Southern Illinois University Carbondale

OpenSIUC

2014

Simulating Univariate and Multivariate Tukey *g*-and-*h* Distributions Based on the Method of Percentiles

Tzu Chun Kuo
Southern Illinois University Carbondale

Todd C. Headrick
Southern Illinois University Carbondale, headrick@siu.edu

Research Article

Simulating Univariate and Multivariate Tukey g-and-h Distributions Based on the Method of Percentiles

Tzu Chun Kuo and Todd C. Headrick

Section on Statistics and Measurement, Department of EPSE, Southern Illinois University Carbondale, P.O. Box 4618, 222-J Wham Building, Carbondale, IL 62901-4618, USA

Correspondence should be addressed to Todd C. Headrick; headrick@siu.edu

This paper derives closed-form solutions for the g-and-h shape parameters associated with the Tukey family of distributions based on the method of percentiles (MOP). A proposed MOP univariate procedure is described and compared with the method of moments (MOM) in the context of distribution fitting and estimating skew and kurtosis functions. The MOP methodology is also extended from univariate to multivariate data generation. A procedure is described for simulating nonnormal distributions with specified Spearman correlations. The MOP procedure has an advantage over the MOM because it does not require numerical integration to compute intermediate correlations. Simulation results demonstrate that the proposed MOP procedure is superior to the MOM in terms of distribution fitting, estimation, relative bias, and relative error.

1. Introduction

The Tukey g-and-h families of univariate and multivariate nonnormal distributions are commonly used for distribution fitting, modeling events, random variable generation, and other applied mathematical contexts such as operational risk, extreme oceanic wind speeds, common stock returns, and solar flare data. See [1–17].

The family of univariate g-and-h distributions can be summarized as follows:

$$Y = q(Z) = q_{g,h}(Z) = g^{-1}\left(e^{gZ} - 1\right)e^{hZ^2/2}, \quad (1)$$

$$Y = q(Z) = q_{g,0}(Z) = \lim_{h \to 0} q_{g,h}(Z) = g^{-1}\left(e^{gZ} - 1\right), \quad (2)$$

$$Y = q(Z) = q_{0,h}(Z) = \lim_{g \to 0} q_{g,h}(Z) = Ze^{hZ^2/2}, \quad (3)$$

where Z is an i.i.d. standard normal random variable with probability density function (pdf), $\phi(z)$, and cumulative distribution function (cdf), $\Phi(z)$. The transformations in (1)–(3) are strictly monotone increasing functions with real-valued constants g and h that produce distributions defined as (i) asymmetric g-and-h ($g \neq 0$, $h > 0$), (ii) log-normal ($g \neq 0$,

$h = 0$), and (iii) symmetric h ($h \geq 0$), respectively. The constant $\pm g$ controls the skew of a distribution in terms of both direction and magnitude. Taking the negative of g will change the direction of the skew but not its magnitude; that is, $q_{-g,h}(Z) = -q_{g,h}(-Z)$. The constant h controls the tail-weight of a distribution where the function $e^{hZ^2/2}$ (i) preserves symmetry, (ii) is increasing for $Z \geq 0$ and $h \geq 0$, and (iii) produces increased tail-weight as the value of h becomes larger. In summary, (1)–(3) are computationally efficient for the purpose of generating nonnormal distributions because they only require the specification of one or two shape parameters (g, h) and an algorithm that produces standard normal random deviates.

The values of g and h associated with (1)–(3) can be determined from either the method of moments (MOM), for example, [8, 10, 13], or the method of percentiles (MOP), for example, [9, 18]. However, these two methods have disadvantages. Specifically, in the context of the MOM, the estimates of conventional skew and kurtosis associated with heavy-tailed or -skewed distributions can be substantially biased, have high variance, or can be influenced by outliers; see, for example, [19–24]. In terms of the MOP, the primary

disadvantage associated with the g-and-h procedure is the laborious "multiplicative iterative" approach needed to determine the h constant for symmetric and asymmetric nonnormal g-and-h distributions; see [9, p. 484–489].

On the other hand, the MOP approach described by Karian and Dudewicz [24–27] in the context of the generalized lambda distribution (GLD) has demonstrated to be an attractive and computationally efficient alternative to the MOM in terms of distribution fitting and computing the GLD shape parameters. Further, it has been demonstrated that the MOP is superior to the MOM over a broad range of combinations of skew and kurtosis for fitting GLDs to theoretical or empirical distributions (see [24]).

A procedure for simulating correlated g-and-h distributions with a specified correlation structure based on (1)–(3) is described by Kowalchuk and Headrick [10] and Headrick [28]. Specifically, the g-and-h procedure makes use of the popular NORTA (NORmal To Anything) approach. That is, the procedure begins with generating multivariate standard normal deviates prior to transformation. However, one limitation arises because the Pearson correlation is not invariant under nonlinear strictly increasing transformations. This is a concern because the transformations in (1)–(3) have this characteristic, for example, [8, 28]. Thus, the initial multivariate normal correlation structure used in the NORTA approach will not be maintained subsequent to any of the transformations in (1)–(3). As such, the NORTA procedure must begin with the computation of an intermediate correlation (IC) matrix, which is different from the specified correlation matrix between the nonnormal distributions. The purpose of the IC matrix is to adjust for the non-normalization effect of a transformation such that the g-and-h procedure can generate nonnormal distributions with a specified correlation matrix. Further, in the context of the g-and-h procedure, NORTA also requires numerical integration techniques to solve for the IC matrix, see [28, p. 143], which can be computationally expensive.

In contradistinction, the Spearman correlation has comparative advantages over the Pearson correlation in the context of the g-and-h procedure for computing ICs. Specifically, the transformations in (1)–(3) must be strictly monotone increasing functions to produce valid pdfs. Thus, subsequent to transformation, the rank order of Y, $R(Y)$, remains the same as the rank order of Z, $R(Z)$; that is, the Spearman correlation remains unchanged. Moreover, there is a straightforward equation that can be used to directly solve for all pairwise ICs, for example, [28, p. 114], and thus does not require the numerical integration techniques associated with the conventional product MOM g-and-h procedure.

In view of the above, the present aim is to obviate the problems associated with the MOM in the context of the family of g-and-h distributions in (1)–(3) by characterizing these distributions through the MOP as described in Karian and Dudewicz [27, p. 172-173]. Specifically, the purpose of this paper is to obviate the laborious "multiplicative iterative" technique given in [9] and develop the methodology and a procedure for simulating g-and-h distributions with specified medians (γ_1), interdecile ranges (γ_2), left-right tail-weight ratios (γ_3, a skew function), and tail-weight factors (γ_4,

a kurtosis function). In terms of simulating multivariate g-and-h distributions, the Spearman correlation will be used in lieu of the Pearson correlation. In summary, the advantages of the proposed MOP procedure are that (i) the MOP parameters (γ_3, γ_4) exist for any distribution, whether the mean exists or not, for example, [25]; (ii) there are less relative bias and less relative standard error when juxtaposed with the MOM procedure; (iii) there are closed-form solutions for the g and h constants; (iv) there is a straightforward equation to obtain pairwise ICs for the purpose of simulating correlated nonnormal g-and-h distributions in (1)–(3).

