Simulating Correlated Multivariate Pseudorandom Numbers

Ali A. Al-Subaihi
Institute of Public Administration
Riyadh 11141, Saudi Arabia
subaihia@ipa.edu.sa

Abstract

A modification of the Kaiser and Dichman (1962) procedure of generating multivariate random numbers with specified intercorrelation is proposed. The procedure works with positive and non-positive definite population correlation matrix. A SAS module is also provided to run the procedure.

Key words: Random Numbers, Multivariate Random Numbers, Random Number Generation.

1. INTRODUCTION

A fundamental task of quantitative researches is to make probability-based inferences about population characteristics, θ, based on an estimator, $\widehat{\theta}$, using a "representative" sample drawn from that population. However, the statistics of classical parametric inference inform us about how the population works to the extent necessary mathematical assumptions are met, and violation of any of the mathematical assumptions may harm the accuracy of the research conclusion(s). In regression, for instance, when the slope of the independent variable (*x*) is statistically significant and BLUE (best linear unbiased estimator), a researcher can clearly expect what happens to the dependent variable (*y*) when *x* changes one unit provided that the usual mathematical assumptions of the regression (independence, homogeneity of variance, and normality) are met. If any of these assumptions is violated, the risk of inference from the ordinary lease squares (OLS) is seriously off the mark (Moony, 1997).

In real world, however, there are numerous situations where data violates at least one of the mathematical assumptions of the inferential statistic that is going to be used in analysis. Therefore, statisticians as well as researchers are interested to know how the mathematical assumptoin(s) violation affects the inferntial statistics' power (the probability of correctly rejecting a false null hypothesis) and/or the Type I error rate (the probability of incorrectly rejecting a true null hypothesis). Fortunately, meaningful investigation of many problems in statistics can be done through Monte Carlo simulations (Brooks et al., 1999).

1.1 Statement of the Problem

Monte Carlo simulation is a computer-based technique that enables one to generate artificially random samples from populations with known characteristics in order to control some variables and manipulate others to investigate the effect of the manipulation on the robustness of a statistic. For more details about Monte Carlo simulation, the reader is referred to Brooks et al., 1999 and Mooney, 1997.

Monte Carlo simulations that require correlated data from normal and nonnormal populations are frequently used to investigate the small sample properties of competing statistics, or the comparison of estimation techniques (Headrick & Sawilowsky, 1999). And, the widely used technique to generate *correlated* data from *normal* population (which is the paper's concentration) is the Kaiser and Dichman (1962) procedure. The procedure uses a component analysis (e.g., principal components, square-root components) of a *positive definite* population correlation matrix (**R**) for generating multivariate random numbers with specified intercorrelations.

The procedure depends mainly on the decomposition of **R** and works *only* with a positive definite **R**. This limitation is an obstacle often faces users because not all real world situations have a positive definite **R**. Thus, a modification for this procedure (or development of a new one) is highly desired.

1.2 Purpose of the Study

The study proposes a modification of the Kaiser and Dichman procedure in order to decrease its limitation effect. That is, the paper presents techniques that enable the Kaiser and Dichman procedure' users to generate multivariate correlated pseudorandom numbers using a non-positive definite **R**.

2. REVIEW OF THE LITERATURE

The importance of the problem of generating a correlated pseudorandom numbers comes up from the considerable attention that has been given to the Monte Carlo studies; Monte Carlo studies have been of interest to applied statisticians and continue to receive large focus in the recent statistical literature.

The first to introduce a method for generating correlated pseudorandom numbers was Hoffman in 1924. He presented a simple technique that is used only to simulate two correlated variables. Then, Kaiser and Dickman (1962) extended the Hoffman's procedure for $m \geq 2$, where m is the number of the correlated variables. They proposed a method for generating sample and population score matrices and sample correlation matrices from a given **R**. The procedure utilizes component analysis for generating multivariate random numbers. The disadvantage of the Kaiser and Dickman procedure is that it depends on matrix factorization that requires positive definiteness, and this condition does not frequently hold.

