DOCUMENT RESUME

ED 463 305 TM 033 737

AUTHOR	Headrick, Todd C.; Beasley, T. Mark
TITLE	A Method for Simulating Correlated Non-Normal Systems of Statistical Equations.
PUB DATE	2002-04-00
NOTE	36p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Correlation; *Equations (Mathematics); Monte Carlo Methods; *Simulation; Statistical Analysis
IDENTIFIERS	*Nonnormal Distributions

ABSTRACT

Real world data often fail to meet the underlying assumptions of normal statistical theory. Many statistical procedures in the psychological and educational sciences involve models that may include a system of statistical equations with non-normal correlated variables (e.g., factor analysis, structural equation modeling, or other complex applications of the general linear model). Monte Carlo techniques were used to test the appropriateness of statistical procedures when the underlying assumptions of these procedures are violated. There is a paucity of methods for generating systems of statistical equations in a simple and efficient manner. Thus, the focus of the current study was to derive a general procedure for simulating correlated non-normal systems of statistical equations with a focus on computational efficiency. The procedure allows for the systematic control of correlated non-normal: (1) stochastic disturbance terms; (2) independent variables; and (3) dependent and independent variables within a system. A numerical example is provided to demonstrate the procedure. The results of a Monte Carlo simulation are provided to demonstrate that the proposed method generates the desired population parameters and intercorrelations. Two appendixes illustrate the derived method. (Contains 1 figure, 3 tables, and 42 references.) (Author/SLD)

A Method for Simulating Correlated Non-normal Systems of Statistical Equations

Todd C. Headrick
Southern Illinois University-Carbondale
Mail Code 4618
222-J Wham Building
Carbondale, IL 62901 USA
headrick@siu.edu

T. Mark Beasley
University of Alabama-Birmingham
343-C Ryals Public Health Building
1665 University BLVD
Birmingham AL 35294 USA
mbeasley@uab.edu

Paper presented at the 2002 Annual Meeting of the American Educational Research Association: New Orleans.

A Method for Simulating Correlated Non-normal Systems of Statistical Equations

Real world data often fail to meet the underlying assumptions of normal statistical theory. Many statistical procedures in the psychological and educational sciences involve models that may include a system of statistical equations with non-normal correlated variables (e.g., factor analysis, structural equation modeling, or other complex applications of the general linear model). Monte Carlo techniques are used to test the appropriateness of statistical procedures when the underlying assumptions of these procedures are violated. There is a paucity of methods for generating systems of statistical equations in a simple an efficient manner. Thus, the focus of the current study is to derive a general procedure for simulating correlated non-normal systems of statistical equations with a focus on computational efficiency. The procedure allows for the systematic control of correlated non-normal (a) stochastic disturbance terms, (b) independent variables, and (c) dependent and independent variables within a system. A numerical example is provided to demonstrate the procedure. The results of a Monte Carlo simulation are provided to demonstrate that the proposed method generates the desired population parameters and intercorrelations.

A Method for Simulating Correlated Non-normal Systems of Statistical Equations

1. Introduction

It has been documented that data sets in the psychological and educational sciences often violate the usual parametric assumptions underlying normal curve theory (Blair, 1981; Bradley, 1968, 1982; Micceri, 1989; Pearson & Please, 1975). Further, many variables of interest (e.g., reaction time) are intrinsically non-normal (Miller, 1988; Zumbo & Coulombe, 1997). In view of these concerns, behavioral and other applied researchers have relied on the results of Monte Carlo studies to aid in the proper application of various statistical techniques. For example, Monte Carlo methods may be used to compare the small sample properties of a test statistic with its competitor(s) or whether these properties are consistent with the statistic's asymptotic approximation (e.g., Headrick & Rotou, 2001; Headrick & Sawilowsky, 2000).

With the advances made in quantitative methods, and the availability of the modern desktop computer, Monte Carlo methods are widely applicable to many areas of statistical research. As a result, more sophisticated methods of simulating data have become available for conducting Monte Carlo studies. For example, Markov chain Monte Carlo methods (e.g., the Gibbs or slice sampler, Gelfand & Smith, 1990; Robert & Casella, 1999) are commonly used to generate posterior distributions to carry out Bayesian analyses in the context of learning or item response theories (e.g., Albert, J. H., 1992; Verguts & De Boeck, 2000). Other applications of modern computer intensive techniques include: the method of approximate bootstrap confidence (ABC) intervals (Efron & Tibishirani, 1998); network-based direct Monte Carlo sampling in the context of conditional logistic regression for evaluating drug withdrawal symptoms (Mehta, Patel, & Senchaudhuri, 2000); and likelihood inference with missing data (Gilks, Richardson, & Spiegelhalter, 1998).

With this plethora of uses for Monte Carlo methods in mind, there may be occasions when it is desirable to investigate the properties of statistics that involve systems of statistical equations under a variety of conditions. For example, latent factor theory of intelligence (e.g., Spearman, 1904) was the impetus for the development of factor analytic methods. The subsequent generalization of factor theory (Thurstone, 1940) resulted in many applications in the social sciences including its common use in the development and validation of psychometric scales (e.g., personality measures). However, the theory underlying these applications was based on normal curve theory that also included the assumption of uncorrelated disturbance terms.

Developments in structural equation modeling (SEM) allow researchers to model correlated disturbance structures. Most SEM software packages use maximum likelihood estimation (MLE) procedures as a default option. Unfortunately, MLE has been demonstrated to be extremely sensitive to departures from multivariate normality (Muthén & Kaplan, 1985). For example, suppose the simple factor analysis model depicted in Figure 1 represents a two-factor model of intelligence where the two latent factors (e.g., performance and verbal intelligence) are assumed to follow normal distributions. In practice, the six manifest variables (Y's) used to measure these constructs may not have normal distributions. Further, the errors from these six tests may also be non-normal and correlated due to response bias. Thus, the factor analytic model in Figure 1 is an example of a system of equations where it may be desired to simulate (a) correlated non-normal manifest variables (Y's), (b) correlated latent variables having normal distributions (X's), and (c) correlated non-normal error terms (ε's).

More generally, consider the p-th equation from a system of T equations as follows:

$$\mathbf{y}_p = \mathbf{x}_p \boldsymbol{\beta}_p + \boldsymbol{\sigma}_p \boldsymbol{\varepsilon}_p , \qquad (1)$$

where $\mathbf{y}_p$ and $\boldsymbol{\varepsilon}_p$ have dimension $(N \times 1)$, $\mathbf{x}_p$ is $(N \times k_p)$, $\boldsymbol{\beta}_p$ is $(k_p \times 1)$, and $\boldsymbol{\sigma}_p$ is a real positive scalar. Combining all T equations of the form in (1) yields a linear system that can generally be expressed as:

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\sigma}\boldsymbol{\varepsilon} , \qquad (2)$$

where $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ have dimension $(TN \times 1)$, $\mathbf{x}$ is $(TN \times K)$, $\boldsymbol{\beta}$ is $(K \times 1)$, where $K = \sum_{p=1}^{T} k_p$, and $\boldsymbol{\sigma}$ represents T scalars associated with each of the T equations. The stochastic disturbances ($\boldsymbol{\varepsilon}$) in (2) are also assumed to have expected values of zero and unit variances.

If the disturbance terms in (2) are contemporaneously correlated (e.g., $\boldsymbol{\varepsilon}_p$ is correlated with $\boldsymbol{\varepsilon}_q$) then a gain in efficiency can be achieved by testing the system jointly using the method of generalized least squares (GLS) (e.g., Judge, Hill, Griffiths, Lutkepohl, & Lee, 1985, p. 447). This approach of joint estimation is perhaps better known as "seemingly unrelated regression equation estimation" (Zellner, 1962). Moreover, the stronger the correlations are between the disturbance terms (or the weaker the correlations are between the independent variables) within (2), the greater the efficiency of GLS relative to ordinary least squares (OLS) (Dwivedi & Srivastava, 1978). Thus, it may be desirable to study the relative Type I error and power properties of the

OLS and GLS estimators under non-normal conditions. Such an investigation would usually be carried out using Monte Carlo methods. To determine any advantages of GLS relative to OLS, a variety of non-normal distributions with varying degrees of correlation between the disturbances would usually be included in the Monte Carlo study.

Generalized linear models (GLMs) and nonparametric tests have been suggested as alternatives to OLS when the stochastic disturbance populations are non-normal. In terms of GLMs, the disturbances can be specified to be some known continuous distribution functions (e.g., exponential) (Hilbe, 1994). As such, statisticians conducting Monte Carlo studies could investigate the validity and robustness of test statistics for various GLMs, including OLS regression (a GLM with an identity link and normal conditional distributions), when the error distributions have been misspecified. For example, suppose the disturbance populations in (2) follow a gamma distribution with parameters that yield typical fan-shaped heteroscedastic patterns. Under these conditions, one could compare OLS and nonparametric (rank) regression procedures with a GLM using various gamma distributions specified as the stochastic disturbance vector in (2). Additionally, if the disturbances in (2) are non-normal and correlated, then a GLM with generalized estimating equations (GEEs) could also be included in the study.

GEEs have gained considerable attention as a technique for analyzing data with dependent non-normal errors (Liang & Zeger, 1986). In terms of the error component in (2), the GEEs are GLS matrices that model their correlational structures while the GLM specifies the distributional shapes of these terms (Horton & Lipsitz, 1999). Other statistical analyses that may involve systems of statistical equations in Monte Carlo studies include: heirachical linear models; time series analysis; and other applications of the GLM.