The remainder of the paper is outlined as follows. In Section 2, a summary of the univariate g-and-h distributions based on the MOM is provided. In Section 3, the requisite information associated with the MOP is provided. Further, the systems of equations for determining the closed-form solutions of the g and h constants are subsequently derived for simulating univariate nonnormal distributions with specified values of γ_1–γ_4. In Section 4, a comparison of the MOM and the MOP is provided by fitting the SPSS data from [29]. In Section 5, the methodology for simulating correlated nonnormal g-and-h distributions with the specified Spearman correlation matrices is provided. In Section 6, the steps for implementing the proposed MOP procedure are described. A numerical example and results of a simulation are also provided to confirm the derivations and compare the proposed procedure with the MOM procedure. In Section 7, the results of the simulation are discussed.

2. Preliminaries for the Tukey Family of g-and-h Distributions

2.1. The Tukey Family of g-and-h Distributions Based on the MOM. The requirement that $q(Z)$ in (1)–(3) be a strictly monotone increasing function implies that an inverse function (q^{-1}) exists and thus $F_{q(z)}(z) = \Phi(z)$, where $F_{q(z)}(z)$ is the general form of the cdf for both the MOM and the MOP. Differentiating both sides with respect to $q(z)$ yields $dF_{q(z)}(z)/dq(z) = f_{q(z)}(z)$, where $f_{q(z)}(z)$ is the general form of the pdf for both the MOM and the MOP. Hence, $dF_{q(z)}(z)/dq(z) = (dF_{q(z)}(z)/dz)/(dq(z)/dz) = \phi(z)/q'(z)$. Whence, the pdf integrates to one because $\phi(z)$ is the standard normal pdf and will be nonnegative on its support $q(z)$ for $z \in (-\infty, +\infty)$ given that $h \geq 0$ and where $\lim_{z \to \pm\infty} \phi(z)/q'(z) = 0$ for the transformations in (1)–(3).

The g and/or h constants associated with (1)–(3) that determine the shape of a distribution are computed using a moment-matching approach that involves the conventional measures of the mean (α_1), variance (α_2), skew (α_3), and kurtosis (α_4). Specifically, the values of g and h in (1) are determined by simultaneously solving (6) and (7) for specified values of skew and kurtosis (e.g., [28, p. 125, Eqs. 5.16-5.17], and [10, (A1)–(A4)]):

$$\alpha_1 = \frac{\left(\exp\left\{g^2/(2-2h)\right\} - 1\right)}{\left(g(1-h)^{1/2}\right)}, \tag{4}$$

$$\alpha_2 = \left[\frac{\left(1 - 2\exp\left\{g^2/(2-4h)\right\} + \exp\left\{2g^2/(1-2h)\right\}\right)}{(1-2h)^{1/2}} \right.$$

$$\left. + \frac{\left(\exp\left\{g^2/2 - 2h\right\} - 1\right)^2}{(h-1)} \right] \times g^{-2},$$

(5)

$$\alpha_3 = \left[\left(\left(3\exp\left\{\frac{g^2}{(2-6h)}\right\} + \exp\left\{\frac{9g^2}{(2-6h)}\right\} \right. \right. \right.$$

$$\left. - 3\exp\left\{\frac{2g^2}{(1-3h)}\right\} - 1 \right) \times (1-3h)^{-1/2}$$

$$- 3\left(1 - 2\exp\left\{\frac{g^2}{(2-4h)}\right\} + \exp\left\{\frac{2g^2}{(1-2h)}\right\}\right)$$

$$\times \left(\exp\left\{\frac{g^2}{(2-2h)}\right\} - 1\right) \times \left((1-2h)^{1/2}(1-h)^{1/2}\right)^{-1}$$

$$\left. + \frac{2\left(\exp\left\{g^2/(2-2h)\right\} - 1\right)^3}{(1-h)^{3/2}} \right]$$

$$\times \left[g^3 \left(\left(\left(1 - 2\exp\left\{\frac{g^2}{(2-4h)}\right\} + \exp\left\{\frac{2g^2}{(1-2h)}\right\}\right) \right. \right. \right.$$

$$\times (1-2h)^{-1/2} + \frac{\left(\exp\left\{g^2/(2-2h)\right\} - 1\right)^2}{(h-1)} \right)$$

$$\left. \left. \times g^{-2} \right)^{3/2} \right]^{-1},$$

(6)

$$\alpha_4 = \left[\exp\left\{\frac{8g^2}{(1-4h)}\right\} \right.$$

$$\times \left(1 + 6\exp\left\{\frac{6g^2}{(4h-1)}\right\} + \exp\left\{\frac{8g^2}{(4h-1)}\right\} \right.$$

$$\left. - 4\exp\left\{\frac{7g^2}{(8h-2)}\right\} - 4\exp\left\{\frac{15g^2}{(8h-2)}\right\}\right)$$

$$\times (1-4h)^{-1/2}$$

$$- 4\left(3\exp\left\{\frac{g^2}{(2-6h)}\right\} + \exp\left\{\frac{9g^2}{(2-6h)}\right\}\right.$$

$$\left. - 3\exp\left\{\frac{2g^2}{(1-3h)}\right\} - 1\right)$$

$$\times \frac{\left(\exp\left\{g^2/(2-2h)\right\} - 1\right)}{(1-3h)^{1/2}(1-h)^{1/2}}$$

$$- \frac{6\left(\exp\left\{g^2/(2-2h)\right\} - 1\right)^4}{(h-1)^2}$$

$$- 12\left(1 - 2\exp\left\{\frac{g^2}{(2-4h)}\right\} + \exp\left\{\frac{2g^2}{(1-2h)}\right\}\right)$$

$$\times \frac{\left(\exp\left\{g^2/(2-2h)\right\} - 1\right)^2}{(1-2h)^{1/2}(h-1)}$$

$$+ 3\left(1 - 2\exp\left\{\frac{g^2}{(2-4h)}\right\} + \exp\left\{\frac{2g^2}{(1-2h)}\right\}\right)^2$$

$$\times (2h-1)^{-1}\right]$$

$$\times \left[\left(1 - 2\exp\left\{\frac{g^2}{(2-4h)}\right\} + \exp\left\{\frac{2g^2}{(1-2h)}\right\}\right)\right.$$

$$\times (1-2h)^{-1/2} + \left(\exp\left\{\frac{g^2}{(2-2h)}\right\} - 1\right)^2$$

$$\left. \times (h-1)^{-1}\right]^{-2}.$$

(7)

The mean and the variance for a g-and-h distribution can be determined by evaluating (4) and (5) using the solutions of g and h obtained from (6) and (7). For any g-and-h distribution, the kth moment will exist if $0 \leq h < 1/k$ (see [8]). As such, (4)–(7) exist if the first four moments are defined. That is, $0 \leq h < 1/4$. Note that the equations of α_1–α_4 for g distributions in (2) and h distributions in (3) can be obtained from [10, Eqs. (A5)–(A8)] and [10, Eqs. (A9)–(A12)], respectively.