Fleishman (1978) noted that real-world distributions of variables are typically characterized by their first four moments (i.e., mean, variance, skew, and kurtosis) and presented a procedure for generating normal as well as nonnormal random numbers with these moments specified. The procedure accomplishes this by simply taking a linear combination of a random number drawn from a normal distribution and its square and cube. Tadikamalla (1980) criticized the Fleishman's procedure for producing distributions of variables for which the exact distribution was known and thus lacked probability-density and cumulative-distribution functions and which, furthermore, could not produce distributions with all possible combinations of skew and kurtosis.

Tadikamalla (1980) proposed five alternative procedures for generating nonnormal random numbers and compared them all with Fleishman power method for speed of execution, simplicity of implementation, and generality of their ability to generate nonnormal distributions.

Vale and Maurelli (1983) extended the Fleishman (1978) and Kaiser and Dickman (1962) procedures to the multivariate case. They proposed a method for generating multivariate normal and nonnormal distributions with specified intercorrelations and marginal means, variances, skews, and kurtoses. However, like Kaiser and Dickman (1962) procedure , the procedure fails to work when **R** is not positive definite. Not only that, but also the procedure fails to generate desired intercorrelations when the conditional distributions possess extreme skew and/or heavy tails, even for sample sizes as large as N = 100 (Headrick and Sawilowsky, 1999).

Headrick and Sawilowsky (1999) proposed a method that combines a generalization of Theorem 1 and 2 from Knapp and Swoyer (1967) with the Fleishman (1978) procedure. The procedure generates multivariate nonnormal distributions with average values of intercorrelations closer to the population parameters for skewed and/or heavy tailed distributions and for small sample sizes. Although the procedure eliminates the necessity of conducting a factorization procedure on **R** that underlies the random deviate, yet it still *requires* the positive definiteness of **R**.

Briefly, all procedures that can be used to generate multivariate pseudorandom numbers require either directly or indirectly positive definite **R** and do not work without it, but the Fleishman’s (1978). This study will present a modified technique that minimizes the reliance on positive definiteness condition for **R**.

3. DESCRIPTION OF THE PROCEDURES

3-1 Kaiser and Dichman procedure

The Kaiser and Dichman (1962) procedure is generally applied to generate multivariate normal random numbers, and uses a matrix decomposition procedure. A Cholesky factorization (or any factorization, for that matter) is performed on **R** that is to underlie the random numbers. To generate a multivariate random number, one random number is generated for each component, and each random variable is defined as the sum of the products of the variable's component loadings and the random number corresponding to each of the components. When the random numbers input are normal, the resulting random numbers are multivariate normal with population intercorrelations equal to those of the matrix originally decomposed.

3-2 The Modified Procedure

The modified procedure is also used to generate multivariate normal pseudorandom numbers utilizing **R**. However, unlike other procedures, it works with a positive as well as a non-positive definite **R**.

To generate a correlated multivariate data matrix $\mathbf{D}^{n\times m} = [\mathbf{Y}^{n\times q}\ \mathbf{X}^{n\times k}]$ with specific **R**, we rewrite **R** as a block matrix such as

$$\mathbf{R} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{bmatrix}$$

where $\mathbf{A}^{q\times q}$ and $\mathbf{C}^{k\times k}$ are the correlation matrices of **Y** and **X**, respectively, and the $\mathbf{B}^{q\times k}$ is the correlation matrix between **Y** and **X**. The division is assumed to occur intentionally to ensure *positive definiteness* of both $\mathbf{A}^{q\times q}$ and $\mathbf{C}^{k\times k}$. This is a necessary condition (i.e., the method dose not work without it).

Next, one of the following techniques may be applied.

3-2-1 When A $^{q \times q}$ and q =1

This technique is applied to generate multivariate pseudorandom numbers when **R** can be partitioned into three matrices: $\mathbf{A}^{1\times 1}$, $\mathbf{B}^{1\times k}$, and $\mathbf{C}^{k\times k}$, where k is the number of **x**'s and k >1 such as

$$\mathsf{R} = \begin{bmatrix} 1 & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{bmatrix} = \left[\begin{array}{c:cccc} 1 & \rho_{\mathbf{yx}_1} & \rho_{\mathbf{yx}_2} & \cdots & \rho_{\mathbf{yx}_\mathbf{k}} \\ \hdashline \rho_{\mathbf{x}_1\mathbf{y}} & 1 & \rho_{\mathbf{x}} & \cdots & \rho_{\mathbf{x}} \\ \rho_{\mathbf{x}_2\mathbf{y}} & \rho_{\mathbf{x}} & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho_{\mathbf{x}} \\ \rho_{\mathbf{x}_\mathbf{k}\mathbf{y}} & \rho_{\mathbf{x}} & \cdots & \rho_{\mathbf{x}} & 1 \end{array}\right]$$

