Adjusting the Tests for Skewness and Kurtosis
for Distributional Misspecifications

Anil K. Bera\textsuperscript{1} and G\textsuperscript{ani}ni Premaratne

\textit{Department of Economics, University of Illinois,}
\textit{1206 S. 6th Street, Champaign, IL 61820, USA.}

\textbf{Abstract}

The standard $\sqrt{b_1}$ test is widely used for testing skewness. However, several studies have demonstrated that this test is not reliable for discriminating between symmetric and asymmetric distributions in the presence of excess kurtosis. The main reason for the failure of the standard $\sqrt{b_1}$ test is that its variance formula is derived under the assumption of no excess kurtosis. In this paper we theoretically derive adjustment to the $\sqrt{b_1}$ test under the framework of Rao’s Score (or the Lagrange multiplier) test principle. Our adjusted test automatically correct the variance formula and does not lead to over- or under-rejection of the correct null hypothesis. In a similar way, we also suggest an adjusted test for kurtosis in the presence of asymmetry. These tests are then applied to both simulated and real data. The finite sample performances of the adjusted tests are far superior compared to those of their unadjusted counterparts.

\textsuperscript{1}Correspondence to: Anil K. Bera, Department of Economics, University of Illinois, 1206 S. 6th Street, Champaign, IL 61820, U.S.A.; email: a-bera@uiuc.edu
1 Introduction

Suppose we have \( n \) independent observations \( y_1, y_2, \ldots, y_n \) on a random variable \( Y \). For simplicity, assume that they are measured from their mean. Then the sample skewness and kurtosis are defined, respectively, as

\[
\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} \tag{1}
\]

and

\[
b_2 = \frac{m_4}{m_2^2} \tag{2}
\]

where \( m_j = n^{-1} \sum_{i=1}^{n} y_i^j \), \( j = 2, 3, 4 \). The use of these coefficients in the statistics literature goes back a long way. Pearson (1895) suggested that one use these coefficients to identify a density within a family of distributions. Fisher (1930) and Pearson (1930) formulated formal tests of normality based on \( \sqrt{b_1} \) and \( b_2 \) separately. Using the Pearson (1895) family of distributions and the Rao (1948) score (RS) test principle, Bera and Jarque (1981) derived the following omnibus statistic for testing normality

\[
RS = n \left[ \frac{\sqrt{b_1^2}}{6} + \frac{(b_2 - 3)^2}{24} \right], \tag{3}
\]

jointly based on \( \sqrt{b_1} \) and \( b_2 \) [see also D’Agostino and Pearson (1973), Cox and Hinkley (1974, p. 42), Bowman and Shenton (1975), and Jarque and Bera (1987)]. Under the null hypothesis of normality, RS is asymptotically distributed as \( \chi^2_2 \).

It is true that the skewness and kurtosis coefficients \( \sqrt{b_1} \) and \( b_2 \) jointly provide a powerful scheme for assessing normality against a wide variety of alternatives. In many practical applications, researchers are using \( n(\sqrt{b_1})^2/6 \) and \( n(b_2 - 3)^2/24 \) separately as the \( \chi^2_1 \) statistic for assessing asymmetry and excess kurtosis [for instance, see Affleck and McDonald (1989), Richardson and Smith (1993), and Christie-David and Chaudhury (2001)]. The purpose of
this paper is to demonstrate that \( n(\sqrt{b_1})^2/6 \) and \( n(b_2 - 3)^2/24 \) are not reliable measures of asymmetry and excess kurtosis, respectively, that can be used as proper test statistics. For instance, the asymptotic variance \( 6/n \) of \( \sqrt{b_1} \) is valid only under the assumption of normality, and the resulting test is not valid when the population kurtosis measure \( \beta_2 \neq 3 \). Similarly, the test statistics \( n(b_2 - 3)^2/24 \) will not provide a reliable inference under asymmetry. Bera and John (1983, p. 104) clearly stated that \( n(\sqrt{b_1})^2/6 \) and \( n(b_2 - 3)^2/24 \) are not pure tests of skewness and kurtosis, since the asymptotic distribution of these statistics is derived using full normality assumption. What is needed is to obtain the correct variances of \( \sqrt{b_1} \) and \( b_2 \) that are valid under distributional misspecification. In this paper we derive correct variance formulae using White's (1982) approach to make inference under misspecified models. To do this, we go back to the original derivation of (3) as given in Bera and Jarque (1981) and Jarque and Bera (1987), using the Pearson family of distribution, and apply White's (1982) modification.

The plan of this paper is as follows: In the next section we derive the Rao (1948) score and White’s (1982) modification to it. Section 3 discusses the adjusted RS test for asymmetry in the presence of excess kurtosis, followed by a finite sample performance study in Section 4. We do the same in sections 5 and 6 for the test of excess kurtosis allowing for asymmetry. Section 7 presents an empirical illustration. In the last section we offer a conclusion.

2 Rao’s (1948) Score Test and White’s (1982) Modification

Suppose there are \( n \) independent observations \( y_1, y_2, \ldots, y_n \) with identical density function \( f(y; \theta) \), where \( \theta \) is a \( p \times 1 \) parameter vector with \( \theta \in \Theta \subset \mathbb{R}^p \). It is assumed that \( f(y; \theta) \) satisfies the regularity condition stated in Rao (1973, p. 364) and Serfling (1980, p. 144). The log-likelihood function, the score function, and the information matrix are then defined,
respectively, as

\[ l(\theta) = \sum_{i=1}^{n} \ln f(y_i; \theta) \]  
\[ s(\theta) = \frac{\partial l(\theta)}{\partial \theta} \]  
\[ I(\theta) = -E \left[ \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'} \right]. \]

Let the hypothesis to be tested be \( H_0 : h(\theta) = 0 \), where \( h(\theta) \) is an \( r \times 1 \) vector function of \( \theta \) with \( r \leq p \). It is assumed that \( H(\theta) = \frac{\partial h(\theta)}{\partial \theta} \) has full rank, i.e., \( \text{rank}[H(\theta)] = r \). Rao’s (1948) score statistic for testing \( H_0 \) can be written as

\[ \text{RS} = s(\hat{\theta})^T I(\hat{\theta})^{-1} s(\hat{\theta}), \]

where \( \hat{\theta} \) is the restricted maximum likelihood estimator (MLE) of \( \theta \). Under \( H_0 \), RS is asymptotically distributed as a \( \chi^2 \). RS is not a valid test statistic when the true data generating process (DGP), \( g(y) \), differs from \( f(y; \theta) \). This is because some of the standard results break down under distributional misspecification. For instance, consider the information matrix equality

\[ E_f \left[ \frac{\partial \ln f(y; \theta)}{\partial \theta}, \frac{\partial \ln f(y; \theta)}{\partial \theta'} \right] = E_f \left[ -\frac{\partial^2 \ln f(y; \theta)}{\partial \theta \partial \theta'} \right], \]

in which \( E_f[\cdot] \) denotes expectation under \( f(y, \theta) \). Let us define

\[ J(\theta_g) = E_g \left[ \frac{\partial \ln f(y; \theta)}{\partial \theta} \cdot \frac{\partial \ln f(y; \theta)}{\partial \theta'} \right] \]
\[ K(\theta_g) = E_g \left[ -\frac{\partial^2 \ln f(y; \theta)}{\partial \theta \partial \theta'} \right], \]

where \( \theta_g \) minimizes the Kullback-Leibler information criterion [see White (1982)]

\[ I_{KL} = E_g \left[ \ln \frac{g(y)}{f(y; \theta)} \right]. \]
One can easily see that $J(\theta_g) \neq K(\theta_g)$, in general. Due to this divergence between $J$ and $K$, and taking expectation in (6) under $f(y, \theta)$ instead of the under of the DGP $g(y)$, in some cases the standard RS test in (7) is not valid. White (1982) suggested the following robust form of the RS statistic [see also Kent (1982)]:

$$RS^* = \frac{1}{n} s(\tilde{\theta})' K(\tilde{\theta})^{-1} H(\tilde{\theta}) H(\tilde{\theta})' K(\tilde{\theta})^{-1} s(\tilde{\theta}),$$  

(12)

where $B(\theta) = K(\theta)^{-1} J(\theta) K(\theta)^{-1}$ and $\tilde{\theta}$ denotes quasi MLE (QMLE). Under $H_0$, $RS^*$ is asymptotically distributed as $\chi^2_r$ even under distributional misspecification, i.e., when the assumed density $f(y; \theta)$ does not coincide with the true DGP $g(y)$.