Most statistics textbook authors discuss the validity of linear models or test statistics in terms of the various assumptions concerning the stochastic disturbance populations (e.g., Cook & Weisberg, 1999; Neter, Kutner, Nachtsheim & Wasserman, 1996). For example, the usual OLS regression procedure has the assumptions that the stochastic disturbance terms be independent and normally distributed with conditional expectations of zero and constant variances. As such, in order to examine the properties of systems of statistical equations using Monte Carlo methods, it is necessary to have an appropriate data generation procedure that would allow for the *a priori* specification of the distributional shapes and correlation structures of the stochastic disturbance populations (such as the vector $\boldsymbol{\varepsilon}$ in equation 2). Further, it is desirable that this procedure be both efficient and general enough to enable the simulation of a variety of statistical problems that may arise e.g., autocorrelation, multicollinearity, non-normality, unequal variances or regression slopes, and other violations of assumptions.

There are procedures that will simulate correlated non-normal distributions (e.g., Headrick & Sawilowsky, 1999; Vale & Maurelli, 1983). These procedures are able to generate *k* correlated non-normal variables for a single equation as in (1) or a system of independent equations. However, these procedures have limitations with respect to their ability to generate a system of correlated non-normal statistical equations. This can be demonstrated by considering six non-normal distributions $X_1,\ldots,X_6$ generated with zero means and unit variances from the Headrick and Sawilowsky (1999) procedure where each of the *X*'s has a unique non-normal distribution and a unique pairwise correlation with the other variables. To form a system of two regression equations, let X_1 and X_2 represent the two dependent measures with X_3 and X_4 as the independent variables for the first equation and X_5 and X_6 as the independent variables for the second equation. Further, let u_1 and u_2 represent the resulting error terms from the OLS regression of the dependent measures on the independent variables for this system.

This approach to creating a system of statistical equations presents a problem if control is desired over u_1 and u_2 in terms of their distributional shapes and correlation. Specifically, the resulting skew, kurtosis, and correlation between the error terms would be difficult to analytically determine. Moreover, the variances of u_1 and u_2 would also be unequal. These problems are exacerbated when more equations are introduced into the system, which may also include various specified degrees of correlation and non-normality.

Another problem with respect to the Headrick and Sawilowsky (1999) and Vale and Maurelli (1983) multivariate power methods is that these procedures are extensions of the Fleishman (1978) univariate power method. As such, not all non-normal (marginal) distributions can be generated in terms of the various possible combinations of skew and kurtosis. For example, the lower boundary point of kurtosis for symmetric distributions is -1.15132 (Headrick & Sawilowsky, 2000). Thus, it is not theoretically possible to simulate a (rectangular) uniform density using these procedures. (See Headrick & Sawilowsky, 2000, for a more complete discussion on the derivation of the boundary for the Fleishman coefficient model.)

To circumvent the aforementioned problem, Headrick (2000) derived a polynomial transformation for the power methods that allows for the additional control of the fifth and sixth moments. As a result, a much larger family of distributions is possible to simulate in terms of the available combinations of skew and kurtosis. Further, simple families of distributions also exist in terms of various combinations of fourth and sixth standardized cumulants.

2. Purpose of the Study

There is a paucity of methods for simulating systems of correlated statistical equations that enable the evaluation of more modern sophisticated statistical procedures as described above. Thus, the purpose of this study is to derive a general procedure that allows for the generation of systems of statistical equations with correlated non-normal distributions with the *least amount of difficulty*. More specifically, the objectives are to (a) extend the Headrick (2000) polynomial transformation to develop a method that generates systems of statistical equations with correlated non-normal dependent and independent variables and correlated non-normal stochastic disturbance terms, and (b) provide *Mathematica* (Wolfram, version 4.0, 1999) notebooks (available from the first author) that solve for power constants, intermediate correlations, and slope coefficients for implementing the procedure.

The procedure is derived generally for simulating a system of T equations. A numerical example is subsequently provided to demonstrate the procedure. The results of a Monte Carlo simulation are also provided to demonstrate that the proposed procedure generates the desired population parameters and intercorrelations.

3. Mathematical Development

Consider the T equations in (2) more explicitly as follows:

$$Y_1 = \beta_{10} + \beta_{11}X_{11} + \cdots + \beta_{1i}X_{1i} + \cdots + \beta_{1j}X_{1j} + \cdots + \beta_{1k}X_{1k} + \sigma_1\varepsilon_1, \qquad (3a)$$

$$\vdots$$

$$Y_p = \beta_{p0} + \beta_{p1}X_{p1} + \cdots + \beta_{pi}X_{pi} + \cdots + \beta_{pj}X_{pj} + \cdots + \beta_{pk}X_{pk} + \sigma_p \varepsilon_p , \qquad (3b)$$

$$Y_q = \beta_{q0} + \beta_{q1}X_{q1} + \cdots + \beta_{qi}X_{qi} + \cdots + \beta_{qj}X_{qj} + \cdots + \beta_{qk}X_{qk} + \sigma_q \varepsilon_q \text{, and} \qquad (3c)$$

$$\vdots$$

$$Y_T = \beta_{T0} + \beta_{T1}X_{T1} + \cdots + \beta_{Ti}X_{Ti} + \cdots + \beta_{Tj}X_{Tj} + \cdots + \beta_{Tk}X_{Tk} + \sigma_T \varepsilon_T . \qquad (3d)$$

The proposed procedure creates $Y_1, \ldots, Y_T$ from the right-hand sides of (3a) through (3d). That is, for all T equations, the Y variables are linear combinations of randomly generated X and ε terms. Because each of the disturbance terms (ε) has unit variance, the scalar terms (σ) are included to allow for the creation of equal or unequal variance conditions. It should also be noted that it is not necessary for each of these equations to have the same number of independent variables. (Eliminating a particular independent variable in the system can be accomplished by setting its associated slope coefficient to zero.)

Because the independent variables and stochastic disturbance terms in (3a) through (3d) are generated and correlated in the same manner, we have selected the independent variables X_{pi} and X_{pj} in (3b) to refer to in deriving the proposed method. Correlations between ε_p and ε_q or between X_{pi} and X_{qi} in (3b) and (3c) can be created in an analogous manner with the method presented below.

More specifically, the X_{pi} and X_{pj} in (3b) and 3c) are generated and correlated using the fifth-order polynomial transformation derived by Headrick (2000, Equation 16) as follows:

$$X_{pi} = c_{0pi} + c_{1pi}X'_{pi} + c_{2pi}X'^{2}_{pi} + c_{3pi}X'^{3}_{pi} + c_{4pi}X'^{4}_{pi} + c_{5pi}X'^{5}_{pi} \text{, and} \qquad (4a)$$

$$X_{pj} = c_{0pj} + c_{1pj}X'_{pj} + c_{2pj}X'^{2}_{pj} + c_{3pj}X'^{3}_{pj} + c_{4pj}X'^{4}_{pj} + c_{5pj}X'^{5}_{pj} , \qquad (4b)$$

where X'_{pi} and X'_{pj} ~ N(0,1). The constant coefficients ($c_{0**},\ldots,c_{5**}$) in (4a) and (4b) are determined by simultaneously solving equations (37) through (42) from Headrick (2000, Appendix 1) such that X_{pi} and X_{pj} have zero means, unit variances, and the third (γ_1), fourth (γ_2), fifth (γ_3), and sixth (γ_4) standardized cumulants from desired probability density functions.

The values of X'_{pi} and X'_{pj} in (4a) and (4b) that are used to create the independent variables (X_{pi}, X_{pj}) in (3b) are correlated at an intermediate level according to the following Lemma:

Lemma 1. Let r_{pi} be real-valued where $|r_{pi}| \in [0, 1] \, \forall_{pi=p0,pk}$, and let Z_1, V, $W_{p1},\ldots,W_{pk}$ ~ iid N(0,1). Further, let $Z_{t+1} = r_{p0}Z_1 + V\sqrt{1-r_{p0}^2}$, where $t=1$ if $r_{p0} < 1$, and $t=0$ if $r_{p0} = 1$. If $X'_{pi} = r_{pi}Z_{t+1} + W_{pi}\sqrt{1-r_{pi}^2}$ and $X'_{pj} = r_{pj}Z_1 + W_{pj}\sqrt{1-r_{pj}^2}$, then X'_{pi} and X'_{pj} ~ N(0,1), with correlation of $\rho_{X'_{pi}X'_{pj}} = r_{p0}r_{pi}r_{pj}$ when $t=1$, and $\rho_{X'_{pi}X'_{pj}} = r_{pi}r_{pj}$ when $t=0$. In particular, note that $\rho_{X'_{pi}X'_{pj}} = r_{p0}$ when $t = r_{pi} = r_{pj} = 1$, and $\rho_{X'_{pi}X'_{pj}} = r^2$ when $r_{pi} = r_{pj}$ and for when $t=0$.

Proof. The result is a consequence the proof of Lemma 1 from Headrick and Sawilowsky (1999).

Note that Lemma 1 enables, for example, the i-th independent variable for each of the T equations in a given system to be the same. This can be shown, from a more general use of Lemma 1, as a special case of where $r_{p0} = r_{pi} = r_{qi} = 1$.