The technique is performed as follows:

A total of k correlation matrices that contain the correlation between **y** and each **x** individually are created such as $\mathbf{B}_i = \begin{bmatrix} 1 & \rho_{yx_i} \\ \rho_{x_i y} & 1 \end{bmatrix}$ ∀ i = 1, 2, …, k. Then, a multivariate normal data matrix ($\mathbf{X}^{n\times k}$) is generated through the equation

$$\mathbf{X} = \boldsymbol{\mu} + \hat{\mathbf{X}}\mathbf{U}$$

where **X** is the multivariate normal data matrix, $\boldsymbol{\mu}$ is the vector containing the variable means, $\hat{\mathbf{X}}$ contains vectors of independent and standard normal variates, **U** is the Cholesky upper triangular matrix of $\mathbf{C}^{k\times k}$. The factorization $\mathbf{C}^{k\times k} = \mathbf{U}\mathbf{U}^T$ is known as the Cholesky factorization.

A data vector ($\mathbf{y}^{n \times 1}$) of standard normal variates is generated and concatenated horizontally with each column of the matrix **X** individually such as

$\mathbf{T}_i$ =[**y** $\mathbf{X}_i$] ∀ i = 1, 2, …, k). This will gives us a total of k matrices each of size (n × 2).

A total of k vectors $\mathbf{Z}_i$ of size (n × 1) are generated independently through the equation $\mathbf{Z}_i = \boldsymbol{\mu} + \mathbf{T}_i\, \mathbf{U}_{ic_2}$ ∀ i = 1, 2, …, k

where $\mathbf{U}_{ic_2}$ is the 2nd column of the Cholesky upper triangular matrix of $\mathbf{B}_i$ which was created in the beginning. The vector **y** is then concatenated horizontally with $\mathbf{Z}_i \ \forall \ i = 1, 2, \ldots, k$ to create data matrix of size ($n \times k$+1) such as $\mathbf{D}^{\,n \times k+1} = [\ \mathbf{y}\ \mathbf{Z}_1\ \mathbf{Z}_2 \ldots \mathbf{Z}_k]$.

The resulting data matrix has a correlation matrix that most likely varies from the given population matrix (**R**) and the change takes place *mainly* in **C**. Thus, to get data matrix that has a correlation matrix close to (i.e. with average intercorrelation among the *x*'s around) the desired **R**, we manipulate the actual correlation value among the *x*'s, and repeat the process.

The average intercorrelation among the *x*'s is assessed using the measure proposed by Kaiser (1968)

$$\gamma = \frac{\lambda - 1}{O - 1} \qquad (3.1)$$

where λ is the largest eigenvalue of the correlation matrix, and *O* is the number of variables (which is *k* when measuring the average intercorrelation among the *x*'s). The larger the value of γ, the greater the correlation; if $\gamma = 0$, there is no correlation. $\gamma = 1$ when the correlations among the variables in the set are quite high. Accordingly, one should expect that γ takes values near the values of ρ_x.

Simulation Example 1

Suppose a data matrix of size (20 × 5) with specific intercorrelations need to be generated, and suppose

$$\mathbf{R} = \left[\begin{array}{c|cccc} \mathbf{y} & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \\ 1 & 0.5 & 0.5 & 0 & 0 \\ \hline 0.5 & 1 & 0.8 & 0.8 & 0.8 \\ 0.5 & 0.8 & 1 & 0.8 & 0.8 \\ 0 & 0.8 & 0.8 & 1 & 0.8 \\ 0 & 0.8 & 0.8 & 0.8 & 1 \end{array}\right]$$

The Kaiser and Dickman procedure cannot be used here to generate the required data because the provided **R** is *not* positive definite. Thus, the new procedure is going to be implemented instead (the SAS module in the Appendix A can be utilized to simulate the data using the modified procedure).

