SIMULATION OF SECURITY RETURNS FOR TESTING PORTFOLIO SELECTION PROCEDURES

George M. Frankfurter*
Syracuse University
Herbert E. Phillips**
State University of New York at Buffalo
John P. Seagle**
State University of New York at Buffalo

Abstract

Portfolio selection procedures typically require estimates of the means, variances, and covariances of returns from the securities. However, sampling error in these estimates is typically ignored. A model is presented to simulate returns that have a multivariate normal distribution, using a regression structure on a set of securities and holding the parameters constant.

Introduction

Present approaches for considering securities for inclusion in investment portfolios are incomplete because they fail to account for error in estimating the required parameters: average returns, and variances and covariances of security returns. Limited analytic work has dealt with the effect of error in estimating means[2] but it has not been possible to consider the effect of error in estimating variances and covariances, let alone the simultaneous effects of all three forms of estimation. In this paper, a simulation model is presented that permits analysis of the simultaneous effects of error in estimating the required parameters for a mean-variance portfolio selection model, and an experiment using the model is described. We begin with a brief review of the mean-variance approach to portfolio selection.

The mean-variance approach to portfolio selection, as described by Markowitz[3] and Sharpe[5] is based on the assumption that investors desire high average returns and low variance of returns. In this approach, a collection of securities is called "efficient" if it:

1) Maximizes expected return for a given level of variance of return.
2) Minimizes variance of return for a given expected return.

Given any set of s securities being considered for inclusion in an investor's portfolio, let the mean returns be represented by the vector

$$\mu = [\mu_1, \mu_2, \ldots, \mu_S] \tag{1}$$

(all vectors in this paper are column vectors), and let the variance-covariance matrix of these returns be represented by the s x s matrix

$$\Sigma = ||\sigma_{ij}|| = ||\rho_{ij}\sigma_i\sigma_j|| \tag{2}$$

A portfolio, which may be represented as a weighted average of the s securities, is defined by a vector

$$\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_S] \tag{3}$$

where λ_i denotes the fraction of the portfolio that is invested in security "i". The expected return "E", and variance of return "V" for any portfolio are obtained from

$$E = \lambda'\mu \tag{4}$$

and

$$V = \lambda'\Sigma\lambda \tag{5}$$

respectively.

The essence of the mean-variance approach is the recognition that security returns are not independent random variables, and thus that simply increasing the number of securities in a portfolio does not necessarily reduce V. Computational techniques for computing efficient E-V combinations exist, but existing procedures[2,5] assume that μ and Σ are known parameters. This is never the case in actual decision situations.

Why Simulation

In practice estimates

$$\tilde{\mu} = [\tilde{\mu}_1, \tilde{\mu}_2, \ldots, \tilde{\mu}_S] \tag{6}$$

and

$$\tilde{\Sigma} = ||\tilde{\sigma}_{ij}|| = ||\tilde{\rho}_{ij}\tilde{\sigma}_i\tilde{\sigma}_j|| \tag{7}$$

are used in portfolio selection models in place of the desired parameters. The analyst may have historical data regarding security returns that is believed to represent the process that will generate future returns that are of interest. Based on such information, however, the analyst can make but one estimation of each of the required parameters. In addition, firms change with time, as do other factors that affect security returns. Thus, the amount of relevant historical data that is available is necessarily small, precluding the possibility of conducting extensive empirical analyses on estimators that have entered into actual portfolio decisions.

*The names of the authors are in alphabetical order.

**Appreciation is expressed to the Research Foundation, State University of New York, for their generous financial support.

Even if one knew the distribution of $\tilde{\mu}$ and $\tilde{\Sigma}$ [see Equations (6) and (7)], it would be a formidable task to relate the distribution of these random variables (which estimate parameters in the model) to the decision vector $\tilde{\lambda}$. Just to identify an efficient portfolio for one given set of values for $\tilde{\mu}$ and $\tilde{\Sigma}$, one must solve for the values of $\tilde{\lambda}$ that optimize a quadratic programming problem. A solution for the entire distribution of $\tilde{\lambda}$ given the distributions of $\tilde{\mu}$ and $\tilde{\Sigma}$ does not seem tractable. A simulation approach, on the other hand, allows us to study directly the distribution of related decision $\tilde{\lambda}$, and the resultant distribution of returns, for a hypothetical investor who follows a mean-variance approach.