For our purposes we need only a special case of $RS^*$. Let us partition $\theta$ as $\theta = (\theta_1', \theta_2')'$, where $\theta_1$ and $\theta_2$ are respectively, $(p-r) \times 1$ and $r \times 1$ vectors. Let $H_0 : \theta_2 = \theta_{20}$, where $\theta_{20}$ is a known quantity. Let $\tilde{\theta}_1$ be the QMLE under $H_0$, and let $\tilde{\theta} = (\tilde{\theta}_1', \theta_{20}')'$. We partition the score vector and other matrices as follows:

$$s(\theta) \equiv s(\theta_1, \theta_2) = [s_1(\theta)' , s_2(\theta)']',$n

$$I(\theta) = \begin{bmatrix} I_{11}(\theta) & I_{12}(\theta) \\ I_{21}(\theta) & I_{22}(\theta) \end{bmatrix},$$

$$I^{-1}(\theta) = \begin{bmatrix} I^{11}(\theta) & I^{12}(\theta) \\ I^{21}(\theta) & I^{22}(\theta) \end{bmatrix},$$  

(13)

and so on. Under these notations the standard RS statistic in (7) for our special case is given by

$$RS = s_2(\tilde{\theta}_1, \theta_{20})' I^{22}(\tilde{\theta}) s_2(\tilde{\theta}_1, \theta_{20}).$$

This can also be written as

$$RS = s_2(\tilde{\theta}_1, \theta_{20})' K^{22}(\tilde{\theta}) s_2(\tilde{\theta}_1, \theta_{20}) = \tilde{s}_2' K^{22} \tilde{s}_2$$  

(14)
after dropping the arguments. Now let us consider $RS^*$ in (12). Here $h(\theta) = \theta_2 - \theta_{x_0}$, $s(\theta) = (0, s_2(\tilde{\theta})')$, and $H(\theta) = \frac{\partial h(\theta)}{\partial \theta} = [0_r \times p_r I_r]$. Hence,

$$s(\tilde{\theta})' K^{-1}(\tilde{\theta}) H(\tilde{\theta})' = \left( \begin{array}{c} 0' \\ s_2(\tilde{\theta})' \end{array} \right) \left( \begin{array}{cc} \tilde{K}^{11} & \tilde{K}^{12} \\ \tilde{K}^{21} & \tilde{K}^{22} \end{array} \right) \left( \begin{array}{c} 0 \\ I_r \end{array} \right)$$

$$= s_2(\tilde{\theta})' \tilde{K}^{22}. \quad (15)$$

We again drop the arguments of $K^{-1}$ and other matrices whenever convenient, and "\(\hat{\cdot}\)" denotes that the quantity is evaluated at $\theta = \hat{\theta}$. Similarly,

$$H \tilde{B} H' = \left( \begin{array}{c} 0 \\ I_r \end{array} \right) \left( \begin{array}{cc} \hat{B}_{11} & \hat{B}_{12} \\ \hat{B}_{21} & \hat{B}_{22} \end{array} \right) \left( \begin{array}{c} 0 \\ I_r \end{array} \right)$$

$$= \hat{B}_{22}. \quad (16)$$

Hence,

$$RS^* = \frac{1}{n} \hat{s'}_2 \hat{K}^{22} \hat{B}_{22} \hat{K}^{22} \hat{s}_2. \quad (17)$$

Note that $K^{22} = (K_{22} - K_{21} K_{12})^{-1}$. In one of our special cases, as will be seen later, $K_{21} = 0$, i.e., $K^{22} = K_{22}^{-1}$. We can then simplify $B$ as

$$B = \left( \begin{array}{cc} K^{11} & 0 \\ 0 & K^{22} \end{array} \right) \left( \begin{array}{cc} J_{11} & J_{12} \\ J_{21} & J_{22} \end{array} \right) \left( \begin{array}{cc} K^{11} & 0 \\ 0 & K^{22} \end{array} \right)$$

$$= \left( \begin{array}{cc} K^{11} J_{11} K^{11} & K^{11} J_{12} K^{22} \\ K^{22} J_{21} K^{11} & K^{22} J_{22} K^{22} \end{array} \right), \quad (18)$$

and hence, $B_{22} = K^{22} J_{22} K^{22} = K_{22}^{-1} J_{22} K_{22}^{-1}$. Therefore, from (17),

$$RS^* = \frac{1}{n} \hat{s'}_2 \hat{K}^{-1}_{22} \hat{J}_{22} \hat{K}^{-1}_{22} \hat{s}_2$$

$$= \frac{1}{n} \hat{s'}_2 \hat{J}_{22} \hat{s}_2. \quad (20)$$

6
Therefore, when the matrix $K(\theta)$ is block diagonal, i.e., $K_{12} = 0$, the robust version of the RS statistic requires calculation of $J_{22}(\theta)$ only. Also, comparing (14) and (20) we see that RS and $RS^*$ have similar forms except that the former uses the Hessian, whereas the latter uses the outer product gradient form for the underlying information matrix.

3 An Adjusted Test for Asymmetry

Let us start with the Pearson (1895) system of distributions,

$$\frac{d \ln f(y)}{dy} = \frac{c_1 - y}{c_0 - c_1 y + c_2 y^2},$$

(21)

where we write $f(y; \theta)$ simply as $f(y)$ with $\theta = (c_0, c_1, c_2)'$. The normal distribution is a special case of this when $c_1 = c_2 = 0$. As discussed in Premaratne and Bera (2000), $c_1$ and $c_2$ could be treated, respectively, as the “asymmetry” and “kurtosis” parameters. Suppose we are interested only in testing asymmetry ignoring excess kurtosis. Then we can start with the above system with $c_2 = 0$, i.e.,

$$\frac{d \ln f(y)}{dy} = \frac{c_1 - y}{c_0 - c_1 y}.$$  

(22)

By integrating (22), it can be shown that (22) leads to gamma density, which is a skewed distribution. Under $c_1 = 0$, (22) becomes the normal density with mean zero and variance $c_0$, which is symmetric. Let us derive the Rao score test for $c_1 = 0$ in (22). That will give us a test for asymmetry without allowing for excess kurtosis. Let us denote

$$\psi(\theta; y) = \int \frac{c_1 - y}{c_0 - c_1 y} dy = \int \frac{c_1 - y}{\nu} dy,$$

with $\theta = (c_0, c_1)'$ and $\nu = c_0 - c_1 y$. Then by integrating (22) for the $i^{th}$ observation, we have

$$\ln f(y_i) = \text{const.} \psi(\theta; y_i)$$

i.e.,

$$f(y_i) = \text{const.} \exp(\psi(\theta; y_i))$$
or
\[ f(y_i) = \frac{\exp(\psi(\theta; y_i))}{\int_{-\infty}^{\infty} \exp(\psi(\theta; y)) \, dy}. \]  
(23)

Note that in the denominator of (23), we don’t use a subscript for \( y \), since this term is purely a constant. Therefore, the log-likelihood function \( l(\theta) = \ln \prod_{i=1}^{n} f(y_i) \) can be written as
\[
l(\theta) = -n \ln \left[ \int_{-\infty}^{\infty} \exp(\psi(\theta; y)) \, dy \right] + \sum_{i=1}^{n} \psi(\theta; y_i) 
= l_1(\theta) + l_2(\theta) \quad \text{(say)}. \]  
(24)