The intermediate correlation between X'_{pi} and X'_{pj} ($\rho_{X'_{pi}X'_{pj}}$) in Lemma 1 is determined by setting the following equation from Headrick (2000, Equation 26) to the desired post-correlation between X_{pi} and X_{pj} ($\rho_{X_{pi}X_{pj}}$) in (4a) and (4b):

$$
\begin{aligned}
\rho_{X_{pi}X_{pj}} &= 3c_{4pi}c_{0pj} + 3c_{4pi}c_{2pj} + 9c_{4pi}c_{4pj} + c_{0pi}(c_{0pj} + c_{2pj} + 3c_{4pj}) + \\
& c_{1pi}c_{1pj}\rho_{X'_{pi}X'_{pj}} + 3c_{3pi}c_{1pj}\rho_{X'_{pi}X'_{pj}} + 15c_{5pi}c_{1pj}\rho_{X'_{pi}X'_{pj}} + 3c_{1pi}c_{3pj}\rho_{X'_{pi}X'_{pj}} + \\
& 9c_{3pi}c_{3pj}\rho_{X'_{pi}X'_{pj}} + 45c_{5pi}c_{3pj}\rho_{X'_{pi}X'_{pj}} + 15c_{1pi}c_{5pj}\rho_{X'_{pi}X'_{pj}} + 45c_{3pi}c_{5pj}\rho_{X'_{pi}X'_{pj}} + \\
& 225c_{5pi}c_{5pj}\rho_{X'_{pi}X'_{pj}} + 12c_{4pi}c_{2pj}\rho^2_{X'_{pi}X'_{pj}} + 72c_{4pi}c_{4pj}\rho^2_{X'_{pi}X'_{pj}} + 6c_{3pi}c_{3pj}\rho^3_{X'_{pi}X'_{pj}} + \\
& 60c_{5pi}c_{3pj}\rho^3_{X'_{pi}X'_{pj}} + 60c_{3pi}c_{5pj}\rho^3_{X'_{pi}X'_{pj}} + 600c_{5pi}c_{5pj}\rho^3_{X'_{pi}X'_{pj}} + 24c_{4pi}c_{4pj}\rho^4_{X'_{pi}X'_{pj}} + \\
& 120c_{5pi}c_{5pj}\rho^5_{X'_{pi}X'_{pj}} + c_{2pi}(c_{0pj} + c_{2pj} + 3c_{4pj} + 2c_{2pj}\rho^2_{X'_{pi}X'_{pj}} + 12c_{4pj}\rho^2_{X'_{pi}X'_{pj}}).
\end{aligned}
\tag{5}
$$

The purpose of correlating X'_{pi} and X'_{pj} at the intermediate level, $\rho_{X'_{pi}X'_{pj}}$, is to control for the non-normalization effect of the constants in (4a) and (4b) such that the resulting X_{pi} and X_{pj} have the desired post-correlation.

Using Lemma 1 more generally, the following theorem can be stated:

Theorem 1. If X_{pi}, X_{pj}, X_{qi}, X_{qj}, ε_p, ε_q have zero means and unit variances from equations of the form in (4a) and (4b), specified correlations $\rho_{X_{pi}X_{pj}}$, $\rho_{X_{qi}X_{qj}}$, $\rho_{X_{pi}X_{qi}}$, $\rho_{\varepsilon_p\varepsilon_q}$ from equations of the form in (5), $\mathrm{cov}(\varepsilon_p, X_{pi}) = 0$, and $\mathrm{cov}(\varepsilon_q, X_{qi}) = 0$, $\forall_{i=1,\ldots k}$, then the following correlations hold with respect to (3b) and (3c):

$$\rho_{Y_p X_{pi}} = \frac{\beta_{pi} + \sum_{pj \neq pi} \beta_{pj}\rho_{X_{pi}X_{pj}}}{\sqrt{\sigma_p^2 + \sum_{pi}\beta_{pi}^2 + 2\sum_{pj \neq pi}\beta_{pi}\beta_{pj}\rho_{X_{pi}X_{pj}}}}, \tag{6}$$

$$\rho_{Y_q X_{pi}} = \frac{\sum_{qi}\beta_{qi}\rho_{X_{pi}X_{qi}}}{\sqrt{\sigma_q^2 + \sum_{qi}\beta_{qi}^2 + 2\sum_{qj \neq qi}\beta_{qi}\beta_{qj}\rho_{X_{qi}X_{qj}}}}, \tag{7}$$

$$\rho_{Y_pY_q} = \frac{\sigma_p\sigma_q\rho_{\varepsilon_p\varepsilon_q} + \sum_{pi}\sum_{qi}\beta_{pi}\beta_{qi}\rho_{X_{pi}X_{qi}}}{\sqrt{\sigma_p^2 + \sum_{pi}\beta_{pi}^2 + 2\sum_{pj\neq pi}\beta_{pi}\beta_{pj}\rho_{X_{pi}X_{pj}}}\sqrt{\sigma_q^2 + \sum_{qi}\beta_{qi}^2 + 2\sum_{qj\neq qi}\beta_{qi}\beta_{qj}\rho_{X_{qi}X_{qj}}}}\text{, and} \tag{8}$$

$$\rho_{Y_p(\sigma_q\varepsilon_q)} = \frac{\sigma_p\rho_{\varepsilon_p\varepsilon_q}}{\sigma_q\sqrt{\sigma_p^2 + \sum_{pi}\beta_{pi}^2 + 2\sum_{pj\neq pi}\beta_{pi}\beta_{pj}\rho_{X_{pi}X_{pj}}}}. \tag{9}$$

Proof. See Appendix 1.

Thus, given exogenous correlations between the independent variables ($\rho_{X_{pi}X_{pj}}$), the pairwise correlation(s) between the dependent variable (Y_p) and the independent variables (X_{pi}) are determined by simultaneously solving a system of k equations (of the form in equation 6) in terms of k unknowns (β_{pk}) for the desired correlation(s) of $\rho_{Y_pX_{pi}}$. The correlations $\rho_{Y_pY_q}$, $\rho_{Y_qX_{pi}}$, and $\rho_{Y_p(\sigma_q\varepsilon_q)}$ can subsequently be determined by evaluating (7), (8), and (9) using the specified correlations and slope coefficients from (5) and (6).

4. Numerical Example

Suppose it is desired to generate a system of three equations of the form in (3b). For $k=2$ independent variables, where each disturbance term has unit variance, the equations are listed as follows:

$$Y_1 = \beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12} + \sigma_1\varepsilon_1, \tag{10}$$

$$Y_2 = \beta_{20} + \beta_{21}X_{21} + \beta_{22}X_{22} + \sigma_2\varepsilon_2, \text{ and} \tag{11}$$

$$Y_3 = \beta_{30} + \beta_{31}X_{31} + \beta_{32}X_{32} + \sigma_3\varepsilon_3, \tag{12}$$

where $\mathrm{var}(\sigma_i\varepsilon_i) = \sigma_i^2(1) = 1,\ \forall_{i=1,2,3}$.

Letting $t=0$ in Lemma 1, for this example, the algorithms to generate correlated normal deviates are as follows:

$$X'_{11} = r_{11}Z_1 + W_{11}\sqrt{1-r_{11}^2}\,, \tag{13}$$

$$X'_{12} = r_{12}Z_1 + W_{12}\sqrt{1-r_{12}^2}\,, \tag{14}$$

$$X'_{21} = r_{21}Z_1 + W_{21}\sqrt{1-r_{21}^2}\,, \tag{15}$$

$$X'_{22} = r_{22}Z_1 + W_{22}\sqrt{1-r_{22}^2}\,, \tag{16}$$

$$X'_{31} = r_{31}Z_1 + W_{31}\sqrt{1-r_{31}^2}\text{, and} \tag{17}$$

$$X'_{32} = r_{32}Z_1 + W_{32}\sqrt{1-r_{32}^2}\,. \tag{18}$$

Similarly, the algorithms to generate the correlated errors based on Lemma 1 are:

$$\varepsilon'_1 = r_1 Z'_1 + W_1\sqrt{1-r_1^2}\,, \tag{19}$$

$$\varepsilon'_2 = r_2 Z'_1 + W_2\sqrt{1-r_2^2}\text{, and} \tag{20}$$

$$\varepsilon'_3 = r_3 Z'_1 + W_3\sqrt{1-r_3^2}\,. \tag{21}$$

(In accordance with Theorem 1, note that Z_1 and Z'_1 are independent.) The resulting $X'_{11},\ldots,\varepsilon'_3$ in (13) through (21) are normally distributed with zero means, unit variances, and have the intermediate correlations according to Lemma 1.

Given this general framework, suppose it is desired to approximate the following correlated non-normal X's and ε's in (10), (11), and (12): a) $X_{11} = X_{21} = X_{31} =$ exponential, b) $X_{12} = X_{22} = X_{32} =$double exponential, and c) $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 =$Cauchy, where $\rho_{X_{11}X_{12}} = .10$, $\rho_{X_{21}X_{22}} = .35$, $\rho_{X_{31}X_{32}} = .70$; $\rho_{Y_1X_{11}} = \rho_{Y_1X_{12}} = .40$, $\rho_{Y_2X_{21}} = \rho_{Y_2X_{22}} = .50$, $\rho_{Y_3X_{31}} = \rho_{Y_3X_{32}} = .60$; and $\rho_{\varepsilon_1\varepsilon_2} = \rho_{\varepsilon_1\varepsilon_3} = \rho_{\varepsilon_2\varepsilon_3} = .40$.

The following steps are taken:

1. Obtain the constants to generate the standardized non-normal distributions (see Headrick, 2000, pg. 57). The standardized cumulants and constants are listed below in Table 1.

2. Determine the values of $r_{11},\ldots,r_{32}$, r_1, r_2, and r_3 to use in (13) through (21). This is accomplished using (5) by setting the left-hand side to the desired post-correlation ($\rho_{X_{11}X_{12}} = .10$; $\rho_{X_{21}X_{22}} = .35$; $\rho_{X_{31}X_{32}} = .70$) and substituting the constants into the right-hand side and then solving for these values. Accordingly, this yields values of $r_{11} = .350171$, $r_{12} = .321446$, $r_{21} = .657772$, $r_{22} = .597454$, $r_{31} = .937894$, $r_{32} = .831441$, and $r_1 = r_2 = r_3 = .749509$.

3. Obtain the slope coefficients to use in (10), (11), and (12). This is accomplished by simultaneously solving two equations of the form in (6) for each of the desired post-correlations ($\rho_{Y_p X_{pi}}$). The coefficients are as follows (see Appendix 2):

   $\beta_{11} = \beta_{12} = 0.431834$, $\beta_{21} = \beta_{22} = 0.466760$, and $\beta_{31} = \beta_{32} = 0.464851$.