Example: An Application of Simulation in a Simple Case

Values for the parameters μ and Σ [see Equations (1) and (2)] were calculated from time series data for three actual securities. The resulting values are shown in Table 1.

Table 1
Securities Used in the Simulation

Security Number	1	2	3
Mean Returns Percent	16.64	6.64	21.35
Covariances:			
With Sec. 1	2102	-115	1115
With Sec. 2	-115	1664	-37
With Sec. 3	1115	-37	2223
	Chrysler	N.Y. Ship.	Bulova

Any number of portfolios can be made up of those securities by specifying a weight vector λ as defined in (3). Incrementing the λ_i in steps of 0.10, 66 different portfolios are defined for the simulation; these are shown in Table 2, ordered according to the mean and variance of returns, E and V respectively, as determined from the assumed security parameters of Table 1. Security returns were simulated using the multivariate normal data generator described in the next section, and the resulting annual returns for the 3 securities were collected in groups, each such group representing a number of accounting periods of data on security returns that may be available to the decision maker.

For each group of annual returns, a vector of means $\tilde{\mu}$, and a variance-covariance matrix $\tilde{\Sigma}$ were estimated, and a mean $\tilde{E}$ and variance $\tilde{V}$ were calculated for each of the 66 portfolios. This procedure was repeated for each of 100 sample trials with simulated data for 5, 10, 25, and 50 independent accounting periods for each trial. The percentage of simulated trials in which each of the 66 portfolios appeared to be efficient was recorded for use in preparing Figure 1, where these frequencies are reported. Each portfolio is assigned to a row and column of Figure 1 corresponding to the values of E and V calculated from the parameters μ and Σ, of Table 1. The circled portfolio, for example, is less desirable to an investor than any portfolio plotted below and to the right of it. Nonetheless, this portfolio appears efficient in from 25 to 42 percent of all simulated groups of sample returns. The sample size of 100 was chosen so that the percentages shown in Figure 1 would not be very sensitive to sampling error; the standard deviations of these percentages range from 3 percentage points at the 10 percent level of expected return to 10 percentage points at the 50 percent level.

Figure 1 shows that portfolios that are nominally inefficient can appear on the efficient frontier (in sample space) a high percentage of the time. A security analyst, using an E-V approach and basing his judgment on small sample sizes, can be seriously misled by error in estimation that is not accounted for by standard portfolio selection models.[2,9] In the next section, we present the methodology of the simulation in a fairly detailed way.

Methodology

An overview of the simulation model is presented in the form of a flow diagram, which is shown in Figure 2.

The annual percentage rates of return

$$R = [R_1, R_2, \ldots, R_s] \tag{8}$$

for any set of s securities, are assumed to have a multivariate normal distribution

$$f(R) = |H|^{\frac{1}{2}} \cdot (2\pi)^{-s/2} \exp[-\tfrac{1}{2}(R-\mu)' H (R-\mu)] \tag{9}$$

where

$$H^{-1} \equiv \Sigma \tag{10}$$

is the variance-covariance matrix of security returns. The parameters of (9), which are μ and Σ (see Equations (1) and (2)), are obtained by statistical methods applied to historical data for a specific collection of s securities. A data generating algorithm, based on Equation (9), is used to generate trial samples of R.