To get the Rao score test, let us first derive the score functions \( \frac{\partial l(\theta)}{\partial c_0} \) and \( \frac{\partial l(\theta)}{\partial c_1} \) under the null hypothesis \( H_0 : c_1 = 0 \). One can easily see that
\[
\frac{\partial l_1(\theta)}{\partial c_1} \bigg|_{c_1=0} = 0, \tag{25}
\]
\[
\frac{\partial l_2(\theta)}{\partial c_1} = \sum_{i=1}^{n} \int \frac{\nu_i - (c_1 - y_i)(-y_i)}{\nu_i^2} \, dy_i \tag{26}
\]
That is,
\[
\frac{\partial l_2(\theta)}{\partial c_1} \bigg|_{c_1=0} = \sum_{i=1}^{n} \int \frac{c_0 y_i^2}{c_0^2} \, dy_i 
= \sum_{i=1}^{n} \left( \frac{y_i}{c_0} - \frac{y_i^3}{3c_0^3} \right). \tag{27}
\]
When evaluated at the restricted MLE \( \tilde{\theta} = (\tilde{c}_0, 0) \), this becomes
\[
\frac{\partial l_2(\tilde{\theta})}{\partial c_1} = \frac{nm_1}{m_2} - \frac{nm_3}{3m_2^2},
\]
where \( m_j = \sum_{i=1}^{n} y_i^j / n, j = 1, 2, 3 \), and \( \tilde{c}_0 = m_2 \) under the null. Therefore, we have
\[
\frac{\partial l(\tilde{\theta})}{\partial c_1} = \frac{\partial l_1(\tilde{\theta})}{\partial c_1} + \frac{\partial l_2(\tilde{\theta})}{\partial c_2} = \frac{nm_1}{m_2} - \frac{nm_3}{3m_2^2}. \tag{28}
\]

To get the information matrix \( I(\theta) \) we use the \( J \) matrix where we take the expectation with respect to the assumed density \( f(y) \), and denote it by, \( J_f \), i.e.,
\[
J_f = E_f \left[ \frac{\partial \ln f(y; \theta)}{\partial \theta} \cdot \frac{\partial \ln f(y; \theta)}{\partial \theta'} \right]. \tag{29}
\]
From (23), we can easily see that, under $H_0 : c_1 = 0$,

$$\frac{\partial \ln f(y)}{\partial c_0} = -\frac{\mu_2}{2c_0^2} + \frac{y^2}{2c_0^2} = -\frac{1}{2\mu_2} + \frac{y^2}{2\mu_2^2}$$  \hspace{1cm} (30)

$$\frac{\partial \ln f(y)}{\partial c_1} = \frac{y}{c_0} - \frac{y^3}{3c_0^2} = \frac{y}{\mu_2} + \frac{y^3}{3\mu_2^2}$$  \hspace{1cm} (31)

where we replace $c_0$ by $\mu_2 = E_f(y^2)$ with the density $f(y)$ being $N(0, c_0)$. Using (30) and (31), the $J_f$ matrix of (29) can be calculated as follows:

$$J_f = E_f \left[ \begin{array}{cc} \frac{\partial \ln f}{\partial c_0} & \frac{\partial \ln f}{\partial c_1} \\ \frac{\partial \ln f}{\partial c_1} & \frac{\partial \ln f}{\partial c_1} \end{array} \right]$$

$$= E_f \left[ \begin{array}{cc} \frac{1}{4\mu_2^2} + \frac{y^4}{4\mu_2^2} - \frac{y^2}{2\mu_2^2} & -\frac{y}{2\mu_2} + \frac{y^3}{6\mu_2} - \frac{y^5}{6\mu_2^3} \\ \frac{y}{\mu_2} + \frac{y^3}{3\mu_2^2} - \frac{y^5}{3\mu_2^3} & \frac{y^2}{\mu_2} - \frac{y^4}{3\mu_2^2} \end{array} \right].$$  \hspace{1cm} (32)

Under symmetry, $E_f(y) = E_f(y^3) = E_f(y^5) = 0$, and under $H_0 : c_1 = 0, E_f(y^2) = \mu_2, E_f(y^4) = 3\mu_2^2, E_f(y^6) = 15\mu_2^3$. Therefore,

$$J_f = \left[ \begin{array}{ccc} \frac{1}{4\mu_2^2} + \frac{3\mu_2^3}{4\mu_2^2} - \frac{\mu_2^2}{2\mu_2^2} & 0 \\ 0 & \frac{\mu_2^3}{\mu_2^2} + \frac{15\mu_2^4}{9\mu_2^3} - \frac{6\mu_2^5}{15\mu_2^4} \\ 0 & 0 & \frac{2}{3\mu_2} \end{array} \right].$$  \hspace{1cm} (33)

From (28) we have the estimated score function $s_2(\hat{\theta})$ under the null hypothesis as

$$s_2(\hat{\theta}) = \frac{\partial \ln(\hat{\theta})}{\partial c_1} = -\frac{nm_3}{3m_2^2}$$  \hspace{1cm} (34)

by putting the sample mean $m_1 = 0$. Since the information matrix in (33) is block diagonal, the standard score test is
\[ RS_{c_1} = \frac{n}{9} m_2^3 \frac{3}{\mu_2} \]
\[ = \frac{n}{6} \left( \frac{m_3}{m_2^{3/2}} \right)^2 \]
\[ = \frac{n}{6} (\sqrt{b_1})^2. \]  

(35)

Note that while deriving this test, we have taken the density (23) as the DGP and have fully utilized it while taking the expectation in (32). If \( f(y) \) is not the DGP, the test \( RS_{c_1} \) will not be valid; in particular, the asymptotic variance formula in (35), \( V(\sqrt{b_1}) = \frac{6}{n} \) is not correct. For instance, in the presence of excess kurtosis, there will be proportionately more outliers. As Pearson (1963) showed, as we move from short-tailed to very long-tailed distributions, the contribution to the moments will come increasingly from the extreme tails of the frequency function. As a result, the variance of a statistic such as \( \sqrt{b_1} \) will increase, and for fat-tailed distributions, \( \frac{6}{n} \) will underestimate the variance.

Let us try to use the modified score statistic in (17). For that we need the \( J_g \) and \( K_g \) matrices. As we noted earlier, if \( K_{12} = 0 \), then the modified score test takes a very simple form (20). After some tedious algebra, we can see that

\[ K_{12} = E_g \left[ \frac{\partial^2 \ln f}{\partial c_0 \partial c_1} \right] = 0. \]  

(36)

To calculate \( RS^* \) in (20), we need to find

\[ J_{22} = E_g \left[ \frac{\partial \ln f}{\partial c_1} \right]^2. \]  

(37)

From our earlier derivation,

\[ \frac{\partial \ln f}{\partial c_1} = \frac{y}{c_0} - \frac{y^3}{3c_0^2}. \]

Hence,

\[ \left( \frac{\partial \ln f}{\partial c_1} \right)^2 = \frac{y^2}{c_0^2} + \frac{y^6}{9c_0^4} - \frac{2y^4}{3c_0^2}, \]
and

\[ E_g \left( \frac{\partial \ln f}{\partial c_1} \right)^2 = \frac{\mu_2}{c_0^2} + \frac{\mu_6}{9c_0^4} - \frac{2\mu_4}{3c_0^6} \]

\[ = \frac{1}{\mu_2} + \frac{\mu_6}{9\mu_2^4} - \frac{2\mu_4}{3\mu_2^6}. \]

Therefore,

\[ J_{22} = \frac{1}{m_2} + \frac{m_6}{9m_2^4} - \frac{2m_4}{3m_2^6}. \] (38)