Table 1 shows the correlation matrix ($\mathbf{R}_i$) used in the procedure and the correlation matrix of the simulated data ($\mathbf{Re}_i$) at attempt $_i$. At the second attempt (when the correlation among the *x*'s was replaced by 0.84 in the original **R**), we had data with correlation matrix ($\mathbf{Re}_2$) that is closer to the original **R** (which equals $\mathbf{R}_1$) than those of the first ($\mathbf{Re}_1$) and the third ($\mathbf{Re}_3$) attempts.

Notice that $\mathbf{Re}_i$ in the Table are full rank (i.e., Rank($\mathbf{Re}_i$) = 5), and usually compared with the original **R**. $\mathbf{R}_2$, $\mathbf{R}_3$, … etc. are *only* used in the method to simulate the needed data.

TABLE 1 The Theoretical and Empirical Correlations Values

The Correlation matrix used							The Correlation matrix gotten						
		y	x_1	x_2	x_3	x_4			y	x_1	x_2	x_3	x_4
	y	1	0.5	0.5	0	0		y	1	0.5013	0.4948	-4E-04	-0.006
	x_1	0.5	1	0.8	0.8	0.8		x_1	0.5013	1	0.8497	0.6954	0.689
$\mathbf{R}_1$	x_2	0.5	0.8	1	0.8	0.8	$\mathbf{Re}_1$	x_2	0.4948	0.8497	1	0.6919	0.6896
	x_3	0	0.8	0.8	1	0.8		x_3	-4E-04	0.6954	0.6919	1	0.7982
	x_4	0	0.8	0.8	0.8	1		x_4	-0.006	0.689	0.6896	0.7982	1
											$\gamma = 0.736$		
		1	0.5	0.5	0	0			1	0.4964	0.5008	0.0011	0.0050
		0.5	1	0.84	0.84	0.84			0.4964	1	0.8819	0.7317	0.7363
$\mathbf{R}_2$		0.5	0.84	1	0.84	0.84	$\mathbf{Re}_2$		0.5008	0.8819	1	0.7272	0.7305
		0	0.84	0.84	1	0.84			0.0011	0.7317	0.7272	1	0.8432
		0	0.84	0.84	0.84	1			0.0050	0.7363	0.7305	0.8432	1
											$\gamma = 0.776$		
		1	0.45	0.45	0	0			1	0.4558	0.4552	0.0068	-0.004
		0.45	1	0.84	0.84	0.84			0.4558	1	0.8761	0.7515	0.7466
$\mathbf{R}_3$		0.45	0.84	1	0.84	0.84	$\mathbf{Re}_3$		0.4552	0.8761	1	0.7517	0.7478
		0	0.84	0.84	1	0.84			0.0068	0.7515	0.7517	1	0.8424
		0	0.84	0.84	0.84	1			-0.004	0.7466	0.7478	0.8424	1
											$\gamma = 0.786$		

Note: $\mathbf{R}_i$ is the correlation matrix used to simulate the required data; $\mathbf{Re}_i$ is the average correlation matrix of the data simulated 1000 times; γ is the Kaiser's Gamma.

The real and cpu times needed for 1000 replications = 0.10 and 0.07 seconds, respectively when a ***hp*** with 1.70GHz processor and 256 KB RAM is used. The standard error matrix of $\mathbf{Re}_2$ is as follows:

$$\mathbf{Re}_2 \quad \begin{array}{c|c|cccc} & y & x_1 & x_2 & x_3 & x_4 \\ y & 0 & 0.0054 & 0.0054 & 0.0069 & 0.0070 \\ \hline x_1 & 0.0054 & 0 & 0.0018 & 0.0036 & 0.0037 \\ x_2 & 0.0054 & 0.0018 & 0 & 0.0038 & 0.0038 \\ x_3 & 0.0069 & 0.0036 & 0.0038 & 0 & 0.0024 \\ x_4 & 0.0070 & 0.0037 & 0.0038 & 0.0024 & 0 \end{array}$$