The data generating algorithm involves a series of regression structures. Let the s securities be arranged 1, 2, ..., s. Then by the logic of Equation (9), the marginal distribution of the first security

$$f(R_1) = (2\pi\sigma_n)^{-\frac{1}{2}} \cdot \exp[-\tfrac{1}{2}(R_1 - \mu_1)\sigma_{11}^{-1}] \tag{11}$$

is univariate normal, where σ_{11} denotes the upper left element of Σ. Returns on Securities 1 and 2 are bivariate normal[4]

$$f(R_1, R_2) = 2\pi^{-1}|\Sigma_{22}|^{-\frac{1}{2}} \cdot \exp\left\{-\tfrac{1}{2}(\underline{R}_2 - \underline{\mu}_2)'\Sigma_{22}^{-1}(\underline{R}_2 - \underline{\mu}_2)\right\} \tag{12}$$

where

$$\underline{R}_2 = [R_1, R_2], \qquad \underline{\mu}_2 \doteq [\mu_1, \mu_2] \tag{13}$$

and Σ_{22} is the variance-covariance matrix for Securities 1 and 2, which is a partition of Σ. Once R_1 has been generated by a standard method, R_2 can be obtained from the regression equation

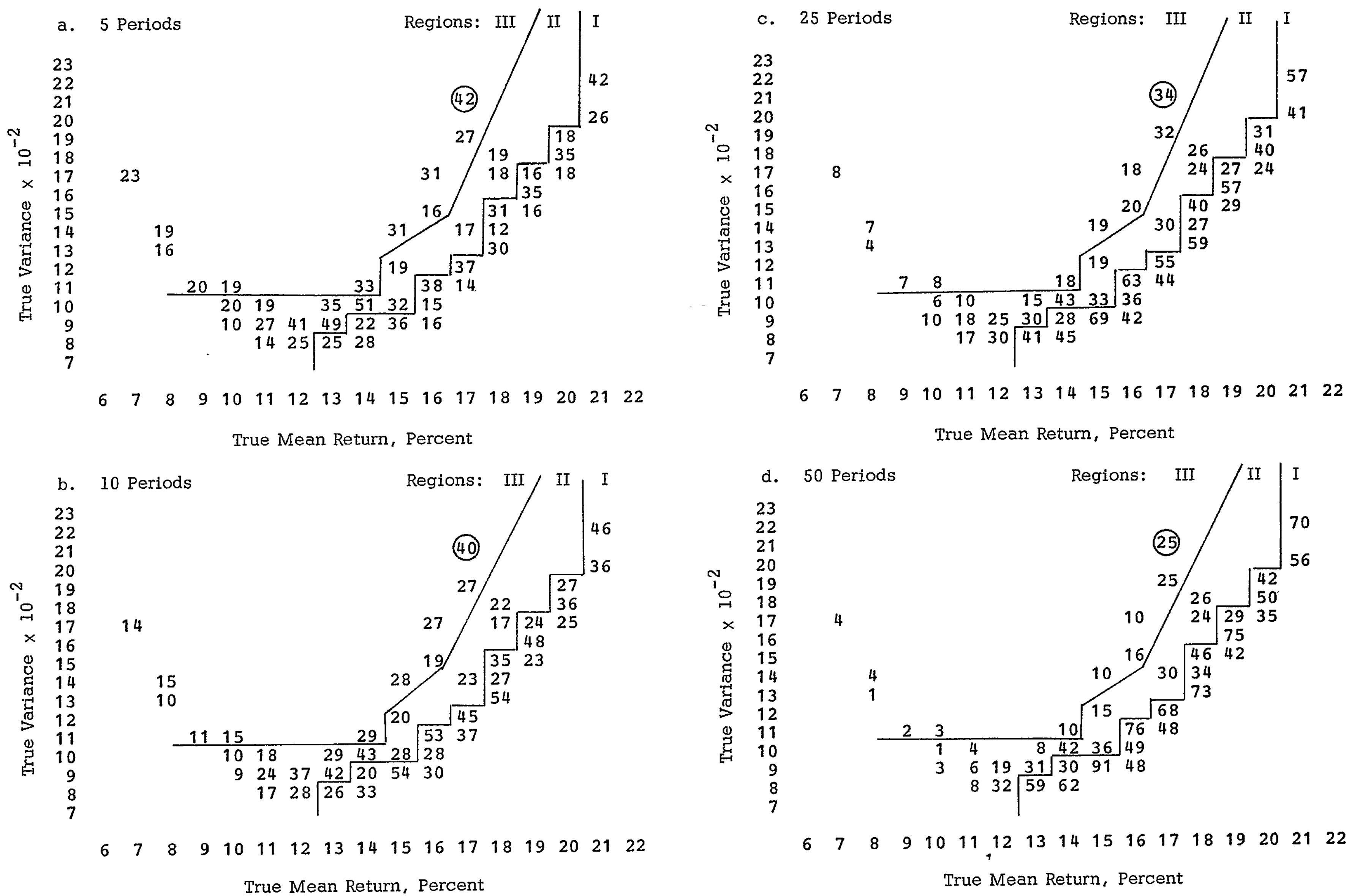
Portfolio Number	Portfolio Weights			Mean Return	Portfolio Variance
1*	0	0	10	21.35	22.22
2*	1	0	9	20.88	20.22
3*	2	0	8	20.41	18.63
4*	3	0	7	19.94	17.46
5	0	1	9	19.88	18.10
6*	4	0	6	19.46	16.72
7*	1	1	8	19.41	16.30
8	5	0	5	18.99	16.38
9*	2	1	7	18.94	14.92
10	6	0	4	18.52	16.47
11*	3	1	6	18.47	13.96
12	0	2	8	18.41	14.77
13	7	0	3	18.05	16.98