Using (34) and (38), \( RS^* \) in (20) for testing \( H_0 : c_1 = 0 \) can be expressed as

\[ RS^*_{c_1} = \frac{n}{9 \left[ m_2^{-1} + \frac{1}{9}m_6m_2^{-4} - \frac{2}{3}m_4m_2^{-3} \right]} \]

\[ = \frac{n}{9 + m_6m_2^{-3} - 6m_4m_2^{-2}} \]

\[ = \frac{(\sqrt{b_1})^2}{9 + m_6m_2^{-3} - 6m_4m_2^{-2}}. \] (39)

Note that if we impose the normality assumption, then the population counterpart of the denominator in (39) is

\[ 9 + m_6m_2^{-3} - 6m_4m_2^{-2} = 9 + 15 - 6.3 = 9 + 15 - 18 = 6, \]

as in (35). Basically, the construction of the adjusted statistic \( RS^*_{c_1} \) indicates that asymptotically, an estimate of the variance of the \( \sqrt{b_1} \) that is valid under excess kurtosis is

\[ \frac{1}{n} [9 + m_6m_2^{-3} - 6m_4m_2^{-2}] . \] (40)

Let us again consider the term

\[ 9 + m_6m_2^{-3} - 6m_4m_2^{-2} . \] (41)

It differs from the variance formula under normality by

\[ 9 + m_6m_2^{-3} - 6m_4m_2^{-2} - 6 = 3 + m_6m_2^{-3} - 6m_4m_2^{-2} . \] (42)
Let us define the generalized notion of kurtosis as

$$\beta_{2r} = \frac{\mu_{2r+2}}{\mu_2^r}, \quad r = 1, 2, \ldots$$

so that

$$\beta_2 = \frac{\mu_4}{\mu_2^2}, \quad \beta_4 = \frac{\mu_6}{\mu_2^3}.$$ 

Then (42) is

$$3 + \beta_4 - 6\beta_2.$$ 

It can be shown that for leptokurtic distributions, $3 + \beta_4 - 6\beta_2 \geq 0$. For example, with $t_7$ density we have $\mu_2 = 7/5$, $\mu_4 = 49/5$, $\mu_6 = 343$; then $3 + \beta_4 - 6\beta_2 = 98$. Therefore, the usual variance formula will underestimate the variance of $\sqrt{b_1}$, and we will reject the null hypothesis of symmetry too frequently in the presence of excess kurtosis.

We can obtain the formula (40) or its population counterpart

$$\frac{1}{n}(9 + \mu_6 \mu_2^{-3} - 6\mu_4 \mu_2^{-2})$$

in an alternative way.

Let $g(w) = g(w_1, w_2, \ldots, w_k)$ be a continuous function of random variables $w_1, w_2, w_3, \ldots, w_k$. A general formula for the asymptotic variance of $g(w)$ is [see, for example, Stuart and Ord (1994, p. 350)]

$$V[g(w)] = \sum_{i=1}^{k} g'_i(w) V(w_i) + \sum_{i \neq j}^{k} g'_i(w)g'_j(w).Cov(w_i, w_j) + o(n^{-1}),$$

(44)

where $g'_i(w) = \frac{\partial g(w)}{\partial w_i}$, $i = 1, 2, \ldots, k$ are in terms of population values. We take $g(w_1, w_2) = m_3/m_2^{3/2}$. Hence using (44),

$$V(\sqrt{b_1}) = \left(\frac{\partial g}{\partial m_3}\right)^2 V(m_2) + \left(\frac{\partial g}{\partial m_2}\right)^2 V(m_3) + 2 \left(\frac{\partial^2 g}{\partial m_3 \partial m_2}\right) Cov(m_2, m_3).$$

Under symmetry, we utilize $\mu_3 = \mu_5 = 0$ and have

$$\frac{\partial g}{\partial m_3} = \frac{1}{\mu_2^{3/2}}$$

12
\[
\frac{\partial g}{\partial m_2} = -\frac{3\mu_3}{2\mu_2^{5/2}} = 0
\]
\[
\frac{\partial^2 g}{\partial m_3 \partial m_2} = -\frac{3}{2\mu_2^{5/2}}.
\]

Hence,
\[V(m_3) = \frac{1}{n}(\mu_6 - \mu_3^2 + 9\mu_2^3 - 6\mu_2\mu_4)\]
\[= \frac{1}{n}(\mu_6 + 9\mu_2^3 - 6\mu_2\mu_4)\]
\[Cov(m_2, m_3) = \frac{1}{n}(\mu_5 - \mu_2\mu_3 + 6\mu_2\mu_1\mu_2 - 2\mu_1\mu_4 - 3\mu_3\mu_2)\]
\[= 0.\]

Therefore,
\[V(\sqrt{b_1}) = \left( \frac{1}{\mu_2^{3/2}} \right)^2 \frac{1}{n}(\mu_6 + 9\mu_2^3 - 6\mu_2\mu_4)\]
\[= \frac{1}{n}(9 + \mu_6\mu_2^{-3} - 6\mu_4\mu_2^{-2}),\] (45)

which is same as (43). Godfrey and Orme (1991) used this approach to develop their test for skewness, which is asymptotically valid under non-normality.

4 Comparison of the Standard and Adjusted Tests for Asymmetry: Some Simulation Results

For our simulation study, we generated data under a variety of scenarios. To study the size properties, we first generated data under the normal distribution, for which the standard \(\sqrt{b_1}\) provides the ideal test. We used the tabulated critical values for the standard \(\sqrt{b_1}\) test [see Pearson and Hartley (1976, p. 183)]. Some of these critical values are interpolated. For our adjusted test \(RS^*_c\) in (39), we used the asymptotic \(\chi_1^2\) critical values. The standard test is somewhat oversized whereas the adjusted test has a size very close to the nominal levels
of 1% and 5%. In fact, the sizes are even lower. This may due to the fact that here we are using an “unnecessary” adjustment [see Equations (42) and (45)] to the variance of $\sqrt{b_t}$ and thereby overestimate the variance which leads to under-rejection. The estimated sizes of the modified test look better as we increase the sample size.

The next results are for Student’s $t_7$-distribution. As explained earlier, for this case, the standard test grossly underestimates the variance, leading to a false rejection of asymmetry too often. The quantity in (43) is $(\nu + \mu_4 \mu_5^2 - 6 \mu_4 \mu_2^2) / \sigma^2 = 104 / n$, which is much larger than the standard value 6/\sigma^2. The rejection probabilities for our adjusted test are again close to the nominal levels 1 and 5% and are in fact, lower than those values. As we noticed for the data generated under normality, the rejection probabilities, in general, get closer to the nominal levels as the sample size increases. The following DGP Beta(2,2) is a platykurtic (i.e., $\beta_2 < 3$) symmetric distribution. Here, possibly 6/\sigma^2 overestimates the variance, and the standard test has much lower estimated sizes; the adjusted test also have lower sizes, but these are much closer to the 1 and 5% values. We then generated data using the procedure of Ramberg, Dudewicz, Tadikamalla, and Mykytka (1979). Their technique can generate data for any given level of skewness and kurtosis. We generated data using $\sqrt{\beta_1} = 0$, $\beta_2 = 7.0$. The results for this distribution are qualitatively similar to those obtained for the $t_7$ distribution. Since for this distribution the excess kurtosis is larger than the $t_7$-density, the performance of the standard test is even worse whereas the behavior of the modified test does not change. The last two DGPs are from symmetric leptokurtic distributions, namely, Laplace and logistic distributions. Again, the standard test rejects the correct hypothesis of symmetry too often whereas our modification corrects this over-rejection.