4. Substitute the values of $r_{11},\ldots,r_{32}$, r_1, r_2, and r_3 obtained from step (2) into equations (13) through (21) to generate X'_{11}, X'_{12}, ε'_1; X'_{21}, X'_{22}, ε'_2; and X'_{31}, X'_{32}, ε'_3. The intermediate correlations are $\rho_{X'_{11}X'_{12}} = r_{11}r_{12} = .112561$, $\rho_{X'_{21}X'_{22}} = r_{21}r_{22} = .392989$, $\rho_{X'_{31}X'_{32}} = r_{31}r_{32} = .779803$, and $\rho_{\varepsilon'_1\varepsilon'_2} = \rho_{\varepsilon'_1\varepsilon'_3} = \rho_{\varepsilon'_2\varepsilon'_3} = r_1r_2 = r_1r_3 = r_2r_3 = .561764$.

5. Substitute the values of X'_{11}, X'_{12}, ε'_1; X'_{21}, X'_{22}, ε'_2; and X'_{31}, X'_{32}, ε'_3 from step (4) into equations of the form in (4a) and (4b) to generate the non-normal

deviates X_{11}, X_{12}, ε_1; X_{21}, X_{22}, ε_2; and X_{31}, X_{32}, ε_3 with the desired post-correlations.

6. Substitute X_{11}, X_{12}, ε_1; X_{21}, X_{22}, ε_2; and X_{31}, X_{32}, ε_3 from step (5) and the slope coefficients from step (3) into (10), (11), and (12) to generate the values of Y_1, Y_2, and Y_3 with the desired post-intercorrelations.

Given the specified correlations from above, the correlations between the dependent variables are follows (see Appendix 2): $\rho_{Y_1Y_2} = .369251$, $\rho_{Y_1Y_3} = .393505$, and $\rho_{Y_2Y_3} = .508267$. The correlations between $\rho_{Y_2X_{11}}$ and $\rho_{Y_2X_{12}}$, for example, are as follows (see Appendix 2): $\rho_{Y_2X_{11}} = .141997$ and $\rho_{Y_2X_{12}} = .138445$.

5. Monte Carlo Simulation

To evaluate the proposed procedure, the population parameters μ, σ^2, γ_1, γ_2, γ_3, γ_4, and the correlations between and within equations (9), (10), and (11) from the numerical example in the previous section were simulated using an algorithm coded in Fortran 77. The algorithm used subroutines NORMB1 and UNI1 from RANGEN (Blair, 1987) to generate pseudo-random normal and uniform deviates. Independent sample sizes of N = (10,10,10), (100,100,100), (1000,1000,1000), and (10000,10000,10000) were generated for simulating the specified standardized cumulants and various correlations. Values of all standardized cumulants and correlations were calculated for each repetition and then averaged across 50,000 repetitions. Thus, the average values of $\mu\,(\hat{\mu})$, σ^2 $(\hat{\sigma}^2)$, $\gamma_1(\bar{\gamma}_1)$, $\gamma_2(\bar{\gamma}_2)$, $\gamma_3(\bar{\gamma}_3)$, $\gamma_4(\bar{\gamma}_4)$, $\rho_{X_{11}X_{12}}(\bar{\rho}_{X_{11}X_{12}})$, $\rho_{X_{21}X_{22}}(\bar{\rho}_{X_{21}X_{22}})$, $\rho_{X_{31}X_{32}}(\bar{\rho}_{X_{31}X_{32}})$, $\rho_{Y_1X_{11}}$ $(\bar{\rho}_{Y_1X_{11}})$, $\rho_{Y_1X_{12}}(\bar{\rho}_{Y_1X_{12}})$, $\rho_{Y_2X_{21}}$, $(\bar{\rho}_{Y_2X_{21}})$, $\rho_{Y_2X_{22}}(\bar{\rho}_{Y_2X_{22}})$,

$\rho_{Y_3X_{31}}$ ($\overline{\rho}_{Y_3X_{31}}$), $\rho_{Y_3X_{32}}$ ($\overline{\rho}_{Y_3X_{32}}$), $\rho_{Y_1Y_2}$ ($\overline{\rho}_{Y_1Y_2}$), $\rho_{Y_1Y_3}$ ($\overline{\rho}_{Y_1Y_3}$), $\rho_{Y_2Y_3}$ ($\overline{\rho}_{Y_2Y_3}$), $\rho_{Y_2X_{11}}$ ($\overline{\rho}_{Y_2X_{11}}$), $\rho_{Y_2X_{12}}$ ($\overline{\rho}_{Y_2X_{12}}$), $\rho_{\varepsilon_1\varepsilon_2}$ ($\overline{\rho}_{\varepsilon_1\varepsilon_2}$), $\rho_{\varepsilon_1\varepsilon_3}$ ($\overline{\rho}_{\varepsilon_1\varepsilon_3}$), and $\rho_{\varepsilon_2\varepsilon_3}$ ($\overline{\rho}_{\varepsilon_2\varepsilon_3}$) computed were based on $N \times 50{,}000$ random deviates. The average values of the population parameters are reported in Table 2 and Table 3.

Inspection of Tables 2 and 3 indicate that the proposed procedure generated average values of the population parameters that were in close agreement with their respective population parameters. In terms of the simulation results concerning the average values of correlation, the procedure produced excellent agreement between these values and their associated population parameters even for sample sizes of $N = 10$.

6. Discussion

The simulation results from the previous section indicated that the proposed procedure generated the desired population parameters and specified intercorrelations. It is also beneficial to provide visual representations of some of the distributions generated by the Headrick (2000) procedure. Specifically, illustrated in Figure 2 are the relative frequency histograms of the approximations of the exponential and double exponential *pdf* s. Inspection of panels A and B in Figure 2 reveals that the proposed procedure approximated the considered theoretical densities very well.

The proposed procedure is also useful for generating other systems of equations based on the GLM. For example, the method could be used to generate *T* independent equations to investigate the statistical properties of competing nonparametric tests in the context of analysis of covariance (ANCOVA) or repeated measures.

With respect to ANCOVA, one attractive feature of the proposed method is that it has an advantage over other competing procedures (e.g., Knapp & Swoyer, 1967) to the extent that it allows for the creation of populations with unequal regression slopes while maintaining the between-group equal variance assumption. This can observed by inspecting equations (3b) and (3c) where the slope coefficient(s) could change (i.e., made unequal) while the error terms remain unchanged. Subsequent to any changes made to the slope coefficients, the variate and covariate correlations can then be determined from equation (6). Thus, this feature of the proposed method would be a remedy to the problem with the algorithms used to correlate data in the Monte Carlo studies by Hamilton (1976) and Peckham (1968). Specifically, the algorithms used in these studies were unable to simulate the unequal slope condition without also simultaneously violating the between-group equal variance assumption. (See Rogosa, 1980, for a discussion on the validity of the Hamilton, 1976, and Peckham, 1968 studies with respect to this issue.)

Many other applications of the proposed procedure to the GLM are possible. From the GLM perspective, the dependent variables (Y_p) could represent the same variable collected under T different conditions or at time points $1,\ldots,T$. In either case, the independent variables (X_{pi}) could represent static covariates (e.g., pre-existing ability measures often used in ANCOVA models). Thus, using the special case of Lemma 1 noted above, the i-th independent variable in each of the T equations would be the same. Conversely, the independent variables (X_{pi}) could also be different for each of the T equations and may be used to represent time-varying covariates (i.e., some variables measured over the T periods along with Y_p). There is also the possibility of using a combination of both static and time-varying covariates (X_{pi}). As such, a variety of analytic procedures are comparable under this data generation process.

The proposed method would also allow the generation of repeated measures data of nonspherical structures with non-normal disturbances and non-normal covariates. It should also be noted that with non-normal stochastic populations one could specify all of these distributions to be the same. This assumption is implicit in parametric analyses of repeated measures data because normal disturbance populations are implied. Likewise, robust estimators such as the rank-based shift model (Lehmann, 1998) also assume that disturbance populations have identical (but not necessarily normal) shapes.

By contrast, fully nonparametric hypotheses (e.g., Akritas & Arnold, 1994) make no assumptions about the distribution of the stochastic disturbance populations and therefore do not require the conditional (disturbance) populations to have identical variances or distributional shapes. Fully nonparametric hypotheses differ from robust estimators because they do not attribute the rejection to location parameters alone but rather to any distributional differences, a concept recently referred to as "stochastic heterogeneity" (Varga & Delaney, 1998). Thus, hypotheses of this form reduce the risk of drawing incorrect conclusions about the likely sources of a statistically significant result, but do so at the cost of not being able to characterize precisely how population distributions differ (Serlin & Harwell, 2001).

The proposed method would also be useful in the context of time series analysis. Specifically, the procedure could be used to model instrumental variables which address one of the problems with certain autoregressive models (e.g., the adaptive expectations model) where the dependent measure from a preceding time period (Y_p) is included as a stochastic independent variable in the subsequent (q-th) period. As such, the vectors Y_p and ε_q are usually correlated. Using the proposed method, a Monte Carlo study could be arranged to simulate non-normal "proxies" correlated at various levels between Y_p and ε_q. The correlations can be determined from equation (9).

Application of the proposed method to the GLM is flexible and has the potential to simulate other types of models where disturbance populations change over time. For example, when data sets with repeated measures have mistimed measures or missing data, the GLM with GEEs and hierarchical linear models are often considered preferable to the usual OLS univariate and multivariate procedures.

It may also be reasonable to consider a model where the correlation structure is a function of time between the observations. Thus, data could be also be generated for Monte Carlo studies involving dynamic regression models that have distributed lags or moving averages.