3. The Percentile Based g-and-h Family of Distributions

3.1. General Considerations. The percentiles (θ_p) associated with a conventional based g-and-h pdf can be obtained by making use of the standard normal cdf, $\Phi(z)$. As such, we will define the following location, scale, and shape parameters as in [27, p. 172-173]

$$\gamma_1 = \theta_{0.50},$$

$$\gamma_2 = \theta_{0.90} - \theta_{0.10},$$

$$\gamma_3 = \frac{\theta_{0.50} - \theta_{0.10}}{\theta_{0.90} - \theta_{0.50}},$$

$$\gamma_4 = \frac{\theta_{0.75} - \theta_{0.25}}{\gamma_2},$$

(8)

where (8) are the (i) median, (ii) interdecile range, (iii) left-right tail-weight ratio (a skew function), and (iv) tail-weight

factor (a kurtosis function), respectively. The parameters in (8) are defined to have the restrictions

$$-\infty < \gamma_1 < +\infty, \qquad \gamma_2 \geq 0, \qquad \gamma_3 \geq 0, \qquad 0 \leq \gamma_4 \leq 1, \tag{9}$$

and where a symmetric distribution implies that $\gamma_3 = 1$.

3.2. The Tukey Family of g-and-h Distributions Based on the MOP. The derivation of a percentile based system of g-and-h pdfs begins by substituting the standard normal distribution percentiles (z_p) into the quantile functions in (1)–(3) and making use of (8) gives

$$\gamma_1 = q(z_{0.50}),$$

$$\gamma_2 = q(z_{0.90}) - q(z_{0.10}),$$

$$\gamma_3 = \frac{q(z_{0.50}) - q(z_{0.10})}{q(z_{0.90}) - q(z_{0.50})}, \tag{10}$$

$$\gamma_4 = \frac{q(z_{0.75}) - q(z_{0.25})}{\gamma_2},$$

where $z_{0.50} = 0$, $z_{0.90} = 1.281\cdots$, $z_{0.75} = 0.6744\cdots$ from the standard normal distribution. Note from symmetry that $z_{0.10} = -z_{0.90}$ and $z_{0.25} = -z_{0.75}$. The explicit forms of (10) are

$$\gamma_1 = 0, \tag{11}$$

$$\gamma_2 = \frac{[(1/2)\exp\{z_{0.90}(hz_{0.75} - 2g)\}(\exp\{2gz_{0.90}\} - 1)]}{g}, \tag{12}$$

$$\gamma_3 = \exp\{-gz_{0.90}\}, \tag{13}$$

$$\gamma_4 = \frac{1}{2}\left[\exp\left\{\frac{(z_{0.75} - z_{0.90})(h(z_{0.90} + z_{0.75}) - 2g)}{2}\right\} \right. $$
$$\left. \times (\exp\{2gz_{0.75}\} - 1)(\coth[gz_{0.90}] - 1)\right]. \tag{14}$$

Simultaneously, solving for the coefficients in (13) and (14) gives the convenient closed-form expressions

$$g = -\frac{\ln(\gamma_3)}{z_{0.90}},$$

$$h = \frac{2\ln\left(\left[\gamma_3^{1-z_{0.75}/z_{0.90}}\left(\gamma_3^{2z_{0.75}/z_{0.90}} - 1\right)\right] / \left[\left(\gamma_3^2 - 1\right)\gamma_4\right]\right)}{z_{0.90}^2 - z_{0.75}^2}. \tag{15}$$

For a symmetric distribution ($g = 0$), the closed-form solution for h can be expressed as

$$h = \frac{2\ln(z_{0.75}/(\gamma_4 z_{0.90}))}{z_{0.90}^2 - z_{0.75}^2}. \tag{16}$$

Estimates of $\gamma_1 - \gamma_4$ based on the percentiles in (10) for a sample of size n can be determined by finding the j and $j+1$

integer values and their corresponding expected values of the order statistics $E[q(Z)_{j:n}]$ and $E[q(Z)_{j+1:n}]$, by making use of the following equation (see [30, 31]):

$$E\left[q(Z)_{j:n}\right] = \frac{n!}{(j-1)!(n-j)!}$$
$$\times \int_{-\infty}^{+\infty} q(z)\phi(z)\{\Phi(z)\}^{j-1}\{1 - \Phi(z)\}^{n-j}dz \tag{17}$$

such that

$$E\left[q(Z)_{j:n}\right] \leq q(z_p) \leq E\left[q(Z)_{j+1:n}\right] \tag{18}$$

and subsequently solve the equation

$$q(z_p) = (u)E\left[q(Z)_{j:n}\right] + (1-u)E\left[q(Z)_{j+1:n}\right] \tag{19}$$

for $0 \leq u \leq 1$. Thus, an empirical estimate of $q(z_p)$ can then be obtained based on the order statistics of a sample of size n as $q(z_p) \simeq q(Z_p) = (u)q(Z)_{j:n} + (1-u)q(Z)_{j+1:n}$.

4. A Comparison of the MOM and the MOP on Distribution Fitting

Presented in Figure 1 are the MOM and the MOP pdfs superimposed on the histogram of the SPSS *customer_dbase* data from [29]. This is a data file that concerns the company's efforts to use the information in its data warehouse to make special offers for customers who are most likely to reply. Specifically, these data are the amount each customer spent on their primary credit card in the last month. The parameters $(\alpha_{3,4}, \gamma_{3,4})$ associated with Figure 1 were based on a sample size of $n = 5{,}000$ participants. The bootstrap estimates, confidence intervals, and relative standard errors of $(\hat{\alpha}_{3,4}, \hat{\gamma}_{3,4})$ are also provided in Figure 1. Note that to fit the g-and-h distributions to the data, a linear transformation has to be imposed on $q(z) = Aq(z) + B$ for both the MOM and the MOP procedures. Specifically, in the context of the MOM, $A = s/\sigma$ and $B = m - A\mu$. In the context of the MOP, $A = \mathrm{IDR}/\gamma_2$ and $B = \mathrm{med}$. The values of the sample and the theoretical means (m, μ), standard deviations (s, σ), medians (med, γ_1), and interdecile range (IDR, γ_2) for the data and g-and-h pdfs are given in Figure 1.