Table 2 provides some simulation results for the power of the tests. The first two DGPs are under positively skewed distributions ($\chi_4^2$ and Beta(1,2)), and the estimated probabilities are the estimated powers. Both the standard and the modified tests have very good power. For
Table 1: Estimated Sizes of the Skewness Tests with 5000 Replications

<table>
<thead>
<tr>
<th>DGP and Sample Sizes</th>
<th>Standard Test 1%</th>
<th>Standard Test 5%</th>
<th>Modified Test 1%</th>
<th>Modified Test 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP: N(0,1) ($\sqrt{\beta_1} = 0, \beta_2 = 3.0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>2.12</td>
<td>9.92</td>
<td>0.18</td>
<td>3.66</td>
</tr>
<tr>
<td>100</td>
<td>2.00</td>
<td>9.90</td>
<td>0.58</td>
<td>4.58</td>
</tr>
<tr>
<td>200</td>
<td>1.70</td>
<td>10.14</td>
<td>0.50</td>
<td>4.66</td>
</tr>
<tr>
<td>250</td>
<td>2.16</td>
<td>10.38</td>
<td>0.70</td>
<td>4.74</td>
</tr>
<tr>
<td>DGP: $t_7$($\sqrt{\beta_1} = 0, \beta_2 = 5.0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>16.18</td>
<td>31.18</td>
<td>0.26</td>
<td>3.42</td>
</tr>
<tr>
<td>100</td>
<td>20.28</td>
<td>35.70</td>
<td>0.42</td>
<td>3.42</td>
</tr>
<tr>
<td>200</td>
<td>25.86</td>
<td>41.70</td>
<td>0.44</td>
<td>3.64</td>
</tr>
<tr>
<td>250</td>
<td>27.06</td>
<td>42.80</td>
<td>0.38</td>
<td>4.10</td>
</tr>
<tr>
<td>DGP: Beta(2,2) ($\sqrt{\beta_1} = 0, \beta_2 = 2.14$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.22</td>
<td>3.30</td>
<td>0.42</td>
<td>3.94</td>
</tr>
<tr>
<td>100</td>
<td>0.16</td>
<td>2.72</td>
<td>0.50</td>
<td>4.38</td>
</tr>
<tr>
<td>200</td>
<td>0.04</td>
<td>2.26</td>
<td>0.74</td>
<td>4.28</td>
</tr>
<tr>
<td>250</td>
<td>0.14</td>
<td>2.26</td>
<td>0.86</td>
<td>4.86</td>
</tr>
<tr>
<td>DGP: Generated Data ($\sqrt{\beta_1} = 0, \beta_2 = 7.0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>26.2</td>
<td>42.78</td>
<td>0.32</td>
<td>3.22</td>
</tr>
<tr>
<td>100</td>
<td>34.2</td>
<td>50.82</td>
<td>0.24</td>
<td>3.28</td>
</tr>
<tr>
<td>200</td>
<td>43.2</td>
<td>57.80</td>
<td>0.36</td>
<td>3.18</td>
</tr>
<tr>
<td>250</td>
<td>43.5</td>
<td>58.74</td>
<td>0.30</td>
<td>3.66</td>
</tr>
<tr>
<td>DGP: Laplace($\lambda = 2$) ($\sqrt{\beta_1} = 0, \beta_2 = 6.0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25.72</td>
<td>43.54</td>
<td>0.38</td>
<td>3.82</td>
</tr>
<tr>
<td>100</td>
<td>31.40</td>
<td>48.78</td>
<td>0.50</td>
<td>3.68</td>
</tr>
<tr>
<td>200</td>
<td>36.26</td>
<td>52.00</td>
<td>0.38</td>
<td>3.94</td>
</tr>
<tr>
<td>250</td>
<td>38.60</td>
<td>53.92</td>
<td>0.28</td>
<td>3.62</td>
</tr>
<tr>
<td>DGP: Logistic(1,2)($\sqrt{\beta_1} = 0, \beta_2 = 4.2$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>12.02</td>
<td>27.14</td>
<td>0.32</td>
<td>3.40</td>
</tr>
<tr>
<td>100</td>
<td>14.88</td>
<td>30.14</td>
<td>0.36</td>
<td>4.00</td>
</tr>
<tr>
<td>200</td>
<td>17.92</td>
<td>33.54</td>
<td>0.42</td>
<td>4.30</td>
</tr>
<tr>
<td>250</td>
<td>19.70</td>
<td>36.72</td>
<td>0.50</td>
<td>3.94</td>
</tr>
</tbody>
</table>
Table 2: Estimated Powers of Skewness Tests with 5000 Replications

<table>
<thead>
<tr>
<th>DGP and Sample Sizes</th>
<th>Standard Test</th>
<th>Modified Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>DGP: $\chi^2$ ($\sqrt{\beta_1} = 1.41, \beta_2 = 6$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>71.02</td>
<td>90.38</td>
</tr>
<tr>
<td>100</td>
<td>98.80</td>
<td>99.90</td>
</tr>
<tr>
<td>200</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>250</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>DGP: Beta(1,2) ($\sqrt{\beta_1} = 0.56, \beta_2 = 3.27$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>13.62</td>
<td>51.72</td>
</tr>
<tr>
<td>100</td>
<td>46.88</td>
<td>85.66</td>
</tr>
<tr>
<td>200</td>
<td>92.00</td>
<td>99.36</td>
</tr>
<tr>
<td>250</td>
<td>98.08</td>
<td>99.94</td>
</tr>
<tr>
<td>DGP: Beta(2,1) ($\sqrt{\beta_1} = -0.56, \beta_2 = 0.67$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>14.52</td>
<td>51.78</td>
</tr>
<tr>
<td>100</td>
<td>45.62</td>
<td>85.10</td>
</tr>
<tr>
<td>200</td>
<td>92.10</td>
<td>99.48</td>
</tr>
<tr>
<td>250</td>
<td>98.32</td>
<td>99.92</td>
</tr>
</tbody>
</table>

the $\chi^2$-data, the standard test has higher estimated powers, whereas for the Beta(1,2) data, the modified test performs better. For the negatively skewed distribution $B(2,1)$, the results are very similar to those for Beta(1,2) although their kurtosis structures are quite different.

These results reinforce our assertion that the standard $\sqrt{\beta_1}$ test is not a reliable test for asymmetry in the presence of excess kurtosis. We will wrongly reject the true null of symmetry too often. On the other hand, our simple variance-adjusted test work remarkably well. It has very good finite sample size and power properties.
5 An Adjusted Test for Excess Kurtosis

Here we start with,

\[
\frac{d \ln f(y)}{dy} = \frac{-y}{c_0 + c_2y^2},
\]

which is the Pearson density [equation (21)] with \(c_1 = 0\). By integrating (46), we get

\[
\ln f(y) = \text{const.} \ln(c_0 + c_2y^2)(-\frac{1}{2c_2}),
\]

i.e.,

\[
f(y) = \text{const.}(c_0 + c_2y^2)^{-\frac{1}{2c_2}},
\]

which is a generalization of Student’s t-density. The normal distribution (no excess kurtosis with \(\beta_2 = 3\)) is a special case of this, with \(c_2 = 0\). Let us first derive the standard RS test for \(H_0 : c_2 = 0\). We have

\[
f(y) = \frac{\exp(\psi(\theta; y_i))}{\int_{-\infty}^{\infty} \exp(\psi(\theta; y)) \, dy},
\]

where we now define \(\psi(\theta; y) = \int_{c_0 + c_2y}^{\infty} \frac{d}{dx} \exp(\psi(\theta; x)) \, dx\) with \(\theta = (c_0, c_2)'\). The log-likelihood function can be written as,

\[
l(c_0, c_2) = -n \ln \left[ \int_{-\infty}^{\infty} \exp(\psi(\theta; y)) \, dy \right] + \sum_{i=1}^{n} \psi(\theta; y_i) \\
= l_1(\theta) + l_2(\theta) \quad \text{(say)}. \quad (48)
\]