3-2-2 When $\mathbf{A}^{q\times q}$ and $q \geq 2$

This technique is applied when **R** can be partitioned into three matrices: $\mathbf{A}^{q\times q}$, $\mathbf{B}^{q\times k}$, and $\mathbf{C}^{k\times k}$, where q >1 and k>1 are numbers of the **y**'s and the **x**'s, respectively, such as

$$\mathbf{R}=\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}' & \mathbf{C}\end{bmatrix}=\left[\begin{array}{cccc|cccc} \mathbf{1} & \boldsymbol{\rho}_{\mathrm{y}} & \cdots & \boldsymbol{\rho}_{\mathrm{y}} & \boldsymbol{\rho}_{\mathrm{y_1x_1}} & \boldsymbol{\rho}_{\mathrm{y_1x_2}} & \cdots & \boldsymbol{\rho}_{\mathrm{y_1x_k}} \\ \boldsymbol{\rho}_{\mathrm{y}} & \mathbf{1} & \ddots & \vdots & \boldsymbol{\rho}_{\mathrm{y_2x_1}} & \boldsymbol{\rho}_{\mathrm{y_2x_2}} & \cdots & \boldsymbol{\rho}_{\mathrm{y_2x_k}} \\ \vdots & \ddots & \ddots & \boldsymbol{\rho}_{\mathrm{y}} & \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{\rho}_{\mathrm{y}} & \cdots & \boldsymbol{\rho}_{\mathrm{y}} & \mathbf{1} & \boldsymbol{\rho}_{\mathrm{y_qx_1}} & \boldsymbol{\rho}_{\mathrm{y_qx_2}} & \cdots & \boldsymbol{\rho}_{\mathrm{y_qx_k}} \\ \hline \boldsymbol{\rho}_{\mathrm{x_1y_1}} & \boldsymbol{\rho}_{\mathrm{x_1y_2}} & \cdots & \boldsymbol{\rho}_{\mathrm{x_1y_q}} & \mathbf{1} & \boldsymbol{\rho}_{\mathrm{x}} & \cdots & \boldsymbol{\rho}_{\mathrm{x}} \\ \boldsymbol{\rho}_{\mathrm{x_2y_1}} & \boldsymbol{\rho}_{\mathrm{x_2y_2}} & \cdots & \boldsymbol{\rho}_{\mathrm{x_2y_q}} & \boldsymbol{\rho}_{\mathrm{x}} & \mathbf{1} & \ddots & \boldsymbol{\rho}_{\mathrm{x}} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{\rho}_{\mathrm{x_ky_1}} & \boldsymbol{\rho}_{\mathrm{x_ky_2}} & \cdots & \boldsymbol{\rho}_{\mathrm{x_ky_q}} & \boldsymbol{\rho}_{\mathrm{x}} & \cdots & \boldsymbol{\rho}_{\mathrm{x}} & \mathbf{1} \end{array}\right]$$

In this case, the technique is performed in this way:

A total of (q × k) correlation matrices that contain the correlation between each **y** and each **x** are created such as $\mathbf{B}_{ij}=\begin{bmatrix}1 & \rho_{y_ix_j}\\ \rho_{x_jy_i} & 1\end{bmatrix}$ $\forall$ i = 1, 2, …, q and j = 1, 2, …, k.

Then a multivariate normal data matrix ($\mathbf{X}^{n\times k}$) is generated through the equation

$$\mathbf{X}=\boldsymbol{\mu}+\hat{\mathbf{X}}\mathbf{U}$$

where **X** is the multivariate normal data matrix, $\boldsymbol{\mu}$ is the vector containing the variable means, $\hat{\mathbf{X}}$ contains vectors of independent and standard normal variates, **U** is the Cholesky upper triangular matrix of $\mathbf{C}^{k\times k}$.

Similarly, a multivariate normal data matrix ($\mathbf{Y}^{n\times q}$) is generated using the Cholesky upper triangular matrix of $\mathbf{A}^{q\times q}$. Next, each column of the matrix **Y** is concatenate horizontally with each column of the matrix **X** individually (i.e., $\mathbf{T}_{ij}$ =[$\mathbf{Y}_i$ $\mathbf{X}_j$] ∀ $i = 1, 2, \ldots, q$ and $j = 1, 2, \ldots, k$), which gives a total of ($q \times k$) matrices each of size ($n \times 2$).