Visual inspection of the approximation in Figure 1 indicates that the MOP pdf provides a more accurate fit to the actual data over the MOM. Further, $\gamma_{3,4}$ have more precision than $\alpha_{3,4}$ because the relative standard errors (RSEs) of $\hat{\gamma}_{3,4}$ are much less than those of $\hat{\alpha}_{3,4}$. To compare the accuracy of the data fitting for the MOP and the MOM, the Euclidean distances are calculated as $d = \sqrt{\sum(O - E)^2}$, where O is the observed proportion in each interval and E is the expected proportion in each interval for both the MOP and the MOM. From Table 1, the MOP has a more accurate data fit because the Euclidean distance (d) of the MOP is less than that of the MOM.

	Data	g-h
	m = 339.636	μ = 0.4122
	s = 248.383	σ = 1.5044
		g = 0.71358
		h = 0.00879

Amount spent on primary credit card in the last month

MOM

Parameter	Estimate	95% bootstrap CI	RSE (%)
α_3 = 3.15505	$\hat{\alpha}_3$ = 3.114	(2.47421, 4.42061)	14.6
α_4 = 22.374	$\hat{\alpha}_4$ = 21.74	(12.4157, 43.3088)	31.7

(a)

	Data	g-h
	med = 278.65	γ_1 = 0
	IDR = 490.95	γ_2 = 3.1649
		g = 0.59325
		h = 0.14171

Amount spent on primary credit card in the last month

MOP

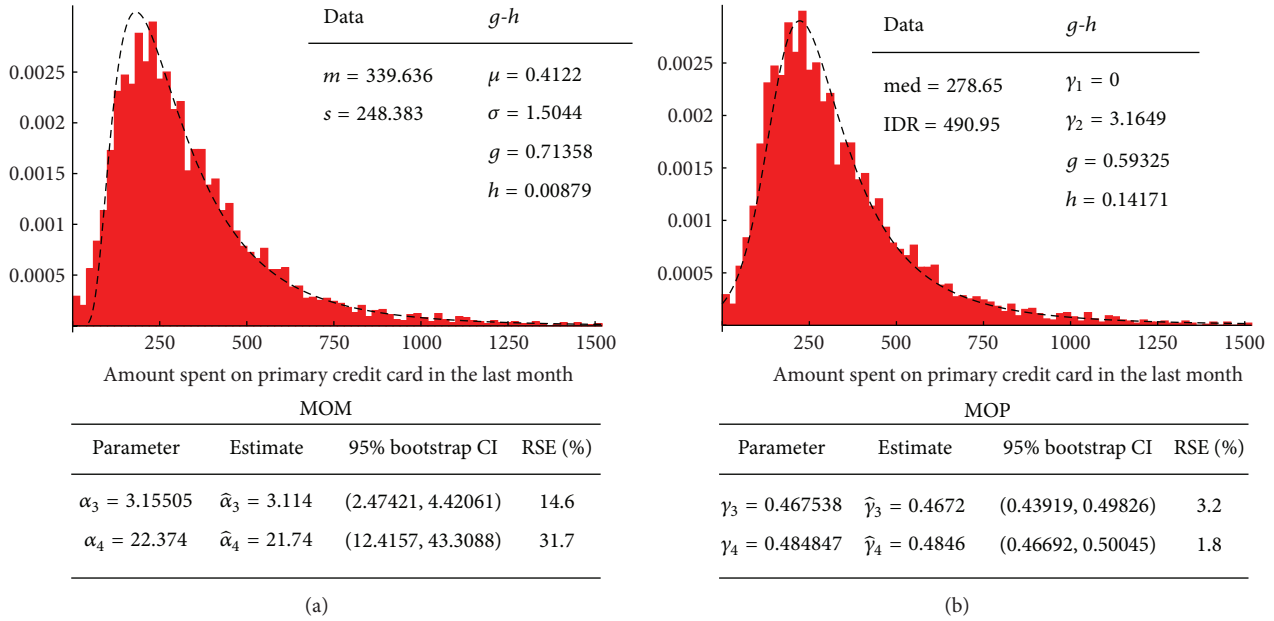| Parameter	Estimate	95% bootstrap CI	RSE (%)
γ_3 = 0.467538	$\hat{\gamma}_3$ = 0.4672	(0.43919, 0.49826)	3.2
γ_4 = 0.484847	$\hat{\gamma}_4$ = 0.4846	(0.46692, 0.50045)	1.8

(b)

FIGURE 1: Histograms and estimates based on the MOM and the MOP for the SPSS credit card *customer_dbase* data from [29].

TABLE 1: Percentiles, expected proportions, observed proportions, and the Euclidean distances (d) for the MOP and the MOM approximations associated with the SPSS *customer_dbase* data (n = 5,000) from [29].

Percentile	Expected prop.	Obs. prop. (MOM)	Obs. prop. (MOP)
10	0.10	0.1240	0.1000
25	0.15	0.1208	0.1658
50	0.25	0.2392	0.2340
75	0.25	0.2564	0.2354
90	0.15	0.1632	0.1646
		d = 0.04196	d = 0.03053

TABLE 2: Specified correlation matrix for the distributions in Figures 2(a)–2(d).

	1	2	3	4
1	1			
2	0.40	1		
3	0.60	0.50	1	
4	0.65	0.70	0.60	1

TABLE 3: Intermediate correlations for the MOM procedure.

	1	2	3	4
1	1			
2	0.418072	1		
3	0.786146	0.646711	1	
4	0.906185	0.796341	0.702012	1

TABLE 4: Intermediate correlations for the MOP procedure (n = 25).

	1	2	3	4
1	1			
2	0.431321	1		
3	0.638650	0.536062	1	
4	0.688961	0.738501	0.638650	1

TABLE 5: Intermediate correlations for the MOP procedure (n = 750).

	1	2	3	4
1	1			
2	0.416344	1		
3	0.618734	0.518259	1	
4	0.668342	0.717483	0.618734	1

TABLE 6: The Cholesky decompositions for the MOM procedure.

a_{11} = 1	a_{12} = 0.418072	a_{13} = 0.786146	a_{14} = 0.736275
0	a_{22} = 0.908414	a_{23} = 0.350111	a_{24} = 0.537778
0	0	a_{33} = 0.509310	a_{34} = −0.127801
0	0	0	a_{44} = 0.390334

TABLE 7: The Cholesky decompositions for the MOP procedure (n = 25).

a_{11} = 1	a_{12} = 0.431321	a_{13} = 0.638650	a_{14} = 0.688961
0	a_{22} = 0.902198	a_{23} = 0.288848	a_{24} = 0.489180
0	0	a_{33} = 0.713227	a_{34} = 0.080404
0	0	0	a_{44} = 0.528745

TABLE 8: The Cholesky decompositions for the MOP procedure ($n = 750$).