Under \(c_2 = 0\),

\[
\frac{\partial l_1(\theta)}{\partial c_2} = -n \int_{-\infty}^{\infty} e^{-\frac{y^2}{2c_0}} \frac{y^4}{c_2^3} \, dy \\
= -\frac{1}{4c_0^2} E(y^4) = -\frac{n\mu_4}{4c_0^2} \\
\frac{\partial l_2(\theta)}{\partial c_2} = \sum_{i=1}^{n} \int_{c_0}^{\infty} \frac{y^3}{c_2^2} \, dy = \frac{\sum_{i=1}^{n} y_i^4}{4c_0^2} = \frac{nm_4}{4c_0^2}.
\]
Hence,
\[ \frac{\partial l(\hat{\theta})}{\partial c_2} = \frac{n}{4m_2^2}(m_4 - 3m_2^2) = \frac{n}{4} \left( \frac{m_4}{m_2^2} - 3 \right) = \frac{n}{4}(b_2 - 3). \] (49)

To obtain the information matrix let us now derive
\[ J_g = E_g \left( \left( \frac{\partial \ln f}{\partial \psi_0} \right)^2 \left( \frac{\partial \ln f}{\partial \psi_2} \right)^2 \right). \] (50)

From (47)
\[ \ln f(y) = -\psi(\theta; y) - \ln \int_{-\infty}^{\infty} \exp(\psi(\theta; y)) dy \] (51)
and hence
\[ \frac{\partial \ln f(y)}{\partial c_0} \bigg|_{c_2 = 0} = \int \frac{y^2}{c_0^2} dy - \frac{\int_{-\infty}^{\infty} \exp(-\frac{y^2}{2c_0^2}) \frac{y^2}{c_0^2} dy}{\int_{-\infty}^{\infty} \exp(-\frac{y^2}{2c_0^2}) dy} \]
\[ = \frac{y^2}{2c_0^2} - \frac{\mu_2}{2c_0^2}, \] (52)
and
\[ \frac{\partial \ln f(y)}{\partial c_2} \bigg|_{c_2 = 0} = \int \frac{y^4}{c_0^2} dy - \frac{\int_{-\infty}^{\infty} \exp(-\frac{y^2}{2c_0^2}) \frac{y^4}{c_0^2} dy}{\int_{-\infty}^{\infty} \exp(-\frac{y^2}{2c_0^2}) dy} \]
\[ = \frac{y^4}{4c_0^2} - \frac{\mu_4}{4c_0^2}. \] (53)

Now
\[ E_g \left( \frac{\partial \ln f(y)}{\partial c_0} \right)^2 = E_g \left( \frac{y^2}{2c_0^2} - \frac{\mu_2}{2c_0^2} \right)^2 \]
\[ = \frac{1}{4c_0^2} \left( \mu_4 + \mu_2^2 - 2\mu_2y^2 \right) \]
\[ = \frac{1}{4\mu_2^2} (\mu_4 + \mu_2^2 - 2\mu_2^2) \]
\[ = \frac{1}{4\mu_2^2} (\mu_4 - \mu_2^2), \] (54)
\begin{align*}
E_g \left( \frac{\partial \ln f(y)}{\partial c_2} \right)^2 & = \frac{1}{16c_0} E_g (y^8 + \mu_4^2 - 2\mu_4 y^4) \\
& = \frac{1}{16\mu_4^4} (\mu_8 - \mu_4^2), \quad (55)
\end{align*}

\begin{align*}
E_g \left[ \frac{\partial \ln f(y)}{\partial c_0} \cdot \frac{\partial \ln f(y)}{\partial c_2} \right] & = \frac{1}{8c_0} E_g [(y^2 - \mu_2)(y^4 - \mu_4)] \\
& = \frac{1}{8\mu_2^2} \left(\mu_6 - \mu_4 \mu_2 \right) + \frac{1}{8\mu_2^4} \left(\mu_8 - \mu_4^2 \right), \quad (56)
\end{align*}

Therefore,

\[ J_g = \begin{bmatrix}
\frac{1}{4\mu_2^2} (\mu_4 - \mu_2^2) \\
\frac{1}{8\mu_2^4} (\mu_6 - \mu_4 \mu_2) \\
\frac{1}{16\mu_2^4} (\mu_8 - \mu_4^2)
\end{bmatrix}. \quad (57)

Evaluating \( J \) under the density \( f(y) \equiv N(0, c_0) \), we have

\begin{align*}
J_f & = \begin{bmatrix}
\frac{1}{2\mu_2} \\
\frac{1}{8\mu_2^3} (15 - 3) \mu_2^3 \\
\frac{1}{16\mu_2^4} (105 - 9) \mu_2^4 \\
\frac{1}{2\mu_2} \\
\frac{3}{2\mu_2} \\
\frac{3}{2\mu_2}
\end{bmatrix} \\
& = \begin{bmatrix}
\frac{1}{2\mu_2} \\
\frac{3}{2\mu_2} \\
\frac{3}{2\mu_2}
\end{bmatrix}, \\
\end{align*} \quad (58)

and

\[ J_f^{22} = \frac{3}{\mu_2^2} - \frac{9}{4\mu_2^4} = \frac{1}{2\mu_2^2} \times \frac{4\mu_2^2}{3} = \frac{2}{3}. \quad (59)\]

Combining (49) and (59), the standard RS statistic is

\[ RSe_2 = \frac{n}{16} (b_2 - 3)^2 \times \frac{2}{3} = n \frac{(b_2 - 3)^2}{24}. \quad (60)\]

To derive the adjusted RS statistic, we need the \( K \) matrix,
\[ K_g = -E_g \left[ \frac{\partial^2 \ln f}{\partial c_0^2}, \frac{\partial^2 \ln f}{\partial c_0 \partial c_2} \right]. \] 

(61)

After some algebra, it can be shown that

\[ E_g \left[ -\frac{\partial^2 \ln f}{\partial c_0^2} \right] = 2\mu_2 + \frac{\mu_4}{4\mu_2^4} - \frac{\mu_2^2}{4\mu_2} - \frac{2\mu_2}{2\mu_2^3} \]

\[ = \frac{\mu_4 - \mu_2^2}{4\mu_2^4}. \] 

(62)

Note that this \( K_{11} \) term is the same as \( J_{11} \) in (57). Under \( H_0 \), we have \( \mu_4 = 3\mu_2^2 \), and this quantity reduces to

\[ E_g \left[ -\frac{\partial^2 \ln f}{\partial c_0^2} \right] = \frac{3}{4\mu_2^2} - \frac{1}{4\mu_2^2} = \frac{1}{2\mu_2^2}. \] 

(63)

Since

\[ \left. \frac{\partial^2 \ln f}{\partial c_2^2} \right|_{c_2=0} = -\frac{g^6}{3c_0^3} - \frac{\mu_8}{16c_0^4} + \frac{2\mu_6}{6c_0^3} + \frac{\mu_4}{4c_2^2} - \frac{\mu_4}{4c_0^2}, \]

(64)

we have

\[ K_{22} = E_g \left[ -\frac{\partial^2 \ln f}{\partial c_2^2} \right] \]

\[ = \frac{\mu_6}{3\mu_2^3} + \frac{\mu_8}{16\mu_2^4} - \frac{\mu_6}{3\mu_2^3} - \frac{\mu_4}{16\mu_2^4} \]

\[ = \frac{1}{16\mu_2^4} (\mu_8 - \mu_4^2) \]

(65)

\[ = J_{22}, \]

as seen from (57). Finally, let us find \( K_{12} \). We have

\[ \left. \frac{\partial^2 \ln f}{\partial c_0 \partial c_2} \right|_{c_2=0} = -\frac{g^4}{2c_0^3} - \frac{\mu_6}{8c_0^4} + \frac{2\mu_4}{4c_0^2} + \frac{\mu_4}{4c_0^2} \]