14*	4	1	5	17.99	13.42
15*	1	2	7	17.94	13.18
16	8	0	2	17.58	17.91
17	5	1	4	17.52	13.29
18*	2	2	6	17.47	12.00
19	9	0	1	17.11	19.25
20	6	1	3	17.05	13.59
21*	3	2	5	17.00	11.25
22	0	3	7	16.94	12.23
23	10	0	0	16.64	21.02
24	7	1	2	16.58	14.30
25*	4	2	4	16.52	10.91
26*	1	3	6	16.47	10.84
27	8	1	1	16.11	15.43
28	5	2	3	16.05	10.99
29*	2	3	5	16.00	09.88
30	9	1	0	15.64	16.98
31	6	2	2	15.58	11.49
32*	3	3	4	15.52	09.32
33	0	4	6	15.47	10.49
34	7	2	1	15.11	12.41
35*	4	3	3	15.05	09.19
36	1	4	5	15.00	09.30
37	8	2	0	14.64	13.75
38	5	3	2	14.58	09.48
39*	2	4	4	14.52	08.54
40	6	3	1	14.11	10.19
41*	3	4	3	14.05	08.20
42	0	5	5	14.00	09.53
43	7	3	0	13.64	11.31
44	4	4	2	13.58	08.27
45	1	5	4	13.53	08.56
46	5	4	1	13.11	08.76
47*	2	5	3	13.05	08.00
48	6	4	0	12.64	09.68
49*	3	5	2	12.58	07.86
50	0	6	4	12.53	09.37
51	4	5	1	12.11	08.14
52	1	6	3	12.06	08.60
53	5	5	0	11.64	08.84
54	2	6	2	11.58	08.25
55	3	6	1	11.11	08.31
56	0	7	3	11.06	10.00
57	4	6	0	10.64	08.80
58	1	7	2	10.58	09.44
59	2	7	1	10.11	09.29
60	3	7	0	09.64	09.56
61	0	8	2	09.58	11.42
62	1	8	1	09.11	11.06
63	2	8	0	08.64	11.12
64	0	9	1	08.12	13.64
65	1	9	0	07.64	13.48
66	0	10	0	06.64	16.64

Note: Mean returns are given in percentages, variances in $(\text{percent})^2 \times 10^{-2}$. Weights are fractions $\times 10$. Efficient portfolios are indicated by an *.