7. Conclusion

Systems of linear statistical equations with correlated non-normal variables are widely applicable in many experimental situations. Some examples include confirmatory factor analysis, hierarchical linear models, time series analysis, and other applications of the general linear model (e.g., analysis of covariance, repeated measures). Thus, the concern of the present study was to develop a procedure that enables the simulation of systems of statistical equations. The procedure allows for the creation of systems with non-normal variables and specified intercorrelations between a) dependent variables, b) independent variables, c) dependent and independent variables, and d) stochastic error terms. The results of a Monte Carlo simulation indicate that the proposed procedure generated the desired population parameters and specified intercorrelations.

References

Akritas, M. G., & Arnold, S. F. (1994). Fully nonparametric hypotheses for factorial designs I: Multivariate repeated-measures designs. *Journal of the American Statistical Association*, **89**, 336-343.

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, **17**, 251-269.

Blair, R. C. (1981). A reaction to "Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance." *Review of Educational Research*, **51**, 499-507.

Blair, R. C. (1987). *RANGEN*, Boca Raton, FL: IBM.

Bradley, J. V. (1968). *Distribution free statistical tests*. Engelwood Cliffs, NJ: Prentice-Hall.

Bradley, J. V. (1982). The insidious L-shaped distribution. *Bulletin of the Psychonomic Society*, **20**, 85-88.

Cook, R. D., & Weisberg, S. (1999). *Applied regression including computing and graphics*. New York: John Wiley.

Dwivedi, T., & Srivastava, K. (1978). Optimality of least squares in the seemingly unrelated regression model. *Journal of Economics*, **7**, 391-395.

Efron, B. & Tibshirani, R. J. (1998). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.

Fleishman, A. I. (1978). A method for simulating univariate non-normal distributions. *Psychometrika*, **43**, 521-532.

Gelfand, A. E. & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1998). *Markov chain Monte Carlo in practice*. Boca Raton, FL: Chapman & Hall/CRC.

Hamilton, B. L. (1976). A Monte Carlo test of the robustness of parametric and non-parametric analysis of covariance against unequal regression slopes. *Journal of the American Statistical Association*, **71**, 864-869.

Headrick, T. C. (2000). Simulating univariate and multivariate nonnormal distributions. *Proceedings of the Statistical Computing Section and Section on Statistical Graphics*, American Statistical Association, Alexandria, VA, 52-57.

Headrick, T. C. & Rotou, O. (2001). An investigation of the rank transformation in multiple regression. *Computational Statistics & Data Analysis*, **38**, 203-215.

Headrick, T. C., & Sawilowsky, S. S. (1999). Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method, *Psychometrika*, **64**, 25-35.

Headrick, T. C. & Sawilowsky, S. S. (2000). Properties of the rank transformation in factorial analysis of covariance. *Communications in Statistics-Simulation and Computation*, **29**, 1059-1087.

Headrick, T. C. & Sawilowsky, S. S. (2000). Weighted simplex procedures for determining boundary points and constants for the univariate and multivariate power methods. *Journal of Educational and Behavioral Statistics*, **25**, 417-436.

Hilbe, J. M. (1994). Generalized linear models. *American Statistician*, **48**, 255-265.

Horton, N. J., & Lipsitz, S. R. (1999). Review of software to fit generalized estimating equation regression models. *American Statistician*, **53**, 160-169.

Judge, G. J., Hill, R. C., Griffiths, W. E., Lutkepohl, H., & Lee, T.C. (1985). *Introduction to the theory and practice of econometrics* (2nd). New York: John Wiley.

Lehmann, E. L., (1998). *Nonparametrics: Statistical methods based on ranks* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Liang, K., Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

Knapp, T. R., & Swoyer, V. H. (1967). Some empirical results concerning the power of Bartlett's test of the significance of a correlation matrix. *American Educational Research Journal*, **4**, 13-17.

Mehta, C. R., Patel, N. R., & Senchaudhuri, P. (2000). Efficient Monte Carlo methods for conditional logistic regression. *Journal of the American Statistical Association*, **95**, 99-108.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, **105**, 156-166.

Miller, J. (1988). A warning about median reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 539-543.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, **38**, 171-189.

Neter, J., Kutner, M. H., Nachtseim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (4th). Boston, MA: WCB/McGraw-Hill.

Pearson, E. S. & Please, N. W. (1975). Relation between the shape of population distribution and the robustness of four simple test statistics. *Biometrika*, **63**, 223-241.

Peckham, P.D., (1968). *An investigation of the effects of non-homogeneity of regression slopes upon the F-test of analysis of covariance*. Unpublished doctoral dissertation, University of Colorado.

Robert, C. P., & Casella, G. (1999). *Monte Carlo statistical methods*. New York: Springer.

Rogosa, D. (1980). Comparing nonparallel regression lines. *Psychological Bulletin*, **88**, 307-321.

Serlin, R. C., & Harwell, M. R. (April, 2001). A review of nonparametric tests for complex experimental designs in educational research. Paper presented at the meeting of the American Educational Research Association. Seattle, WA.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology*, **15**, 201-293.

Thurstone, L. L. (1940). Current issues in factor analysis. *Psychological Bulletin*, **37**, 189-236.

Vale C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, **48**, 465-471.

Wolfram, S. (1999). *The Mathematica book* (4th). Wolfram-Media/Cambridge University Press.

Vargha, A., & Delaney, H. D. (1998). The Kruskal-Wallis test and stochastic homogeneity. *Journal of Educational and Behavioral Statistics*, **23**, 170-192.

Verguts, T. & De Boeck, P. (2000). A Rasch model for detecting learning while solving an intelligence test. *Applied Psychological Measurement*, **24**, 151-162.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. *Journal of the American Statistical Association*, **57**, 348-368.

Zumbo, B. D. & Coulombe, D. (1997). Investigation of robust rank-order test for non-normal populations with unequal variances: The case of reaction time. *Canadian Journal of Experimental Psychology*, **51**, 139-149.

TABLE 1. Values of $c_0,\ldots,c_5$ that were used to simulate the desired non-normal distributions[a].

Dist.	c_0	c_1	c_2	c_3	c_4	c_5
1	-0.307740	0.800560	0.318764	0.033500	-0.003675	0.000159
2	0.000000	0.727709	0.000000	0.096303	0.000000	-0.022320
3	0.000000	-0.159685	0.000000	0.355036	0.000000	-0.009473

[a]The three distributions are described as follows: (1) approximate exponential ($\gamma_1 = 2$, $\gamma_2 = 6$, $\gamma_3 = 24$, $\gamma_4 = 120$); (2) approximate double exponential ($\gamma_1 = 0$, $\gamma_2 = 3$, $\gamma_3 = 0$, $\gamma_4 = 30$); and (3) approximate Cauchy ($\gamma_1 = 0$, $\gamma_2 = 25$, $\gamma_3 = 0$, $\gamma_4 = 1500$).

TABLE 2. Values of average correlation from the simulation[a].

Population Correlation	Average Correlation	$N=10$	$N=10^2$	$N=10^3$	$N=10^4$
.4000	$\bar{\rho}_{Y_1X_{11}}$	.3999	.3998	.3998	.3999
.4000	$\bar{\rho}_{Y_1X_{12}}$	.3998	.4000	.4001	.4000
.3935	$\bar{\rho}_{Y_1Y_3}$	.3933	.3933	.3931	.3934
.5000	$\bar{\rho}_{Y_2X_{21}}$	.4998	.4998	.4999	.5000
.5000	$\bar{\rho}_{Y_2X_{22}}$	.5003	.5000	.5001	.4999
.3693	$\bar{\rho}_{Y_1Y_2}$	.3689	.3690	.3690	.3692
.1420	$\bar{\rho}_{Y_2X_{11}}$	.1421	.1420	.1421	.1419
.1384	$\bar{\rho}_{Y_2X_{12}}$	.1383	.1383	.1385	.1384
.6000	$\bar{\rho}_{Y_3X_{31}}$	.5994	.5995	.5995	.5999
.6000	$\bar{\rho}_{Y_3X_{32}}$	.5999	.5999	.6000	.5999
.5083	$\bar{\rho}_{Y_2Y_3}$	.5083	.5082	.5082	.5083
.1000	$\bar{\rho}_{X_{11}X_{12}}$	.0999	.0999	.1000	.1000
.3500	$\bar{\rho}_{X_{21}X_{22}}$	.3501	.3500	.3502	.3499
.7000	$\bar{\rho}_{X_{31}X_{32}}$	.6995	.6995	.6996	.6999
.4000	$\bar{\rho}_{\varepsilon_1\varepsilon_2}$	.3993	.3997	.3995	.4000
.4000	$\bar{\rho}_{\varepsilon_1\varepsilon_3}$	.3992	.3999	.3994	.4000
.4000	$\bar{\rho}_{\varepsilon_2\varepsilon_3}$	.3997	.3999	.4000	.4000

[a]The population parameters for the variables are: (1) $X_{11} = X_{21} = X_{31} = (\gamma_1 = 2$, $\gamma_2 = 6$, $\gamma_3 = 24$, $\gamma_4 = 120)$; (2) $X_{12} = X_{22} = X_{32} = (\gamma_1 = 0,\ \gamma_2 = 3,\ \gamma_3 = 0,\ \gamma_4 = 30)$; and (3) $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = (\gamma_1 = 0,\ \gamma_2 = 25,\ \gamma_3 = 0,\ \gamma_4 = 1500)$.

TABLE 3. Average values of mean($\hat{\mu}$), variance($\hat{\sigma}^2$), and the third($\bar{\gamma}_1$), fourth($\bar{\gamma}_2$), fifth($\bar{\gamma}_3$), and sixth($\bar{\gamma}_4$) standardized cumulants from the simulation[a].