(66)

and hence,
\[ K_{12} = E_g \left[ -\frac{\partial^2 \ln f}{\partial c_0 \partial c_2} \right] = \frac{\mu_4}{2\mu_2^2} + \frac{\mu_6}{8\mu_2^3} - \frac{\mu_4}{2\mu_2^2} - \frac{\mu_4}{8\mu_2^3} \]
\[ = \frac{1}{8\mu_2^3}(\mu_6 - 4\mu_2^4) \]
\[ = J_{12}, \]

as in (57). Thus, we have

\[ J_g = K_g = \left[ \begin{array}{c} \frac{\mu_4 - \mu_2^4}{4\mu_2^2} \\ \frac{1}{8\mu_2^3} (\mu_6 - 4\mu_2^4) \end{array} \right]. \quad (67) \]

Under the null hypothesis of no excess kurtosis, i.e., under \( \mu_4 = 3\mu_2^4 \), the above matrix reduces to

\[ \left[ \begin{array}{c} \frac{1}{2\mu_2^2} \\ \frac{1}{8\mu_2^3} (\mu_6 - 3\mu_2^6) \end{array} \right]. \quad (68) \]

Now we use expression (17) to obtain \( RS_{c2}^t \). First, note that since \( J_g = K_g \), \( K^{22}B_{22}K^{22} = K^{22} \). Therefore, (17) can be written as

\[ RS_{c2}^t = \frac{1}{n} \tilde{s}_2K^{22} \tilde{s}_2, \quad (69) \]

where \( K^{22} = (K_{22} - K_{21}K_{11}^{-1}K_{12})^{-1} \). We already had \( \tilde{s}_2 = \frac{n}{4}(b_2 - 3) \) in (49). Now

\[ K_{22} - K_{21}K_{11}^{-1}K_{12} = \frac{1}{16\mu_2^4} (\mu_8 - 9\mu_2^4) - \frac{2\mu_2^2}{64\mu_2^6} (\mu_6 - 3\mu_2^6)^2 \]
\[ = \frac{1}{32\mu_2^6} [2\mu_8\mu_2^2 - 27\mu_6^2 - \mu_6^2 + 6\mu_6\mu_2^6]. \quad (70) \]

Let us simplify (70) further by defining \( \beta_{2r} = \frac{\mu_{2r+2}}{\mu_2^r} \), \( r = 1, 2, \ldots \), i.e., \( \beta_4 = \frac{\mu_6}{\mu_2^4} \), and \( \beta_6 = \frac{\mu_8}{\mu_2^6} \). Thus, we have

\[ K_{22} - K_{21}K_{11}^{-1}K_{12} = \frac{1}{32} [2\mu_8 \beta_4 - 27 - (\mu_6 \beta_4)^2 + 6\mu_6 \beta_4]. \quad (71) \]

\[ = \frac{1}{32} [2\beta_6 - 27 - \beta_4^2 + 6\beta_4]. \]

21
Therefore, from (69), the adjusted RS statistic is given by

\[
RS_{c_2}^* = \frac{n}{16} (b_2 - 3)^2 \cdot \frac{32}{2b_6 - 27 - b_2^4 + 6b_4} = \frac{n}{2b_6 - 27 - b_2^4 + 6b_4}.
\]  

(72)

If we impose normality, then \(\beta_6 = 105, \beta_4 = 15\), and we can replace \(2b_6 - 27 - b_2^4 + 6b_4\) by \((2 \times 105 - 27 - 15 \times 15 + 6 \times 15) = 48\). Then \(RS_{c_2}^*\) reduces to \(n(b_2 - 3)^2/24\) the same as our unadjusted RS test for kurtosis in (60). There is one problem with this adjusted form. The variance formula (71) does not explicitly take into account the possible presence of asymmetry through \(\beta_1\), although it does allow for some other higher moment (sixth and eights) violations. Therefore, we try our previous general formula (44) to obtain an alternative expression for \(V(b_2)\). After some derivations, we have [see, for example, Stuart and Ord (1994, p. 349)]

\[
V(b_2) = \frac{1}{n} \left[ \frac{1}{\mu_2^2} + 64\beta_1 - \frac{8\mu_5\mu_3}{\mu_2^4} - 12\frac{\mu_6}{\mu_2^5} + 99 \right].
\]  

(73)

Utilizing \(\mu_8 = 105\mu_2^4\) and \(\mu_6 = 15\mu_2^4\) but still allowing for asymmetry through \(\mu_3\) and \(\mu_5\), we can write

\[
V(b_2) = \frac{1}{n} [24 + 64\beta_1 - 8\frac{\mu_5\mu_3}{\mu_2^4}].
\]  

(74)

Let us define \(\beta_{2r+1} = \mu_3\mu_{2r+3}/\mu_2^{r+3}\), i.e., \(\beta_1 = \mu_3^2/\mu_2^3\) and \(\beta_3 = \mu_3\mu_5/\mu_2^4\). Then we have a simple form for \(V(b_2)\),

\[
V(b_2) = \frac{1}{n} [24 + 64\beta_1 - 8\beta_3].
\]

Therefore, the extra quantity due to asymmetry is \(n^{-1}(64\beta_1 - 8\beta_3)\). Hence, the RS statistic for testing excess kurtosis that takes into account the possible presence of asymmetry can be written as

\[
RS_{c_2}^{**} = n \frac{(b_2 - 3)^2}{24 + 64\beta_1 - 8\beta_3}.
\]  

(75)

We will use this form of the adjusted test rather than that given in (72) in our simulation study. We should, however, note that although we do not study \(RS_{c_2}^{**}\) any further since it
does not serve our special purpose of taking account of asymmetry, this statistic might be better than the unadjusted test $RS_{x'}$ in certain circumstances.

6 Comparison of the Standard and Adjusted Tests for Excess Kurtosis: Some Simulation Results

In Table 3 we present the simulation results with 5000 replications. The design of our simulation study is very similar to that in Table 1, except that here we study tests for excess kurtosis instead of asymmetry. Again, for the standard $b_2$ test [see equation (60)], we use the tabulated values from Pearson and Hartley (1976, p. 184), and for the adjusted test [see equation (75)], we use asymptotic $\chi^2$ values. When the data are generated under the normal distribution, the standard test has rejection probabilities nearly twice the nominal sizes of 1 and 5%. The modified test is also oversized for the 1% level but is undersized when the nominal level is 5%. Under the t-distribution, when the standard test should have been ideal (since there is no asymmetry), as expected, the test has reasonable power. The adjusted test uses some redundant adjustments but still its estimated powers are very close to those of the standard test and, in some cases, are even higher. Therefore, we do not loose much power in using the adjusted test when the standard test would have been good enough. The next data set is generated according to Ramberg et al. (1979). This data set is from a non-normal distribution but with the same structure as the normal, i.e., symmetric and with no excess kurtosis. It is not surprising that the results are very similar to those of the normal distribution. We then used generated data with no excess kurtosis but with a modest amount of asymmetry ($\sqrt{\beta_1} = 0.85$). The results reveal the drawback of the standard test—it rejects the true null too often, whereas the adjusted test also rejects more than 1 or 5%, but not excessively. In the last two data sets, we have the same amount of excess kurtosis ($\beta_2 - 3 = 4$) but different skewness structures. The estimated powers of the
adjusted test seem to be slightly higher than those of the standard test. Also, the adjusted version appears to be less influenced by the presence of asymmetry ($\beta_1 = 0.5$) in terms of having less fluctuations in the estimated powers, as move from the symmetric to asymmetric data. Although the finite sample performance of our adjusted test is very good both in terms of size and power, we should note that it requires the existence of higher moments.