$a_{11} = 1$	$a_{12} = 0.416343$	$a_{13} = 0.618734$	$a_{14} = 0.668342$
0	$a_{22} = 0.909207$	$a_{23} = 0.286682$	$a_{24} = 0.483084$
0	0	$a_{33} = 0.731424$	$a_{34} = 0.091215$
0	0	0	$a_{44} = 0.558237$

5. The Spearman Correlations for the System of the MOP

We assume that the variates $Y_i = q(Z_i)$ and $Y_j = q(Z_j)$ in (1)–(3) produce valid pdfs and are thus increasing monotonic transformations in Z_i and Z_j. This implies that the rank orders of Y_i ($R(Y_j)$) and Z_i ($R(Z_j)$) are identical and thus will have rank correlations of $\rho_{R(Y_i),R(Z_i)} = \rho_{R(Y_j),R(Z_j)} = 1$.

Given these assumptions, suppose it is desired to simulate a T-variate distribution from the quantile functions in (1)–(3) with a specified $T \times T$ Spearman correlation matrix and where each distribution has specified γ_3 and γ_4. Let Z_i and Z_j have univariate normal pdfs $\phi(z_i)$ and $\phi(z_j)$, the Pearson correlation ρ_{z_i,z_j}, and standard normal bivariate density of

$$
\begin{aligned}
f_{ij} &:= f_{z_i z_j}\left(z_i, z_j, \rho_{z_i,z_j}\right) \\[2mm]
&= \left(2\pi \sqrt{1 - \rho_{z_i,z_j}^2}\right)^{-1} \\[2mm]
&\quad \times \exp\left\{-\left(2(1 - \rho_{z_i,z_j}^2)\right)^{-1}\left(z_i^2 - 2\rho_{z_i,z_j} z_i z_j + z_j^2\right)\right\}.
\end{aligned}
\tag{20}
$$

The correlation $\rho_{R(Y_i),R(Y_j)}$ can be obtained from the derivation of $\rho_{R(Z_i),R(Z_j)}$ given in [32]. That is, because (1)–(3) are strictly increasing monotonic transformations, $\rho_{R(Y_i),R(Y_j)} = \rho_{R(Z_i),R(Z_j)}$ and thus we have from Headrick [28, Eq. 4.34]

$$
\begin{aligned}
&\rho_{R(Y_i),R(Y_j)} \\[2mm]
&= \left(\frac{6}{\pi}\right)\left[\left(\frac{n-2}{n+1}\right)\sin^{-1}\left(\frac{\rho_{z_i,z_j}}{2}\right) + \left(\frac{1}{n+1}\right)\sin^{-1}\left(\rho_{z_i,z_j}\right)\right],
\end{aligned}
\tag{21}
$$

where $\rho_{R(Y_i),R(Y_j)}$ is the specified Spearman correlation and ρ_{z_i,z_j} is the intermediate correlation (IC). For a specified value of $\rho_{R(Y_i),R(Y_j)}$ and a finite sample size n, the value of the IC, ρ_{z_i,z_j}, can be obtained by numerically solving (21).

6. The Procedure for Simulation and Monte Carlo Study

To implement the method for simulating g-and-h, g, h distributions with specified γ_3, γ_4 and the Spearman correlations, we suggest the following six steps.

(1) Specify the values of γ_3 and γ_4 for the T transformations of the forms in (1)–(3), that is, $q_1(Z_1), \ldots,$ $q_T(Z_T)$ and obtain the constants of g and/or h by solving (15)-(16) using the specified values of γ_3 and γ_4 for each distribution. Specify a $T \times T$ matrix of the Spearman correlations between $q_j(Z_j)$ and $q_k(Z_k)$, where $j < k \in \{1, 2, \ldots, T\}$.

(2) Compute the ICs ρ_{z_i,z_j} by substituting the solutions of the constants from Step 1 into (21) and then solve for ρ_{z_i,z_j}. Repeat this step separately for all $T(T-1)/2$ pairwise combinations of correlations.

(3) Assemble the ICs into a $T \times T$ matrix and decompose this matrix using a Cholesky factorization. Note that this step requires the IC matrix to be positive definite.

(4) Use the results of the Cholesky factorization from Step 3 to generate T standard normal variables $(Z_1, \ldots, Z_T)$ correlated at the intermediate levels as follows:

$$
\begin{aligned}
Z_1 &= a_{11} V_1 \\
Z_2 &= a_{12} V_1 + a_{22} V_2 \\
&\vdots \\
Z_j &= a_{1j} V_1 + a_{2j} V_2 + \cdots + a_{ij} V_i + \cdots + a_{jj} V_j \\
&\vdots \\
Z_T &= a_{1T} V_1 + a_{2T} V_2 + \cdots + a_{iT} V_i + \cdots \\
&\quad + a_{jT} V_j + \cdots + a_{TT} V_T,
\end{aligned}
\tag{22}
$$

where $V_1, \ldots, V_T$ are independent standard normal random variables and a_{ij} represents the element in the ith row and the jth column of the matrix associated with the Cholesky factorization performed in Step 3.

(5) Substitute $Z_1, \ldots, Z_T$ from step 4 into T equations of the form in (1)–(3), as noted in Step 1, to generate the g and/or h distributions with the specified values of γ_3 and γ_4 and the Spearman correlations.

To demonstrate the previous steps and evaluate the proposed procedure, a comparison between the MOP and the MOM procedures is subsequently described. Specifically, the distributions in Figure 2 are used as a basis for a comparison using the specified correlation matrices in Table 2. Tables 3, 4, and 5 give the solved IC matrices for the MOM and the MOP procedures with samples of sizes 25 and 750, respectively. Note that the ICs for the MOM were computed by using the Mathematica source code as in [10, Table 1] and the ICs for the MOP are based on (21). Tables 6, 7, and 8 give the results of the Cholesky decompositions on the IC matrices, which are then used to create $Z_1, \ldots, Z_4$ with the specified ICs by making use of the formulae given in (22) of step 4 with $T = 4$. The values of $Z_1, \ldots, Z_4$ are subsequently substituted into equations of the form in (1)–(3) to produce $q_1(Z_1), \ldots, q_T(Z_T)$ for both procedures.