	X_{11}	X_{12}	ε_1	X_{21}	X_{22}	ε_2	X_{31}	X_{32}	ε_3
N=10									
$\hat{\mu}$	-0.00004	-0.00038	-0.00021	0.00010	0.00007	-0.00008	-0.00034	-0.00001	-0.00008
$\hat{\sigma}^2$	0.99293	1.00042	0.99736	0.99928	1.00083	0.99856	0.99872	0.99998	1.09866
$\bar{\gamma}_1$	1.99887	-0.00352	-0.00480	1.99949	0.00023	0.00450	1.99901	0.00101	-0.00170
$\bar{\gamma}_2$	5.95550	3.00977	24.8245	5.95497	3.02343	24.8869	5.96749	2.99814	24.7938
$\bar{\gamma}_3$	23.8746	-0.07745	-0.18880	23.8247	-0.00188	0.81641	23.9470	0.01741	-0.55553
$\bar{\gamma}_4$	119.649	30.4301	1474.255	118.388	31.2483	1484.101	119.152	29.9682	1475.551
N=10^2									
$\hat{\mu}$	-0.00005	-0.00008	-0.00007	0.00002	-0.00012	-0.00001	-0.0001	-0.00001	-0.00001
$\hat{\sigma}^2$	0.99921	1.00020	0.99960	0.99918	1.00005	0.99973	0.99891	0.99990	1.00975
$\bar{\gamma}_1$	1.99896	-0.00129	-0.00321	1.99860	-0.00008	0.00633	1.99785	0.00040	0.00433
$\bar{\gamma}_2$	5.99765	3.00426	24.9644	5.98813	3.00426	24.9696	5.98946	3.00009	24.9383
$\bar{\gamma}_3$	24.0212	-0.03029	-0.38283	23.8938	0.00304	0.10304	23.9333	-0.02657	0.09315
$\bar{\gamma}_4$	120.265	30.2585	1481.675	118.601	30.2260	1495.552	119.097	30.0729	1492.868
N=10^3									
$\hat{\mu}$	-0.00007	-0.00018	-0.00024	0.00010	0.00003	-0.00013	-0.0001	0.00002	-0.00013
$\hat{\sigma}^2$	0.99915	1.00033	0.998555	0.99938	1.00042	1.00030	0.99885	1.00026	1.00128
$\bar{\gamma}_1$	1.99811	-0.00242	-0.00200	1.99967	0.00012	0.00341	1.99809	0.00029	-0.00275
$\bar{\gamma}_2$	5.99076	3.00629	24.8942	5.99425	3.01366	25.0105	5.99569	3.00495	24.9178
$\bar{\gamma}_3$	23.9727	-0.06022	-0.00207	23.9342	-0.02068	0.09304	24.0306	-0.00248	-0.07326
$\bar{\gamma}_4$	120.061	30.2127	1486.226	118.974	30.1845	1502.074	120.518	30.1534	1493.898
N=10^4									
$\hat{\mu}$	-0.00003	-0.00003	-0.00002	0.00000	-0.00002	0.00000	-0.00001	0.00001	0.00000
$\hat{\sigma}^2$	0.99984	0.99997	0.99996	0.99985	1.00002	0.99990	0.99908	1.00000	0.99999
$\bar{\gamma}_1$	1.99902	-0.00025	0.00050	2.00030	-0.00024	-0.00020	1.99803	0.00016	0.00201
$\bar{\gamma}_2$	6.00722	2.99966	24.9951	5.99153	3.00039	24.9932	5.99649	2.99952	25.0004
$\bar{\gamma}_3$	23.9842	-0.00484	0.01022	24.0127	-0.00897	0.06241	23.9822	-0.00160	0.19972
$\bar{\gamma}_4$	120.016	29.9999	1499.177	119.971	30.0030	1499.134	119.993	29.9869	1499.332

[a]The population parameters for the variables are: (1) $X_{11} = X_{21} = X_{31} = (\gamma_1 = 2,\ \gamma_2 = 6,\ \gamma_3 = 24,\ \gamma_4 = 120)$; (2) $X_{12} = X_{22} = X_{32} = (\gamma_1 = 0,\ \gamma_2 = 3,\ \gamma_3 = 0,\ \gamma_4 = 30)$; and (3) $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = (\gamma_1 = 0,\ \gamma_2 = 25,\ \gamma_3 = 0,\ \gamma_4 = 1500)$.

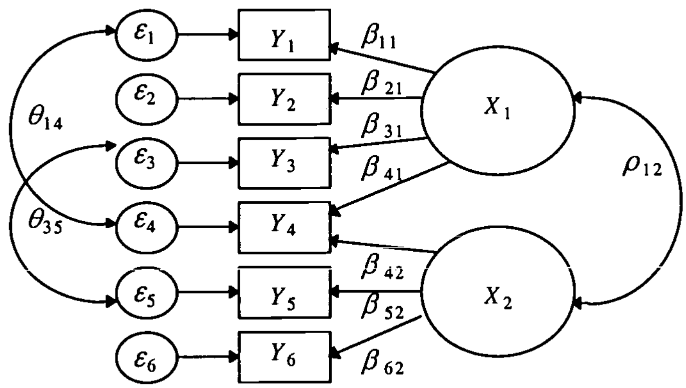
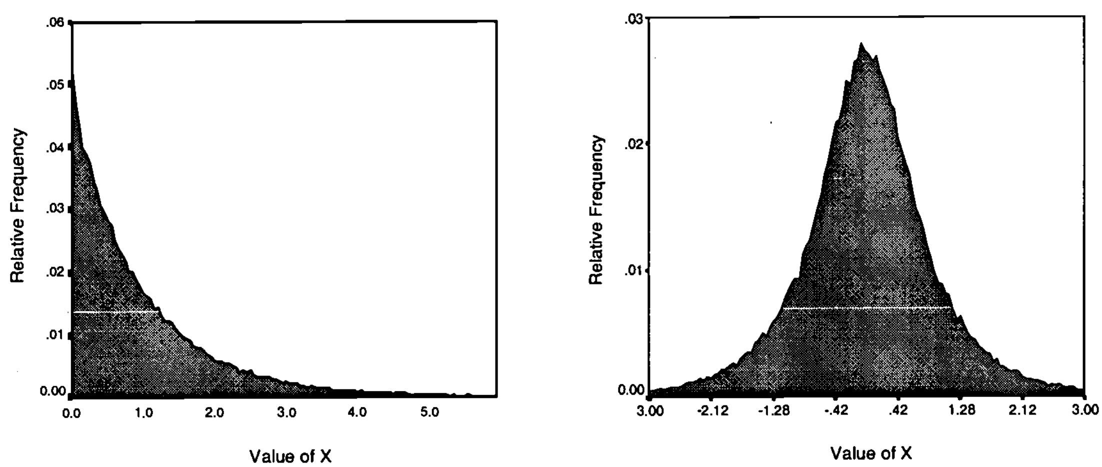


FIGURE 1. A two factor confirmatory factor analysis model with cross-loadings and correlated error terms (ε's). The parameters ρ and θ denote correlations between the X's and between the ε's.



A) Exponential

B) Double Exponential

FIGURE 2. Approximations of the exponential and double exponential pdf's generated by the Headrick (2000) polynomial transformation from equations (4a) and (4b). The constants used to simulate the densities are listed in Table 1. The sample size used was 10,000. For amenability to the standard exponential pdf, a constant of 1.0 was added to each value of X in panel A.

Appendix 1

Without loss of generality, Theorem 1 can be shown from the independent use of Lemma 1 to create X_{pi}, X_{qi}; and ε_{p}, ε_q such that $\text{cov}(\varepsilon_p, X_{pi}) = 0$ and $\text{cov}(\varepsilon_q, X_{qi}) = 0$, $\forall_{i=1,\ldots k}$, and with random variables that follow a standard normal distribution.

PROOF: $$\rho_{Y_p X_{pi}} = \frac{E[Y_p X_{pi}] - E[Y_p]E[X_{pi}]}{\{(E[Y_p^2] - (E[Y_p])^2) \times (E[X_{pi}^2] - (E[X_{pi}])^2)\}^{1/2}}, \tag{22}$$

$$\rho_{Y_q X_{pi}} = \frac{E[Y_q X_{pi}] - E[Y_q]E[X_{pi}]}{\{(E[Y_q^2] - (E[Y_q])^2) \times (E[X_{pi}^2] - (E[X_{pi}])^2)\}^{1/2}}, \tag{23}$$

$$\rho_{Y_p Y_q} = \frac{E[Y_p Y_q] - E[Y_p]E[Y_q]}{\{(E[Y_p^2] - (E[Y_p])^2) \times (E[Y_q^2] - (E[Y_q])^2)\}^{1/2}}, \text{ and} \tag{24}$$

$$\rho_{Y_p(\sigma_q \varepsilon_q)} = \frac{E[(Y_p)(\sigma_q \varepsilon_q)] - E[Y_p]E[\sigma_q \varepsilon_q]}{\{(E[Y_p^2] - (E[Y_p])^2) \times (E[(\sigma_q \varepsilon_q)^2] - (E[\sigma_q \varepsilon_q])^2)\}^{1/2}}. \tag{25}$$

Setting $c_{1**} = 1$ and all other constants $(c_{0**}, c_{2**}, \ldots, c_{5**})$ to zero in equations of the form in (4a), (4b), and (5) gives the standard normal case where $X_{pi} = X'_{pi}$, $X_{pj} = X'_{pj}$, $X_{qi} = X'_{qi}$, $X_{qj} = X'_{qj}$, $\varepsilon_p = \varepsilon'_p$, $\varepsilon_q = \varepsilon'_q$, $\rho_{X_{pi}X_{pj}} = \rho_{X'_{pi}X'_{pj}}$, $\rho_{X_{qi}X_{qj}} = \rho_{X'_{qi}X'_{qj}}$, $\rho_{X_{pi}X_{qi}} = \rho_{X'_{pi}X'_{qi}}$, and $\rho_{\varepsilon_p\varepsilon_q} = \rho_{\varepsilon'_p\varepsilon'_q}$. Further, in Lemma 1, (3b), and (3c), let all X_{pi} be a function of Z_1 and let all X_{qi} be a function of Z_2. Similarly, let ε_p be a function of Z'_1 and ε_q be a function of Z'_2, where $\text{cov}(Z_1, Z'_1) = \text{cov}(Z_2, Z'_2) = \text{cov}(V, V') = 0$.