7 An Empirical Illustration

Our empirical illustration is done with some financial data. Considerable attention has been paid to the asymmetry and excess kurtosis of return distributions. Beanless and Simkowitz (1980), Singleton and Wingender (1986), and Alles and King (1994) have all studied the symmetry of returns using the $\sqrt{b_1}$ test. DeFusco, Kareles, and Muralidhar (1996) correctly argued that the standard $\sqrt{b_1}$ test is not accurate enough to provide correct information on the skewness of the distribution. However, they did not suggest any modification to the $\sqrt{b_1}$ test that could be used in the presence of excess kurtosis. Mandelbrot (1963) was probably the first to systematically study the excess kurtosis in the financial data, as he noted that the price changes were usually too peaked and thick tailed compared to samples from the normal distribution. To take into account this excess kurtosis, Mandelbrot (1963, 1967), Fama (1965) and others suggested the use of stable distribution. Blattberg and Gonedes (1974) and Tucker (1992) examined the use of Student’s $t$-distribution. More recently, Premaratne and Bera (2000) advocated the use of the Pearson type IV distribution, which is a special case of (21) that takes care of both asymmetry and excess kurtosis in a simple way. However, there still remains the question of testing for the presence of asymmetry and excess kurtosis separately while allowing for the presence of the other effect. In our empirical illustration we use tests discussed in the earlier sections.

We considered daily S & P 500 returns from the Center for Research in Security Prices (CRSP) database from August 27, 1990. We tried two sample sizes, 300 and 500. The results
Table 3: Estimated Sizes and Powers of the Two Tests for Excess Kurtosis with 5000 Replications

<table>
<thead>
<tr>
<th>DGP and Sample Sizes</th>
<th>Standard Test 1%</th>
<th>Standard Test 5%</th>
<th>Modified Test 1%</th>
<th>Modified Test 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>N(0,1) (β₁ = 0, β₂ = 3.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.46</td>
<td>9.46</td>
<td>1.92</td>
<td>3.16</td>
</tr>
<tr>
<td>100</td>
<td>1.80</td>
<td>9.66</td>
<td>2.18</td>
<td>3.60</td>
</tr>
<tr>
<td>200</td>
<td>1.68</td>
<td>9.58</td>
<td>1.72</td>
<td>3.90</td>
</tr>
<tr>
<td>250</td>
<td>2.12</td>
<td>10.48</td>
<td>1.70</td>
<td>4.52</td>
</tr>
<tr>
<td>t₁₀ (β₁ = 0, β₂ = 4.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>10.72</td>
<td>24.76</td>
<td>7.20</td>
<td>10.92</td>
</tr>
<tr>
<td>100</td>
<td>16.76</td>
<td>34.56</td>
<td>14.90</td>
<td>21.74</td>
</tr>
<tr>
<td>200</td>
<td>30.56</td>
<td>51.36</td>
<td>29.76</td>
<td>39.42</td>
</tr>
<tr>
<td>250</td>
<td>36.52</td>
<td>57.92</td>
<td>35.62</td>
<td>47.12</td>
</tr>
<tr>
<td>Generated Data (β₁ = 0, β₂ = 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.44</td>
<td>8.16</td>
<td>1.40</td>
<td>2.40</td>
</tr>
<tr>
<td>100</td>
<td>1.54</td>
<td>9.08</td>
<td>1.62</td>
<td>3.32</td>
</tr>
<tr>
<td>200</td>
<td>1.26</td>
<td>8.26</td>
<td>1.22</td>
<td>3.32</td>
</tr>
<tr>
<td>250</td>
<td>1.12</td>
<td>7.84</td>
<td>0.86</td>
<td>3.08</td>
</tr>
<tr>
<td>Generated Data (β₁ = 0.85, β₂ = 3.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>3.98</td>
<td>15.46</td>
<td>2.98</td>
<td>4.42</td>
</tr>
<tr>
<td>100</td>
<td>3.36</td>
<td>11.86</td>
<td>3.34</td>
<td>5.66</td>
</tr>
<tr>
<td>200</td>
<td>3.38</td>
<td>14.80</td>
<td>2.46</td>
<td>5.58</td>
</tr>
<tr>
<td>250</td>
<td>3.34</td>
<td>14.40</td>
<td>2.30</td>
<td>5.40</td>
</tr>
<tr>
<td>Generated data (β₁ = 0, β₂ = 7.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>11.06</td>
<td>36.70</td>
<td>20.38</td>
<td>29.06</td>
</tr>
<tr>
<td>100</td>
<td>38.60</td>
<td>65.30</td>
<td>48.70</td>
<td>58.96</td>
</tr>
<tr>
<td>200</td>
<td>72.47</td>
<td>77.06</td>
<td>91.44</td>
<td>81.70</td>
</tr>
<tr>
<td>250</td>
<td>88.20</td>
<td>96.18</td>
<td>90.84</td>
<td>94.88</td>
</tr>
<tr>
<td>Generated Data (β₁ = 0.5, β₂ = 7.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>9.52</td>
<td>22.56</td>
<td>18.18</td>
<td>26.34</td>
</tr>
<tr>
<td>100</td>
<td>36.22</td>
<td>52.14</td>
<td>46.14</td>
<td>56.56</td>
</tr>
<tr>
<td>200</td>
<td>74.66</td>
<td>80.98</td>
<td>90.64</td>
<td>87.98</td>
</tr>
<tr>
<td>250</td>
<td>84.62</td>
<td>95.00</td>
<td>87.98</td>
<td>93.26</td>
</tr>
</tbody>
</table>
are given in Table 4. For the standard $\sqrt{b_1}$ test, the critical values at 1% were 0.329 and 0.255, respectively, for $n = 300$ and 500 [see Pearson and Hartley (1976, p. 183)]. Therefore, we rejected symmetry. However, our adjusted $\sqrt{b_1}$ test supported the null hypothesis of symmetry using $\chi^2_1$ values. As we noted in our simulation study, the standard $\sqrt{b_1}$ test over-rejects the null of symmetry in the presence of excess kurtosis. The two-sided critical values for $b_2$ at 1% are (2.46, 3.79) and (2.57, 3.60), respectively, for $n = 300$ and 500 [see Pearson and Hartley (1976, p. 184)]. The two kurtosis tests gave similar results implying excess kurtosis in the data for both sample sizes. We get similar results possibly due to either the absence of asymmetry (as revealed by our adjusted $\sqrt{b_1}$ test) or the strong presence of excess kurtosis.

Table 4: Use of Tests for Daily Returns of the S & P 500 Index

<table>
<thead>
<tr>
<th>Observations</th>
<th>Test for Asymmetry</th>
<th>Test for Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sqrt{b_1}$</td>
<td>Adjusted $\sqrt{b_1}$</td>
</tr>
<tr>
<td>300</td>
<td>0.376*</td>
<td>3.10</td>
</tr>
<tr>
<td>500</td>
<td>0.264*</td>
<td>1.36</td>
</tr>
</tbody>
</table>

*Significant at the 1% level.

8 Conclusion

In this paper we suggested some adjustments to the standard $\sqrt{b_1}$ and $b_2$ tests, allowing for excess kurtosis and asymmetry, respectively. From our simulation results, we noticed that the adjusted tests performed very well compared to their unadjusted counterparts. The adjusted tests were quite immune to the misspecifications that can arise through thick tail and the lack of symmetry. Of course, more simulation work and empirical applications are needed to further confirm the good properties of our suggested tests.
References


Bowman, K. O. and L. R. Shenton (1975), Omnibus contours for departures from normality based on $\sqrt{b_1}$ and $b_2$, *Biometrika, 62*, 243-250.


D’Agostino, R.D., and E.S. Pearson (1973), Tests for departure from normality: Empirical results for the distributions of $b_2$ and $\sqrt{b_1}$, *Biometrika, 60*, 613-622.


Mandelbrot, B. (1967), The variation of some other speculative prices, *Journal of Business*, 40, 393-413.

Pearson, E.S. (1930), A further development of tests for normality, *Biometrika*, 22, 239-249.