A total of ($q \times k$) vectors $\mathbf{Z}_{ij}$ of size ($n \times 1$) are generated through the equation

$\mathbf{Z}_{ij} = \mu + \mathbf{T}_{ij}\, \mathbf{U}_{ijc_2}\ \forall\ i = 1, 2, \ldots, q$ and $j = 1, 2, \ldots, k$

where $\mathbf{U}_{ijc_2}$ is the 2nd column of the Cholesky upper triangular of the matrix $\mathbf{B}_{ij}$ which was created at first. Afterward, the vectors $\mathbf{Z}_j\ \forall\ j = 1, 2, \ldots, k$ are concatenate horizontally to create data matrix of size ($n \times k$) for each $\mathbf{Y}_i$ individually such as $\widetilde{\mathbf{X}}_i^{n\times k} = [\mathbf{Z}_{i1}\ \mathbf{Z}_{i2} \ldots \mathbf{Z}_{ik}]$.

The matrices $\widetilde{\mathbf{X}}_j^{n\times k}\ \forall\ i = 1, 2, \ldots, q$ are then summed together to create data matrix ($\sum_{i=1}^{q} \widetilde{\mathbf{X}}_i^{n\times k}$), and then the matrix **Y** is concatenate horizontally with the matrix

$\overline{\widetilde{\mathbf{X}}} = \dfrac{\sum_{i=1}^{q} \widetilde{\mathbf{X}}_i^{n\times k}}{q}$ to create the required data matrix ($\mathbf{D}^{n\times q+k}$).

Similar to the first case, the resulting data matrix ($\mathbf{D}^{n\times q+k}$) has a correlation matrix that most likely varies from the population matrix and the change, once again, takes place *mainly* in **C**. Thus, to get data matrix that has a correlation matrix close to (i.e. with average intercorrelation among the x's around) the desired population correlation matrix, we manipulate the actual correlation value among the x's, and repeat the process.

Simulation Example 2

Suppose a data matrix of size (20×6) with specific intercorrelation need to be generated and suppose

$$\mathbf{R} = \left[\begin{array}{ccc|ccc} 1 & 0.5 & 0.5 & 0.7 & 0.7 & 0.1 \\ 0.5 & 1 & 0.5 & 0.7 & 0.7 & 0.1 \\ 0.5 & 0.5 & 1 & 0.7 & 0.7 & 0.1 \\ \hline 0.7 & 0.7 & 0.7 & 1 & 0.4 & 0.4 \\ 0.7 & 0.7 & 0.7 & 0.4 & 1 & 0.4 \\ 0.1 & 0.1 & 0.1 & 0.4 & 0.4 & 1 \end{array}\right]$$

Again, the Kaiser and Dickman procedure does not work in this situation because the provided **R** is *not* positive definite. Consequently, the modified procedure is going to be implemented as an alternative (the SAS module in Appendix A also can be utilized to simulate the data using the modified procedure).

Table 2 shows the correlation matrix ($\mathbf{R}_i$) used in the procedure and the correlation matrix of the simulated data ($\mathbf{Re}_i$) at attempt $_i$. At the third attempt (when the correlation among the **x**'s was replaced by 0.84 in the original population correlation matrix), we had data with correlation matrix ($\mathbf{Re}_3$) that is closer to the original population correlation matrix ($\mathbf{R}_1$) than those of the first ($\mathbf{Re}_1$) and the second ($\mathbf{Re}_2$) attempts.