In terms of the simulation, a Fortran algorithm was written for both methods to generate 25,000 independent sample

TABLE 9: Skew (α_3) and kurtosis (α_4) results for the MOM procedure ($n = 25$).

Dist.	Parameter	Estimate	95% bootstrap CI	Standard error	Relative bias %
1	$\alpha_3 = 0$	0.0034	−0.0035, 0.01065	0.0036	—
	$\alpha_4 = 0$	−0.1785	−0.1896, −0.1673	0.0056	−5.93
2	$\alpha_3 = 0$	−0.0009	−0.0158, 0.0142	0.0075	—
	$\alpha_4 = 25$	1.297	1.2668, 1.3246	0.0151	−94.81
3	$\alpha_3 = 10$	1.763	1.7466, 1.7778	0.0082	−82.37
	$\alpha_4 = 1000$	3.471	3.3925, 3.5469	0.0392	−99.65
4	$\alpha_3 = 3$	1.436	1.4227, 1.4487	0.0066	−52.13
	$\alpha_4 = 21$	2.068	2.0184, 2.1203	0.0270	−90.15

TABLE 10: Skew (α_3) and kurtosis (α_4) results for the MOM procedure ($n = 750$).

Dist.	Parameter	Estimate	95% bootstrap CI	Standard error	Relative bias %
1	$\alpha_3 = 0$	−0.0003	−0.0017, 0.0013	0.0007	—
	$\alpha_4 = 0$	−0.0149	−0.0175, −0.0123	0.0013	—
2	$\alpha_3 = 0$	0.0043	−0.0043, 0.0141	0.0045	—
	$\alpha_4 = 25$	6.231	6.1678, 6.2958	0.0313	−75.08
3	$\alpha_3 = 10$	4.325	4.3022, 4.3491	0.0121	−56.75
	$\alpha_4 = 1000$	30.82	30.4145, 31.1905	0.2079	−96.92
4	$\alpha_3 = 3$	2.524	2.5145, 2.5335	0.0051	−15.87
	$\alpha_4 = 21$	10.58	10.4758, 10.6732	0.0506	−49.62

estimates for the specified parameters of (i) conventional skew (α_3), kurtosis (α_4), and the Pearson correlation; (ii) left-right tail-weight ratio (γ_3), tail-weight factor (γ_4), and the Spearman correlation. All estimates were based on sample sizes of $n = 25$ and $n = 750$. The formulae used for computing estimates of $\alpha_{3,4}$ were based on Fisher's k-statistics, that is, the formulae currently used by most commercial software packages such as SAS, SPSS, and Minitab, for computing indices of skew and kurtosis (where $\alpha_{3,4} = 0$ for the standard normal distribution). The formulae used for computing estimates of $\gamma_{3,4}$ were based on (10). Note that the estimates of percentiles were based on (17). The estimate for the Pearson correlations were based on the usual formula for the Pearson product-moment of correlation statistic and the estimate for the Spearman correlations were computed based on usual formula for the Spearman rank of correlation statistic. The estimates for the Pearson and Spearman correlations were both transformed using Fisher's z transformation. Bias-corrected accelerated bootstrapped average (median) estimates, confidence intervals (CIs), and standard errors were subsequently obtained for the estimates associated with the parameters using 10,000 resamples via the commercial software package Spotfire S+ [33]. The bootstrap results for the estimates of the medians and CIs associated with the Pearson and Spearman correlations were transformed back to their original matrices (i.e., estimates for the Pearson and Spearman correlations). Further, if a parameter (P) was outside its associated bootstrap CI, then an index of relative bias (RB) was computed for the estimate (E) as RB = $(((E − P))/P) \times 100$. Note that the small amount of bias associated with any bootstrap CI containing a parameter was considered negligible and thus not reported. The results of the simulation are reported in Tables 9, 10, 11, and 12 and are discussed in the next section.

7. Discussion and Conclusion

One of the advantages that the MOP has over the MOM is that the MOP can be much less biased when sampling is from distributions with more severe departures from normality. Moreover, inspection of the simulation results in Tables 9 and 12 clearly indicates this result. That is, the superiority that the MOP estimates (γ_3, γ_4) have over their corresponding MOM counterparts (α_3, α_4). For example, with samples size of $n = 25$, the estimates of skew and kurtosis for Distribution 3 (heavy-skewed and heavy-tailed) were, on average, 82.37% and 99.65% below their associated population parameters, whereas the estimates of γ_3 and γ_4 were 3.02% and 11.06% over their respective parameters. It is also evident from comparing Tables 9 and 12 that γ_3 and γ_4 are more efficient estimators as their relative standard errors RSE = (standard error/estimate)×100 are considerably smaller than those of the MOM estimators of skew and kurtosis. For example, with samples size of $n = 25$, in terms of Distribution 3, inspection of Tables 9 and 12 indicates RSE measures of RSE($\hat{\alpha}_3$) = 0.47% and RSE($\hat{\alpha}_4$) = 1.13% compared with RSE ($\hat{\gamma}_3$) = 0.36% and RSE($\hat{\gamma}_4$) = 0.20%. This demonstrates that γ_3 and γ_4 have more precision because they have less variance around their estimates.

Presented in Tables 13, 14, 15, and 16 are the results associated with the conventional Pearson and the Spearman correlations. Overall inspection of these tables indicates that the Spearman correlation is superior to the Pearson correlation

TABLE 11: Left-right tail-weight ratio (γ_3) and tail-weight factor (γ_4) results for the MOP procedure ($n = 25$).

Dist	Parameter	Estimate	95% bootstrap CI	Standard error	Relative bias %
1	$\gamma_3 = 1$	0.9978	0.9908, 1.0040	0.0034	—
	$\gamma_4 = 0.526307$	0.5294	0.5279, 0.5310	0.0008	0.5877
2	$\gamma_3 = 1$	1.001	0.9940, 1.0087	0.0039	—
	$\gamma_4 = 0.469319$	0.4762	0.4745, 0.4778	0.0008	1.4662
3	$\gamma_3 = 0.387801$	0.3995	0.3967, 0.4024	0.0014	3.02
	$\gamma_4 = 0.440929$	0.4895	0.4876, 0.4913	0.0010	11.06
4	$\gamma_3 = 0.432409$	0.4452	0.4418, 0.4484	0.0017	2.96
	$\gamma_4 = 0.477822$	0.4894	0.4876, 0.4913	0.0009	2.42

TABLE 12: Left-right tail-weight ratio (γ_3) and tail-weight factor (γ_4) results for the MOP procedure ($n = 750$).

