

Chapter 7

Parameter Estimation for Multivariate Probability Distributions

Univariate probability distributions are for individual random variables so multivariate (MV) probability distributions are for two or more random variables that are dependent on one another. Multivariate distributions are the rule in economic analysis models because most variables are correlated to each other. The purpose of this chapter is to describe and demonstrate how to estimate and apply parameters for multivariate distributions. This chapter builds on Chapter 6, which describes how to simulate univariate probability distributions.

The chapter is separated into three parts: multivariate normal (MVN) distributions, multivariate empirical (MVE) distributions, and simulating very large MVE distributions. The MVN and MVE sections deal with correlating random variables within years or intra-temporal correlation. For the problem of simulating inter-temporally correlated random variables see Chapter 8, after working through this chapter. Chapter 8 provides a comprehensive treatment of intra- and inter- temporal correlation and is recommended for advanced work in simulation.

Ignoring Correlation

If two random variables are correlated and their correlation is ignored in simulation the model will either over or under state the variance and mean for the system's KOVs. The direction of the bias introduced on the variance is inversely related to the correlation. Ignoring a positive correlation between \tilde{X} and \tilde{Y} will understate the variance for Z if $Z = \tilde{X} + \tilde{Y}$. Ignoring a negative correlation between \tilde{X} and \tilde{Y} will overstate the variance for Z in the same case.

The reason why the variance of Z is inversely biased relative to the correlation between X and Y is due to the variance formula for variable Z :

Let $Z = \tilde{X} + \tilde{Y}$ where \tilde{X} and \tilde{Y} are random variables,

the expected value of Z is

$$E(Z) = E(X) + E(Y)$$

the variance for Z is

$$V(Z) = \sigma_X^2 + \sigma_Y^2 + 2 \text{Cov}(X, Y) = \sigma_X^2 + \sigma_Y^2 + 2 \rho_{xy} * \sigma_x * \sigma_y$$

where ρ_{xy} is the correlation between X and Y .

When X and Y are negatively correlated the $\text{Cov}(X, Y)$, or ρ_{xy} is negative and reduces $V(Z)$ so ignoring the correlation overstates the true variance of Z . The opposite is true when X and Y are positively correlated because the $\text{Cov}(X, Y)$ or ρ_{xy} is positive. In both cases the mean, $E(Z)$, is unbiased by ignoring the correlation between X and Y .

On the other hand if the KOV is a function of products of random variables, as $Z = \tilde{X} * \tilde{Y}$, the mean will be biased if correlation between \tilde{X} and \tilde{Y} is ignored. In this case the expected value and variance of Z is:

$$E(Z) = E(X) + E(Y) + \text{Cov}(X, Y) = E(X) + E(Y) + 2 \rho_{xy} * \sigma_x * \sigma_y$$

$$V(Z) = \sigma_x^2 + \sigma_y^2 + 2 \text{Cov}(X, Y) = \sigma_x^2 + \sigma_y^2 + 2 \rho_{xy} * \sigma_x * \sigma_y$$

The mean and variance for Z are over or under estimated, inversely with respect to the sign on the correlation coefficient, if correlation is ignored in simulation.

As demonstrated in this chapter simulating a multivariate probability distribution is very easy and automatically corrects for the potential of biasing the mean and variance. The procedure described for simulating multivariate distributions insures that the random variables are “appropriately” correlated, meaning that the historical correlation is maintained in the simulation process.

Multivariate Normal (MVN) Distribution

Two or more normally distributed random variables that are correlated must be simulated as a MVN distribution to prevent biasing the model results. Check for correlation of the random variables by calculating the simple correlation coefficients among the variables. If the correlation coefficients are significantly different from zero the variables must be simulated MVN. A Student-t test is used to test each correlation coefficient in the correlation matrix to determine if it is statistically different from zero at, say, the 95 percent level. Simetar provides a Student-t test of correlation coefficients when calculating the correlation matrix (Figure 7.1). (See Chapter 16 and Correlation Demo.XLS for an example of this test.) The example in Figure 7.1 uses a critical t value of 2.20 for a 95% confidence test. The calculated t-statistics (in the lower matrix) which are larger than the critical value indicate their corresponding correlation coefficient is statistically different from zero. For ease of interpretation these calculated t values are bold.

	A	B	C	D	E	F	G
20	Correlation Matrix						
21		Corn P	Wheat P	Sorg P	Corn Y	Wheat Y	Sorg Y
22	Corn P	1	0.87	0.98	-0.30	-0.18	-0.53
23	Wheat P		1	0.82	-0.16	-0.37	-0.34
24	Sorg P			1	-0.29	-0.17	-0.56
25	Corn Y				1	0.53	0.59
26	Wheat Y					1	0.47
27	Sorg Y						1
28							
29	Correlation Coefficient t-values.						
30	Significanc	95%		t-critical	2.20		
31		Corn P	Wheat P	Sorg P	Corn Y	Wheat Y	Sorg Y
32	Corn P		5.83	16.79	1.04	0.60	2.06
33	Wheat P			4.68	0.54	1.31	1.21
34	Sorg P				1.00	0.58	2.27
35	Corn Y					2.08	2.40
36	Wheat Y						1.78
37	Sorg Y						
38	Bold values indicate statistical significance at the specified level.						

Figure 7.1. Statistical Test of Correlation Coefficients.