TABLE 2 The Theoretical and Empirical Correlations Values

The Correlation matrix used								The Correlation matrix gotten							
		y_1	**y_2**	**y_3**	**x_1**	**x_2**	**x_3**			**y_1**	**y_2**	**y_3**	**x_1**	**x_2**	**x_3**
R1	**y_1**	1	0.5	0.5	0.7	0.7	0.1	Re1	**y_1**	1	0.4959	0.4931	0.5165	0.5101	0.0700
	y_2	0.5	1	0.5	0.7	0.7	0.1		**y_2**	0.4959	1	0.4972	0.5079	0.5079	0.0730
	y_3	0.5	0.5	1	0.7	0.7	0.1		**y_3**	0.4931	0.4972	1	0.5119	0.5054	0.0635
	x_1	0.7	0.7	0.7	1	0.4	0.4		**x_1**	0.5165	0.5079	0.5119	1	0.6329	0.3546
	x_2	0.7	0.7	0.7	0.4	1	0.4		**x_2**	0.5101	0.5079	0.5054	0.6329	1	0.3672
	x_3	0.1	0.1	0.1	0.4	0.4	1		**x_3**	0.0700	0.0730	0.0635	0.3546	0.3672	1
													$\gamma = 0.4585$		
R2		1	0.5	0.5	0.9	0.9	0.1	Re2		1	0.5072	0.5061	0.7071	0.7015	0.0666
		0.5	1	0.5	0.9	0.9	0.1			0.5072	1	0.5098	0.7074	0.7063	0.0726
		0.5	0.5	1	0.9	0.9	0.1			0.5061	0.5098	1	0.7067	0.7068	0.0748
		0.9	0.9	0.9	1	0.4	0.4			0.7071	0.7074	0.7067	1	0.8448	0.2731
		0.9	0.9	0.9	0.4	1	0.4			0.7015	0.7063	0.7068	0.8448	1	0.2768
		0.1	0.1	0.1	0.4	0.4	1			0.0666	0.0726	0.0748	0.2731	0.2768	1
													$\gamma = 0.4982$		
R3		1	0.5	0.5	0.9	0.9	0.1	Re3		1	0.4984	0.4969	0.7028	0.7015	0.0649
		0.5	1	0.5	0.9	0.9	0.1			0.4984	1	0.4965	0.6989	0.7002	0.0604
		0.5	0.5	1	0.9	0.9	0.1			0.4969	0.4965	1	0.7002	0.6995	0.0631
		0.9	0.9	0.9	1	0.1	0.1			0.7028	0.6989	0.7002	1	0.7613	0.1193
		0.9	0.9	0.9	0.1	1	0.1			0.7015	0.7002	0.6995	0.7613	1	0.1163
		0.1	0.1	0.1	0.1	0.1	1			0.0649	0.0604	0.0631	0.1193	0.1163	1
													$\gamma = 0.3981$		

Note: **R**$_i$ is the correlation matrix used to simulate the required data; **Re**$_i$ is the average correlation matrix of the data simulated 1000 times; γ is the Kaiser's Gamma; the real and cpu times needed for 1000 replications = 0.57 and 0.10 seconds, respectively.

The real and cpu times needed for 1000 replications = 0.57 and 0.10 seconds, respectively when a ***hp*** with 1.70GHz processor and 256 KB RAM is used. The standard error matrix of **Re**$_2$ is as follows:

$$
\mathrm{Se} = \begin{array}{c|ccc|ccc}
 & \mathbf{y_1} & \mathbf{y_2} & \mathbf{y_3} & \mathbf{x_1} & \mathbf{x_2} & \mathbf{x_3} \\
\mathbf{y_1} & 0 & 0.0055 & 0.0056 & 0.0039 & 0.0039 & 0.0073 \\
\mathbf{y_2} & 0.0055 & 0 & 0.0057 & 0.0039 & 0.0041 & 0.0071 \\
\mathbf{y_3} & 0.0056 & 0.0057 & 0 & 0.0040 & 0.0041 & 0.0073 \\
\hline
\mathbf{x_1} & 0.0039 & 0.0039 & 0.0040 & 0 & 0.0033 & 0.0073 \\
\mathbf{x_2} & 0.0039 & 0.0041 & 0.0041 & 0.0033 & 0 & 0.0072 \\
\mathbf{x_3} & 0.0073 & 0.0071 & 0.0073 & 0.0073 & 0.0072 & 0
\end{array}
$$

4. CONCLUSION

The paper has suggested a modification to the Kaiser and Dickman (1962) procedure of generating multivariate normal pseudorandom numbers. As demonstrated in the numerical examples, the procedure can produce random numbers with intercorrelation near the population correlation matrix when other procedures do not work at all. Although there are few situations (especially when **A** or **C** are not positive definite) where the procedure does not work, still it might be used to generate numbers close to the preferred intercorrelations by manipulating the correlations among the *y*'s or the *x*'s beside the correlations between the *y*'s and the *x*'s.

