixture models*. New York, NY: Wiley.
Neyman, J. 1937. Smooth test for goodness of fit. *Scandinavian Actuarial Journal* 1937 (3–4):149–99. doi:10.1080/03461238.1937.10404821.
Pearson, K. 1894. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London A* 185:71–114.
Pearson, E., and M. Stephens. 1964. The ratio of range to standard deviation in the same normal sample. *Biometrika* 51:484–7.
Poitras, G., and J. Heaney. 2015. Classical ergodicity and modern portfolio theory. *Chinese Journal of Mathematics* 2015:737905. doi:10.1155/2015/737905.2015.
Poitras, G. 2006. More on the correct omnibus test for normality. *Economics Letters* 90 (3):304–9. doi:10.1016/j.econlet.2005.08.016.
Romão, X., R. Delgado, and A. Costa. 2010. An empirical power comparison of univariate goodness-of-fit tests for normality. *Journal of Statistical Computation and Simulation* 80 (5):545–91. doi:10.1080/00949650902740824.
Rayner, J., O. Thas, O, and D. Best. 2009. *Smooth tests of goodness of fit: Using R*. 2nd ed. Singapore: Wiley.
Schilling, M., A. Watkins, and W. Watkins. 2002. Is human height bimodal? *The American Statistician* 56 (3):223–9. doi:10.1198/00031300265.
Schmitt, N., and F. Westerhoff. 2017. On the bimodality of the distribution of the S&P 500's distortion: Empirical evidence and theoretical explanations. *Journal of Economic Dynamics and Control* 80:34–53. doi:10.1016/j.jedc.2017.05.002.
Schlögl, E. 2013. Option pricing where the underlying assets follow a Gram/Charlier density of arbitrary order. *Journal of Economic Dynamics and Control* 37 (3):611–32. doi:10.1016/j.jedc.2012.10.001.
Shapiro, S., M. Wilk, and H. Chen. 1968. A comparative study of various tests for normality. *Journal of the American Statistical Association* 63 (324):1343–72. doi:10.1080/01621459.1968.10480932.
Silverman, B. 1986. *Density estimation*. London, UK: Chapman and Hall.
Spiring, F. 2011. The refined positive definite and unimodal regions for the Gram-Charlier and edgeworth series expansion. *Advances in Decision Sciences* 2011:1–18. doi:10.1155/2011/463097.
Thas, O. 2010. *Comparing distributions*. New York, NY: Springer.
Thode, H. 2002. *Testing for normality*. New York, NY: Marcel Dekker.
Thode, H., L. Smith, and S. Finch. 1983. Power of tests of normality for detecting scale contaminated normal samples. *Communications in Statistics - Simulation and Computation* 12 (6):675–95. doi:10.1080/03610918308812352.

Uhm, T., and S. Yi. 2021. A comparison of normality testing methods by empirical power and distribution of *P*-values. *Communications in Statistics-Simulation and Computation* 2021:1–14. doi:10.1080/03610918.2021.1963450.

Urzúa, C. 1996. On the correct use of omnibus tests for normality. *Economics Letters* 53 (3):247–51. doi:10.1016/S0165-1765(96)00923-8.

Uthoff, V. 1970. An optimum test property of two well-known statistics. *Journal of the American Statistical Association* 65 (332):1597–600. doi:10.2307/2284342.

Venegas, O., H. Salinas, D. Gallardo, H. Bolfarine, and H. Gómez. 2018. Bimodality based on the generalized skew-normal distribution. *Journal of Statistical Computation and Simulation* 88 (1):156–81. doi:10.1080/00949655.2017.1381698.

Vollmer, S. 2009. A likelihood ratio test for bimodality in two-component mixtures. Essay 1 in S. Vollmer, *A contribution to the empirics of economic and human development*. Frankfurt: Peter Lang.

Vollmer, S., H. Holzmann, F. Ketterer, and S. Klasen. 2013. Distribution dynamics of regional GDP per employee in unified Germany. *Empirical Economics* 44 (2):491–509. doi:10.1007/s00181-011-0543-3.

Wijekularathna, D., A. Manage, and S. Scariano. 2022. Power analysis of several normality tests: A Monte Carlo simulation study. *Communications in Statistics - Simulation and Computation* 51 (3):757–73. doi:10.1080/03610918.2019.1658780.