Table 2

Portfolio Summary Report

Figure I
Relative Frequency (percent) in Which Portfolios
Appear Efficient in Sample Trials



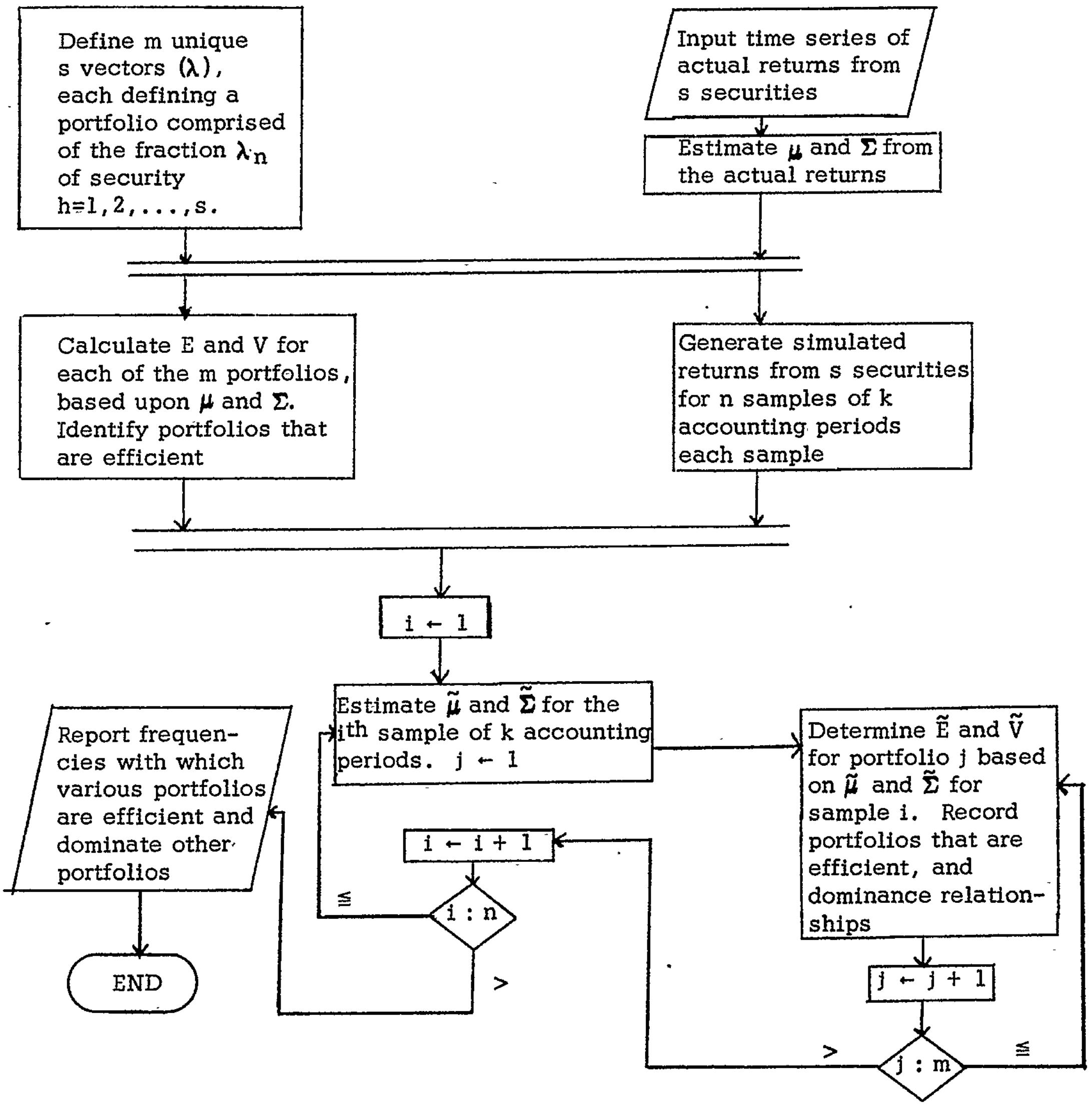


Figure 2

Flow Chart of Simulation

$$R_2 = \mu_2 + \rho_{12} \frac{\sigma_2}{\sigma_1} (R_1 - \mu_1) + \Sigma_{2.1} \quad (14)$$

where

$$\Sigma_{2.1} \sim N(0, \sigma_{22.1}) \quad (15)$$

and

$$\sigma_{22.1} = \sigma_{22} - (\sigma_{12})^2 / \sigma_{11} \quad (16)$$

Returns $R_3, R_4, \ldots, R_s$ can be sequentially generated in an analogous manner, using a series of multiple regression equations.