It follows from expressing Z_2 as a function of Z_1 in Lemma 1 that:

$$E[X_{qi}] = r_{qi}E[Z_2] + E[W_{qi}]\sqrt{1 - r_{qi}^2} = r_{p0}E[Z_1] + E[V]\sqrt{1 - r_{p0}^2}) + E[W_{qi}]\sqrt{1 - r_{qi}^2} = 0,$$

because $E[Z_1] = E[V] = E[W_{qi}] = 0$.

It follows analogously that:

$$E[\sigma_q \varepsilon_q] = \sigma_q (r_{\varepsilon_q} E[Z_2'] + E[W_{\varepsilon_q}]\sqrt{1-r_{\varepsilon_q}^2})$$

$$= \sigma_q (r_{\varepsilon_q 0} E[Z_1'] + E[V']\sqrt{1-r_{\varepsilon_q 0}^2}) + E[W_{\varepsilon_q}]\sqrt{1-r_{\varepsilon_q}^2}) = 0, \text{ because}$$

$E[Z_1'] = E[V'] = E[W_{\varepsilon_q}] = 0$.

Define the variances of X_{qi} and $\sigma_q \varepsilon_q$ as:

$$\mathrm{var}[X_{qi}] = E[X_{qi}^2] - (E[X_{qi}])^2, \text{ and} \tag{26}$$

$$\mathrm{var}[\sigma_q \varepsilon_q] = \sigma_q^2 \,\mathrm{var}[\varepsilon_q] = \sigma_q^2 (E[\varepsilon_q^2] - (E[\varepsilon_q])^2), \text{ since } \sigma_q \text{ is a constant scalar.} \tag{27}$$

Applying Lemma 1 to both (26) and (27) gives:

$$\begin{aligned}\mathrm{var}[X_{qi}] = E[X_{qi}^2] = E[& r_{qi}^2 V^2 - r_{p0}^2 r_{qi}^2 V^2 + 2 r_{qi} V W_{qi} \sqrt{1-r_{p0}^2}\sqrt{1-r_{qi}^2} + W_{qi}^2 - r_{qi}^2 W_{qi}^2 + \\ & 2 r_{p0} r_{qi}^2 V Z_1 \sqrt{1-r_{p0}^2} + 2 r_{p0} r_{qi} W_{qi} Z_1 \sqrt{1-r_{qi}^2} + r_{p0}^2 r_{qi}^2 Z_1^2] - 0\end{aligned} \tag{28}$$

$$\begin{aligned}\mathrm{var}[\varepsilon_q] = E[\varepsilon_q^2] = E[& r_{\varepsilon_q}^2 V'^2 - r_{\varepsilon_q 0}^2 r_{\varepsilon_q}^2 V'^2 + 2 r_{\varepsilon_q} V' W_{\varepsilon_q} \sqrt{1-r_{\varepsilon_q 0}^2}\sqrt{1-r_{\varepsilon_q}^2} + W_{\varepsilon_q}^2 - r_{\varepsilon_q}^2 W_{\varepsilon_q}^2 + \\ & 2 r_{\varepsilon_q 0} r_{\varepsilon_q}^2 V' Z_1' \sqrt{1-r_{\varepsilon_q 0}^2} + 2 r_{\varepsilon_q 0} r_{\varepsilon_q} W_{\varepsilon_q} Z_1' \sqrt{1-r_{\varepsilon_q}^2} + r_{\varepsilon_q 0}^2 r_{\varepsilon_q}^2 Z_1'^2] - 0,\end{aligned} \tag{29}$$

because $E[X_{qi}] = E[\varepsilon_q] = 0$.

Taking expectations in (28) and (29) yields:

$\mathrm{var}[X_{qi}] = E[X_{qi}^2] = 1$, and

$\mathrm{var}[\varepsilon_q] = E[\varepsilon_q^2] = 1$, because

$E[Z_1^2] = E[Z_1'^2] = E[W_{\varepsilon_q}^2] = E[W_{qi}^2] = E[V^2] = E[V'^2] = 1$, and

$E[Z_1] = E[Z_1'] = E[W_{\varepsilon_q}] = E[W_{qi}] = E[V] = E[V'] = 0$.

Thus,

$\mathrm{var}[\sigma_q \varepsilon_q] = \sigma_q^2 E[\varepsilon_q^2] = \sigma_q^2 (1) = \sigma_q^2$.

It follows from analogous arguments that:

$E[X_{pi}] = E[\varepsilon_p] = 0,$

$\text{var}[X_{pi}] = E[X_{pi}^2] = 1$, and

$\text{var}[\sigma_p \varepsilon_p] = \sigma_p^2 E[\varepsilon_p^2] = \sigma_p^2(1) = \sigma_p^2.$

Thus,

$E[Y_p] = \beta_{p0}$ and $E[Y_q] = \beta_{q0}$, because $E[\varepsilon_p] = E[\varepsilon_q] = E[X_{qi}] = E[X_{pi}] = 0$, $\forall_{i=1,\ldots k}$.

Hence,

$$\rho_{Y_p X_{pi}} = \frac{E[Y_p X_{pi}] - (\beta_{p0})(0)}{\sqrt{E[Y_p^2] - \beta_{p0}^2}\sqrt{1}}, \tag{30}$$

$$\rho_{Y_q X_{pi}} = \frac{E[Y_q X_{pi}] - (\beta_{q0})(0)}{\sqrt{E[Y_q^2] - \beta_{q0}^2}\sqrt{1}}, \tag{31}$$

$$\rho_{Y_p Y_q} = \frac{E[Y_p Y_q] - \beta_{p0}\beta_{q0}}{\sqrt{E[Y_p^2] - \beta_{p0}^2}\sqrt{E[Y_q^2] - \beta_{q0}^2}}, \text{ and} \tag{32}$$

$$\rho_{Y_p(\sigma_q \varepsilon_q)} = \frac{E[(Y_p)(\sigma_q \varepsilon_q)] - (\beta_{p0})(0)}{\sqrt{E[Y_p^2] - \beta_{p0}^2}\sqrt{\sigma_q^2}}. \tag{33}$$

Expressing $Z_2\,(Z_2')$ as a function of $Z_1\,(Z_1')$ from Lemma 1 yields:

$$Y_p^2 = (\beta_{p0} + \sigma_p(r_{\varepsilon_p} Z_1' + W_{\varepsilon_p}\sqrt{1 - r_{\varepsilon_p}^2}) + \sum_{pi} \beta_{pi}(r_{pi} Z_1 + W_{pi}\sqrt{1 - r_{pi}^2}))^2, \tag{34}$$

$$\begin{aligned} Y_q^2 = (\beta_{q0} + \sigma_q(r_{\varepsilon_p 0} Z_1' + V'\sqrt{1 - r_{\varepsilon_p 0}^2} + W_{\varepsilon_q}\sqrt{1 - r_{\varepsilon_q}^2}) + \\ \sum_{qi} \beta_{qi}(r_{qi}(r_{p0} Z_1 + V\sqrt{1 - r_{p0}^2}) + W_{qi}\sqrt{1 - r_{qi}^2}))^2, \end{aligned} \tag{35}$$

$$\begin{aligned} Y_p X_{pi} = (\beta_{p0} + \sigma_p(r_{\varepsilon_p} Z_1' + W_{\varepsilon_p}\sqrt{1 - r_{\varepsilon_p}^2}) + \sum_{pi} \beta_{pi}(r_{pi} Z_1 + W_{pi}\sqrt{1 - r_{pi}^2})) \times \\ (r_{pi} + W_{pi}\sqrt{1 - r_{pi}^2}), \end{aligned} \tag{36}$$

$$Y_q X_{pi} = (\beta_{p0} + \sigma_q (r_{\varepsilon_p 0} Z_1' + V'\sqrt{1 - r_{\varepsilon_p 0}^2} + W_{\varepsilon_q}\sqrt{1 - r_{\varepsilon_q}^2}) + \sum_{qi} \beta_{qi} (r_{qi} (r_{p0} Z_1 + V\sqrt{1 - r_{p0}^2}) + W_{qi}\sqrt{1 - r_{qi}^2})) \times (r_{pi} + W_{pi}\sqrt{1 - r_{pi}^2}), \tag{37}$$

$$Y_p Y_q = (\beta_{p0} + \sigma_p (r_{\varepsilon_p} Z_1' + W_{\varepsilon_p}\sqrt{1 - r_{\varepsilon_p}^2}) + \sum_{pi} \beta_{pi} (r_{pi} Z_1 + W_{pi}\sqrt{1 - r_{pi}^2})) \times (\beta_{q0} + \sigma_q (r_{\varepsilon_p 0} Z_1' + V'\sqrt{1 - r_{\varepsilon_p 0}^2} + W_{\varepsilon_q}\sqrt{1 - r_{\varepsilon_q}^2}) + \sum_{qi} \beta_{qi} (r_{qi} (r_{p0} Z_1 + V\sqrt{1 - r_{p0}^2}) + W_{qi}\sqrt{1 - r_{qi}^2})), \text{ and} \tag{38}$$

$$(Y_p)(\sigma_q \varepsilon_q) = (\beta_{p0} + \sigma_p (r_{\varepsilon_p} Z_1' + W_{\varepsilon_p}\sqrt{1 - r_{\varepsilon_p}^2}) + \sum_{pi} \beta_{pi} (r_{pi} Z_1 + W_{pi}\sqrt{1 - r_{pi}^2})) \times \sigma_q (r_{\varepsilon_p 0} Z_1' + V'\sqrt{1 - r_{\varepsilon_p 0}^2} + W_{\varepsilon_q}\sqrt{1 - r_{\varepsilon_q}^2}). \tag{39}$$