Dist.	Parameter	Estimate	95% bootstrap CI	Standard error	Relative bias %
1	$\gamma_3 = 1$	1	0.9992, 1.0014	0.0005	—
	$\gamma_4 = 0.526307$	0.5264	0.5261, 0.5267	0.0002	—
2	$\gamma_3 = 1$	0.9994	0.9980, 1.0009	0.0007	—
	$\gamma_4 = 0.469319$	0.4696	0.4693, 0.4699	0.0002	—
3	$\gamma_3 = 0.387802$	0.3882	0.3876, 0.3887	0.0003	—
	$\gamma_4 = 0.440929$	0.4746	0.4742, 0.4749	0.0002	7.64
4	$\gamma_3 = 0.432409$	0.4326	0.4321, 0.4332	0.0003	—
	$\gamma_4 = 0.477822$	0.4784	0.4780, 0.4788	0.0002	—

TABLE 13: The Pearson correlation (corr.) results for the MOM procedure ($n = 25$).

Corr. parameter	Estimate	95% bootstrap CI	Standard error	Relative bias %
0.40	0.4188	0.4158, 0.4212	0.0016	4.69
0.60	0.7092	0.7077, 0.7108	0.0016	18.19
0.65	0.6879	0.6862, 0.6893	0.0015	5.83
0.50	0.5738	0.5718, 0.5760	0.0016	14.76
0.70	0.7347	0.7331, 0.7364	0.0018	4.96
0.60	0.6580	0.6558, 0.6605	0.0021	9.66

TABLE 14: The Pearson correlation (corr.) results for the MOM procedure ($n = 750$).

Corr. parameter	Estimate	95% bootstrap CI	Standard error	Relative bias %
0.40	0.4017	0.4012, 0.4022	0.0003	0.43
0.60	0.6337	0.6330, 0.6342	0.0006	5.61
0.65	0.6544	0.6541, 0.6548	0.0003	0.68
0.50	0.5193	0.5187, 0.5200	0.0005	3.86
0.70	0.7047	0.7042, 0.7051	0.0005	0.67
0.60	0.6173	0.6164, 0.6181	0.0007	2.88

TABLE 15: The Spearman correlation (corr.) results for the MOP procedure ($n = 25$).

Corr. parameter	Estimate	95% bootstrap CI	Standard error	Relative bias %
0.40	0.4138	0.4115, 0.4169	0.0017	3.45
0.60	0.6139	0.6115, 0.6162	0.0020	2.31
0.65	0.6660	0.6646, 0.6692	0.0019	2.46
0.50	0.5129	0.5108, 0.5161	0.0017	2.57
0.70	0.7156	0.7131, 0.7169	0.0019	2.23
0.60	0.6137	0.6115, 0.6169	0.0020	2.29

TABLE 16: The Spearman correlation (corr.) results for the MOP procedure ($n = 750$).

Corr. parameter	Estimate	95% bootstrap CI	Standard error	Relative bias %
0.40	0.39995	0.3996, 0.4005	0.0003	—
0.60	0.6002	0.5998, 0.6006	0.0003	—
0.65	0.6502	0.6499, 0.6506	0.0003	—
0.50	0.5004	0.4999, 0.5008	0.0003	—
0.70	0.7005	0.7002, 0.7008	0.0003	0.072
0.60	0.6000	0.5997, 0.6004	0.0003	—
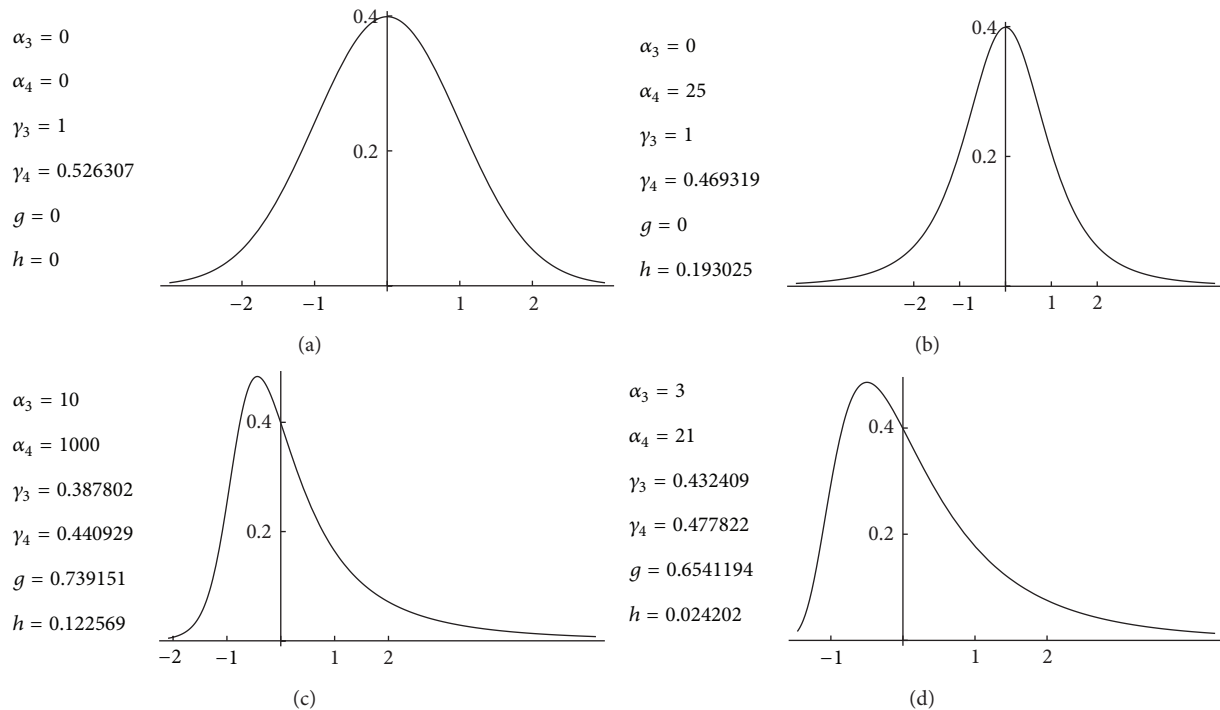
FIGURE 2: Two symmetric h distributions (a)-(b) and two asymmetric g-and-h distributions (c)-(d) with their MOM and MOP parameters of skew (α_3) and left-right tail-weight ratio (γ_3), kurtosis (α_4) and tail-weight factor (γ_4), and corresponding shape parameters for (3) and (1).

in terms of RB. For example, with samples of size $n = 25$, the RB for the two heavy-tailed distributions (i.e., distributions 3 and 4) was 9.66% for the Pearson correlation compared with only 2.29% for the Spearman correlation. Further, for large sample sizes ($n = 750$), the Spearman correlation bootstrap CIs contained most of the population parameters, whereas the Pearson correlation CIs contained none of the parameters. It is also noted that the variability of the Spearman correlation appears to be more stable than that of the Pearson correlation both within and across the different conditions.