APPENDIX (A)

```
      OPTIONS ls=100 ps=60 nodate nonumber;

proc iml;

/*********    The Parameters         **************/

NS=20;                           /* No. of subjects */

/* The population correlation matrix is entered as YY YX, XY XX */

      PopCor={1 .5 .5 .7 .7 .1,
              .5 1 .5 .7 .7 .1,
              .5 .5 1 .7 .7 .1,
              .7 .7 .7 1 .2 .2,
              .7 .7 .7 .2 1 .2,
              .1 .1 .1 .2 .2 1};
%Let NY=3;                          /*    No. of the y's   */
%Let NX=3;                          /*    No. of the x's   */
%Let NPC=9;                         /* No. of yx correlations = NY*NX   */

/****************************************************/

NV=&NY+&NX;                         /* No. of Variables    */
CorY= PopCor[1:&NY,1:&NY];          /* Corr. among the y's */
CorX= PopCor[&NY+1:NV,&NY+1:NV];    /* Corr. among the x's */
CorYX= PopCor[&NY+1:NV,1:&NY];      /* Corr. betw. the y's & the x's */

do i=1 to ncol(CorYX);  /* Corr. betw. the y's & the x's as a column*/
CorYXs=CorYXs//CorYX[,i];
end;

%macro loop(NPC);
      %Do i=1 %to &NPC;       /* Bi's Correlation matrices */
      Cryx&i=I(2);
      Cryx&i[1,2]=CorYXs[&i,1];
      Cryx&i[2,1]=CorYXs[&i,1];
      %end;

%loop (&npc);

X=Rannor(Repeat(0,NS,&NX))*root(CorX);    /* The X data matrix    */
y=Rannor(Repeat(0,NS,&NY))*root(CorY);    /* The Y data matrix    */

            DaXs=0*j(ns,&NX);

%macro loop2 (NY);
            %Let k=0;
                  %do j= 1 %to &NY;
            %do i=1 %to &NX;
            %Let c=%eval(&i+&K);
            %put c=&c;
            dat=(Y[,&j]||X[,&i])*(root(CrYX&c));
            dat2=dat2||dat[,2];
            %end;
            %Let k=&c;
            daXs=daXs+Dat2;
            free dat2;
                  %end;
```

```
%mend loop2;
%loop2 (&NY );

            daXs=daXs*(1/&NY);
            data=Y||daXs;     /* The final data matrix            */


      eg=eigval(corr(DaXs));
CXs=(eg[<>,1]-1)/(&NX-1);     /* The average Correlations among all x's */
      eg=eigval(corr(Y));
CYs=(eg[<>,1]-1)/(&NY-1);     /* The average Correlations among all y's */
Call=corr(data);              /* Correlations among all data */
ca=call[1:&NY,(&NY+1):(&NY+&NX)];

print       'The Correlations between Xs and Ys',ca,
            'The average Correlations among all Xs =  ' CXs,
            'The average Correlations among all Xs =  'CYs,
            'The total correlation matrix of the data', call;

quit;
```

REFERENCES

[1] Brooks, G. P.,& Robey, R. R. (1999), "Monte Carlo Simulation for Perusal and Practice", *Paper presented at the meeting of the American Educational Research Association*, Motereal, Quebec, Canada.

[2] Fleishman, A. (1978), "A Method for Simulating Non-Normal Distributions", *Psychomerika*, **43** (4), 521-532.

[3] Headrick, T. C., & Sawilowsky, S. S. (1999), “Simulating Correlated Multivariate Nonnormal Distributions Extending the Fleishman Power Method”, *Psychomerika*, **64** (1), 25-35.

[4] Hoffman, P. J. (1924), "Generating Variables with Arbitrary Parameters" *Psychomerika*, **24**, 265-267.

[5] Kaiser, H. F., & Dickman, K. (1962), "Sample and Population Score Matrices and Sample Correlation Matrices From an Arbitrary Population Correlation Matrix", *Psychomerika*, **27** (2), 179-182.

[6] Moony, C. Z. (1997), "*Monte Carlo Simulation*", Sage Publications, Thousand Oaks, CA.

[7] Tadikamalla, P. R. (1980), "On Simulating Non-Normal Distributions", *Psychomerika*, **45** (2), 273-279.

[8] Vale, C. D., & Maurelli, V. A. (1983), "Simulating Multivariate Nonnormal Distributions", *Psychomerika*, **48** (3), 465-471.

[9] 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** (1), 13-17.