Expanding the right-hand sides of (34) through (39) and taking expectations gives:

$$E[Y_p^2] = \sigma_p^2 + \beta_{p0}^2 + \sum_{pi} \beta_{pi}^2 + 2\sum_{pj \neq pi} \beta_{pi}\beta_{pj}\rho_{X_{pi}X_{pj}}, \tag{40}$$

$$E[Y_q^2] = \sigma_q^2 + \beta_{q0}^2 + \sum_{qi} \beta_{qi}^2 + 2\sum_{qj \neq qi} \beta_{qi}\beta_{qj}\rho_{X_{qi}X_{qj}}, \tag{41}$$

$$E[Y_p X_{pi}] = \beta_{pi} + \sum_{pj \neq pi} \beta_{pj}\rho_{X_{pi}X_{pj}}, \tag{42}$$

$$E[Y_q X_{pi}] = \sum_{qi} \beta_{qi}\rho_{X_{qi}X_{pi}}, \tag{43}$$

$$E[Y_p Y_q] = \sigma_p \sigma_q \rho_{\varepsilon_p \varepsilon_q} + \beta_{p0}\beta_{q0} + \sum_{pi}\sum_{qi} \beta_{pi}\beta_{qi}\rho_{X_{pi}X_{qi}}, \text{ and} \tag{44}$$

$$E[(Y_p)(\sigma_q \varepsilon_q)] = \sigma_p \rho_{\varepsilon_p \varepsilon_q}. \tag{45}$$

Substituting (40) through (45) into (30) through (33) and simplifying gives the expressions in (6), (7), (8), and (9) as follows:

$$\rho_{Y_p X_{pi}} = \frac{\beta_{pi} + \sum_{pj \neq pi} \beta_{pj}\rho_{X_{pi}X_{pj}}}{\sqrt{\sigma_p^2 + \sum_{pi} \beta_{pi}^2 + 2\sum_{pj \neq pi} \beta_{pi}\beta_{pj}\rho_{X_{pi}X_{pj}}}},$$

$$\rho_{Y_q X_{pi}} = \frac{\sum_{qi} \beta_{qi} \rho_{X_{qi} X_{pi}}}{\sqrt{\sigma_q^2 + \sum_{qi} \beta_{qi}^2 + 2 \sum_{qj \neq qi} \beta_{qi} \beta_{qj} \rho_{X_{qi} X_{qj}}}},$$

$$\rho_{Y_p Y_q} = \frac{\sigma_p \sigma_q \rho_{\varepsilon_p \varepsilon_q} + \sum_{pi} \sum_{qi} \beta_{pi} \beta_{qi} \rho_{X_{pi} X_{qi}}}{\sqrt{\sigma_p^2 + \sum_{pi} \beta_{pi}^2 + 2 \sum_{pj \neq pi} \beta_{pi} \beta_{pj} \rho_{X_{pi} X_{pj}}} \sqrt{\sigma_q^2 + \sum_{qi} \beta_{qi}^2 + 2 \sum_{qj \neq qi} \beta_{qi} \beta_{qj} \rho_{X_{qi} X_{qj}}}}, \text{ and}$$

$$\rho_{Y_p (\sigma_q \varepsilon_q)} = \frac{\sigma_p \rho_{\varepsilon_p \varepsilon_q}}{\sigma_q \sqrt{\sigma_p^2 + \sum_{pi} \beta_{pi}^2 + 2 \sum_{pj \neq pi} \beta_{pi} \beta_{pj} \rho_{X_{pi} X_{pj}}}}.$$

which completes the proof.

Appendix 2

For the desired correlations of $\rho_{Y_1X_{11}} = \rho_{Y_1X_{12}} = .40$, $\rho_{Y_2X_{21}} = \rho_{Y_2X_{22}} = .50$, and $\rho_{Y_3X_{31}} = \rho_{Y_3X_{32}} = .60$, the slope coefficients β_{11} and β_{12}; β_{21} and β_{22}; β_{31} and β_{32} are determined by simultaneously solving the following three independent sets of two equations for the unknowns (β_{*1}, β_{*2}):

$$.40 = \frac{\beta_{11} + \beta_{12}(.10)}{\sqrt{1 + \beta_{11}^2 + \beta_{12}^2 + 2\beta_{11}\beta_{12}(.10)}}, \text{ and} \tag{46}$$

$$.40 = \frac{\beta_{12} + \beta_{11}(.10)}{\sqrt{1 + \beta_{11}^2 + \beta_{12}^2 + 2\beta_{11}\beta_{12}(.10)}}; \tag{47}$$

$$.50 = \frac{\beta_{21} + \beta_{22}(.35)}{\sqrt{1 + \beta_{21}^2 + \beta_{22}^2 + 2\beta_{21}\beta_{22}(.35)}}, \text{ and} \tag{48}$$

$$.50 = \frac{\beta_{22} + \beta_{21}(.35)}{\sqrt{1 + \beta_{21}^2 + \beta_{22}^2 + 2\beta_{21}\beta_{22}(.35)}}; \tag{49}$$

$$.60 = \frac{\beta_{31} + \beta_{32}(.70)}{\sqrt{1 + \beta_{31}^2 + \beta_{32}^2 + 2\beta_{31}\beta_{32}(.70)}}, \text{ and} \tag{50}$$

$$.60 = \frac{\beta_{32} + \beta_{31}(.70)}{\sqrt{1 + \beta_{31}^2 + \beta_{32}^2 + 2\beta_{31}\beta_{32}(.70)}}; \tag{51}$$

which gives $\beta_{11} = \beta_{12} = 0.431834$; $\beta_{21} = \beta_{22} = 0.466760$; and $\beta_{31} = \beta_{32} = 0.464851$.

Given the solutions for β_{11}, β_{12}, β_{21}, β_{22}, β_{31}, and β_{32} from (46) through (51), the correlations $\rho_{Y_1Y_2}$, $\rho_{Y_1Y_3}$, $\rho_{Y_2Y_3}$, $\rho_{Y_2X_{11}}$, and $\rho_{Y_2X_{12}}$ are determined from equations (7) and (8) as follows:

$$\rho_{Y_1Y_2} = .40 + (0.431834)(0.466760)(.197432) + (0.431834)(0.466760)(.187942) + (0.431834)(0.466760)(.185900) + (0.431834)(0.466760)(.185860) / \left\{\left(1 + 2(0.431834)^2 + 2(0.431834)^2(.10)\right) \times \left(1 + 2(0.466760)^2 + 2(0.466760)^2(.35)\right)\right\}^{1/2}, \quad (52)$$

$$\rho_{Y_1Y_3} = .40 + (0.431834)(0.464851)(.287319) + (0.431834)(0.464851)(.268181) + (0.431834)(0.464851)(.258961) + (0.431834)(0.464851)(.258954) / \left\{\left(1 + 2(0.431834)^2 + 2(0.431834)^2(.10)\right) \times \left(1 + 2(0.464851)^2 + 2(0.464851)^2(.70)\right)\right\}^{1/2}, \quad (53)$$

$$\rho_{Y_2Y_3} = .40 + (0.464851)(0.466760)(.197432) + (0.464851)(0.466760)(.187942) + (0.464851)(0.466760)(.185900) + (0.464851)(0.466760)(.185860) / \left\{\left(1 + 2(0.466760)^2 + 2(0.466760)^2(.35)\right) \times \left(1 + 2(0.464851)^2 + 2(0.464851)^2(.70)\right)\right\}^{1/2}, \quad (54)$$

$$\rho_{Y_2X_{11}} = \frac{(0.466760)(.197432) + (0.466760)(.185960)}{\sqrt{\left(1 + 2(0.466760)^2 + 2(0.466760)^2(.35)\right)}}, \quad (55)$$

$$\rho_{Y_2X_{12}} = \frac{(0.466760)(.187942) + (0.466760)(.185860)}{\sqrt{\left(1 + 2(0.466760)^2 + 2(0.466760)^2(.35)\right)}}, \text{ and} \quad (56)$$

Simplifying (52) through (56) gives: $\rho_{Y_1Y_2} = .369251$, $\rho_{Y_1Y_3} = .393505$, $\rho_{Y_2Y_3} = .508267$, $\rho_{Y_2X_{11}} = .141997$, and $\rho_{Y_2X_{12}} = .138445$.

Note that the correlations in the numerators of (52) through (56) were determined by evaluating equation (5) from substituting the prespecified correlations from step 4 of the numerical example and constants from Table 1. For example, the correlation $\rho_{X_{11}X_{21}} = .197432$ in equation (52) was determined by evaluating (5) from substituting the values of $r_{11} = .350171$, $r_{21} = .657772$, and the constants from Table 1 representing the exponential and double exponential distributions.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM033737

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A METHOD FOR ~~SIMULATING~~ SIMULATING CORRELATED NONNORMAL SYSTEMS...

Author(s): TODD C. HEADRICK, T. MARK BEASLEY

Corporate Source: Southern Illinois University - Carbondale

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
☒	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Todd C. Headrick	Printed Name/Position/Title: Todd C. Headrick, Ph.D. / Ass. Professor
Organization/Address: SIUC. MAILCODE 4618, CARBONDALE, IL 62901-4618	Telephone: 618-453-1818 FAX: /
	E-Mail Address: headrick@siu.edu Date: 3-20-02

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)