In summary, the proposed MOP procedure is an attractive alternative to the traditional MOM procedure. In particular, the MOP procedure has distinct advantages when distributions with large departures from normality are used.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] J. Algina, H. J. Keselman, and R. D. Penfield, "Confidence interval coverage for Cohen's effect size statistic," *Educational and Psychological Measurement*, vol. 66, no. 6, pp. 945–960, 2006.

[2] J. Algina, H. J. Keselman, and R. D. Penfield, "Confidence intervals for an effect size when variances are not equal," *Journal of Modern Applied Statistical Methods*, vol. 5, pp. 2–13, 2006.

[3] S. G. Badrinath and S. Chatterjee, "A data-analytic look at skewness and elongation in common-stock return distributions," *Journal of Business and Economic Statistics*, vol. 9, pp. 223–233, 1991.

[4] D. J. Dupuis and C. A. Field, "Large wind speeds: modeling and outlier detection," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 9, pp. 105–121, 2004.

[5] C. Field and M. G. Genton, "The multivariate g-and-h distribution distribution," *Technometrics*, vol. 48, no. 1, pp. 104–111, 2006.

[6] G. M. Goerg, *The Lambert Way to Gaussianize skewed, heavy tailed data with the inverse of Tukey's h transformation as a special case [Ph.D. thesis]*, Cornell University, 2011.

[7] D. Guegan and B. Hassani, "A modified Panjer algorithm for operational risk capital calculations," *Journal of Operational Risk*, vol. 4, pp. 53–72, 2009.

[8] T. C. Headrick, R. K. Kowalchuk, and Y. Sheng, "Parametric probability densities and distribution functions for Tukey *g*-and-*h* transformations and their use for fitting data," *Applied Mathematical Sciences*, vol. 2, no. 9–12, pp. 449–462, 2008.

[9] D. C. Hoaglin, F. Mosteller, and J. W. Tukey, *Summarizing Shape Numerically: the g-and-h Distributions*, Wiley, New York, NY, USA, 1985.

[10] R. K. Kowalchuk and T. C. Headrick, "Simulating multivariate *g*-and-*h* distributions," *The British Journal of Mathematical and Statistical Psychology*, vol. 63, no. 1, pp. 63–74, 2010.

[11] H. J. Keselman, R. K. Kowalchuk, and L. M. Lix, "Robust nonorthogonal analyses revisited: an update based on trimmed means," *Psychometrika*, vol. 63, pp. 145–163, 1998.

[12] H. J. Keselman, L. M. Lix, and R. K. Kowalchuk, "Multiple comparison procedures for trimmed means," *Psychological Methods*, vol. 3, pp. 123–141, 1998.

[13] J. Martinez and B. Iglewicz, "Some properties of the Tukey *g*-and-*h* family of distributions," *Communications in Statistics A*, vol. 13, no. 3, pp. 353–369, 1984.

[14] S. Morgenthaler and J. W. Tukey, "Fitting quantiles: doubling, *HR*, *HQ*, and *HHH* distributions," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 180–195, 2000.

[15] J. W. Tukey, "Modern techniques in data analysis," in *NSF Sponsored Regional Research Conference at Southern Massachusetts University*, North Dartmouth, Mass, USA, 1977.

[16] R. R. Wilcox, "Comparing the slopes of two independent regression lines when there is complete heteroscedasticity," *British Journal of Mathematical and Statistical Psychology*, vol. 50, pp. 309–317, 1997.

[17] R. R. Wilcox, "Detecting nonlinear associations, plus comments on testing hypotheses about the correlation coefficient," *Journal of Educational and Behavioral Statistics*, vol. 26, pp. 73–83, 2001.

[18] D. C. Hoaglin, *Encyclopedia of Statistical Sciences*, Wiley, New York, NY, USA, 1983.

[19] F. A. Hodis, T. C. Headrick, and Y. Sheng, "Power method distributions through conventional moments and *L*-moments," *Applied Mathematical Sciences*, vol. 6, no. 41-44, pp. 2159–2193, 2012.

[20] T. C. Headrick, "A characterization of power method transformations through *L*-moments," *Journal of Probability and Statistics*, vol. 2011, Article ID 497463, 22 pages, 2011.

[21] T. C. Headrick and M. D. Pant, "Characterizing Tukey *h* and *hh*-distributions through *L*-moments and the *L*-correlation," *ISRN Applied Mathematics*, vol. 2012, Article ID 980153, 20 pages, 2012.

[22] T. C. Headrick and M. D. Pant, "Simulating non-normal distributions with specified *L*-moments and *L*-correlations," *Statistica Neerlandica*, vol. 66, no. 4, pp. 422–441, 2012.

[23] M. D. Pant and T. C. Headrick, "A method for simulating Burr Type III and Type XII distributions through *L*-moments and *L*-correlations," *ISRN Applied Mathematics*, vol. 2013, Article ID 191604, 14 pages, 2013.

[24] Z. A. Karian and E. J. Dudewicz, "Comparison of gld fitting methods: superiority of percentile fits to moments in L^2 norm," *Journal of the Iranian Statistical Society*, vol. 2, pp. 171–187, 2003.

[25] E. J. Dudewicz and Z. A. Karian, "Fitting the generalized lambda distribution (GLD) system by a method of percentiles, II: tables," *American Journal of Mathematical and Management Sciences*, vol. 19, pp. 1–73, 1999.

[26] Z. A. Karian and E. J. Dudewicz, "Fitting the generalized lambda distribution to data: a method based on percentiles," *Communications in Statistics: Simulation and Computation*, vol. 28, pp. 793–819, 1999.

[27] Z. A. Karian and E. J. Dudewicz, *Handbook of Fitting Statistical Distributions with R*, Chapman & Hall, 2011.

[28] T. C. Headrick, *Statistical Simulation: Power Method Polynomials and Other Transformations*, Chapman & Hall/CRC, 2010.

[29] IBM Corp. IBM SPSS Statistics for Windows, version 20.0. Armonk, NY, USA, IBM Corp., 2011.

[30] T. C. Headrick and M. D. Pant, "On the order statistics of standard normal-based power method distributions," *ISRN Applied Mathematics*, vol. 2012, Article ID 945627, 13 pages, 2012.

[31] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, John Wiley, New York, NY, USA, 1994.

[32] P. A. P. Moran, "Rank correlation and product-moment correlation," *Biometrika*, vol. 35, pp. 203–206, 1948.

[33] TIBCO Software, *TIBCO Spotfire S+ 8.1 for Windows*, Palo Alto, Calif, USA, 2008.

The Scientific World Journal