To establish the validity of the method described above, consider the situation where returns on the first q ($q \geq 1$) of the s securities have been properly generated, letting

$$\underline{R}_q = [R_1, R_2, \ldots, R_q] \quad (17)$$

be the s x 1 vector of known returns. The variance-covariance matrix is then partitioned

$$\Sigma = \left[\begin{array}{c|c} \Sigma_{qq} & \Sigma_{qs} \\ \hline \Sigma_{sq} & \Sigma_{ss} \end{array}\right] \quad (18)$$

such that Σ_{qq} is the variance-covariance matrix for the first q securities. Let

$$\underline{R}_s = [R_{q+1}, R_{q+2}, \ldots, R_s]$$

represent the returns not yet generated. It has been shown by Mood and Graybill[4] that $\underline{R}_s$ is distributed multivariate normal with mean

$$\underline{\mu}_{s.q} \equiv \underline{\mu}_s - \Sigma_{sq} \Sigma_{qq}^{-1} (\underline{R}_q - \underline{\mu}_q) \quad (19)$$

and variance

$$\Sigma_{ss.q} = \Sigma_{ss} - \Sigma_{sq} \Sigma_{qq}^{-1} \Sigma_{qs} \quad (20)$$

where $\underline{\mu}_q$ is the vector of means for $\underline{R}_q$ and $\underline{\mu}_s$ is the vector of means for $\underline{R}_s$.

To generate R_{q+1}, one needs only to determine a univariate normal density marginal to the distribution of $\underline{R}_s$. From (19), R_{q+1} has mean

$$\mu_{q+1.q} = \mu_{q+1} - \underline{b}'_{q+1.q} \cdot [\underline{x}_q - \underline{\mu}_q] \quad (21)$$

where

$$b'_{q+1.q} \equiv \text{first row of } \Sigma_{sq} \Sigma_{qq}^{-1} \quad (22)$$

and can be calculated as the coefficients of a multiple regression of R_{q+1} against $\underline{R}_q$, where the assumed μ and Σ are substituted for sample estimates. R_{q+1} has variance

$$\sigma_{q+1, q+1.q} = \sigma_{q+1, q+1} - b'_{q+1.q} \cdot [\sigma_{1, q+1}, \sigma_{2, q+1}, \ldots, \sigma_{q, q+1}] \quad (23)$$

and is obtained from the modified regression model described above. Notice that degree of freedom adjustments must be eliminated when standard regression codes are used, as (assumed) parameters are used in place of the usual sample values.

The remaining steps in the simulation are more straightforward, as is clear from Figure 2. These steps were applied in the example presented in the previous section.

Conclusion

In this paper, a model is presented for generating multivariate normal returns from a group of correlated securities. A regression structure is used, but the logic is precisely reversed relative to a usual regression analysis, and allows the random variables to be generated with little more computational effort than for uncorrelated variables. The parameters of the multivariate normal distribution are preserved.

References

1. Frankfurter, G.M., H.E. Phillips, and J.P. Seagle, "Portfolio Selection: The Effects of Uncertain Means, Variances and Covariances." Working Paper Series No. 115, State University of New York at Buffalo, School of Management, June 1971.

2. Kalymon, Basil A., "Estimation Risk in the Portfolio Selection Model," Journal of Financial and Quantitative Analysis, 4:559-582, January 1971.

3. Markowitz, Harry M., Portfolio Selection, New York: John Wiley & Sons, Inc., 1959.

4. Mood, Alexander and Franklyn Graybill, Introduction to the Theory of Statistics, New York: McGraw-Hill Book Company, Inc., 1963.

5. Sharpe, William F., "A Simplified Model for Portfolio Analysis," Management Science, 9:277-293, January 1963.