A MVN distribution has three parameters (components) to be quantified and is described here for a model with four random variables. The three components for a four variable MVN distribution are:

- Deterministic component for each of the four variables is the mean, or forecast, or \hat{X}_j for $j = 1, 2, 3, 4$.
- Stochastic component for each of the four variables is the standard deviation about the mean or forecast, or $\hat{\sigma}_{\hat{e}_j}$ for $j = 1, 2, 3, 4$
- Multivariate component for the four variables is represented by a 4x4 correlation matrix, or ρ (or the covariance matrix or Σ).

- **Parameters for a MVN Distribution**

The deterministic component of a MVN can be the mean or the predicted value from a trend regression, multiple regression, or time series model for each of the random variables, such as:

$$\hat{X}_{ij} = \hat{a} + \hat{b}_1 T_i + \hat{b}_2 X_{i-1} + \hat{b}_3 Z_i$$

or simply the mean

$$\hat{X}_{ij} = \bar{X}_{ij}$$

where \hat{X}_{ij} are the predicted values for all random variables $X_j, j = 1, 2, 3, \dots m$, and i denotes the periods (years, months, etc.) over which the variable is to be simulated.

The stochastic component for the MVN distribution is the measure of the dispersion about the deterministic component. The dispersion measure for a normal distribution is the standard deviation ($\sigma_{\hat{e}}$). The standard deviation is calculated using the residuals about the mean or forecast and is defined for each j variable in the distribution as:

$$\hat{e}_{ij} = X_{ij} - \hat{X}_{ij}$$

$$\hat{\sigma}_{\hat{e}_j} = \text{standard deviation for the } \hat{e}_{ij} \text{'s.}$$

where $\hat{\sigma}_{\hat{e}_j}$ is the standard deviation of the residuals for each of the random variables $X_j, j = 1, 2, 3, \dots m$.

The $\hat{\sigma}_{\hat{e}_j}$ is calculated over the T historical periods used to calculate the deterministic component, the \hat{X}_j .

The multivariate component for the MVN distribution is generally the correlation matrix of rank m for the m random variables. The correlation matrix must be calculated using the residuals (\hat{e}_{ij}), i.e., the stochastic component. For a 4 variable model the ρ matrix is:

$$\rho = \begin{bmatrix} 1.0 & \rho_{12} & \rho_{13} & \rho_{14} \\ & 1.0 & \rho_{23} & \rho_{24} \\ & & 1.0 & \rho_{34} \\ & & & 1.0 \end{bmatrix}$$

An alternative method for simulating a MVN distribution uses the covariance matrix for the multivariate component. For a 4 variable MVN model the covariance matrix, Σ , is:

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_{22}^2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_{33}^2 & \sigma_{34} \\ & & & \sigma_{44}^2 \end{bmatrix}$$

– Parameter Estimation for the MVN Distribution

The steps for estimating the parameters for a MVN distribution are:

1. Calculate the best model possible to predict each of the random variables, whether this is simply the mean, a trend regression, a multiple regression, or a time series model.

$$\hat{X}_{ij} = \text{econometric model}$$

2. Calculate the residuals, $\hat{\epsilon}_{ij}$, from the econometric forecast for each random variable.
3. Calculate the standard deviations, $\hat{\sigma}_j$, for each random variable using their residuals.
4. Calculate the correlation matrix (ρ) and calculate the covariance matrix (Σ) for the random variables using the residuals. (Note: Use the residuals to calculate the matrices because the residuals are the stochastic component of the variables to be correlated. Calculating the correlation and covariance matrices from the actual data is equivalent to calculating the multivariate measures about the mean which is not the same as the correlation for the residuals.)

– Parameter Estimation Using Simetar

The Simple Statistics and Multiple Regression options in Simetar will most often be used to estimate the deterministic components for MVN distributions. When the Multiple Regression function is used, Simetar forecasts the random variable, \hat{X}_{ij} , and estimates the standard deviation for the residuals, $\hat{\sigma}_j$. Additionally, Simetar calculates the standard error of prediction, $\hat{\sigma}_{jp}$, or SEP_j which should be used in place of a standard deviation of the residuals for simulating a variable distributed normal.

Use the residuals provided in the Multiple Regression function's output to calculate the

correlation matrix and the covariance matrix. Simetar provides a Correlation function for calculating the ρ matrix and the Σ matrix and testing the correlation coefficients for significance.

An example of estimating the parameters for a 3 variable MVN distribution is included in Multivariate Normal Distribution Demo.XLS. Each of the four steps for MVN parameter estimation are identified and the distribution is simulated different ways using Simetar. The example begins with the data for the three random variables in rows 8-34. In Step 1, OLS regression results show significant trends for all three random variables. The residuals from trend for each random variable are calculated using the Simple Regression option in Simetar in Step 1. Standard deviations for the residuals are calculated using an Excel function in line 108. The unsorted residuals are used in Step 2 to calculate the correlation matrix.

– **Simulating a MVN Distribution**

Three methods for simulating a MVN probability distribution are presented here. The technical description of what is involved in simulating the MVN distribution is provided in an Appendix at the end of this Chapter.

The first method for simulating a MVN distribution uses the correlation matrix to simulate CUSD's. An example of this method for a three variable MVN distribution is presented in Figure 7.2 (see Multivariate Normal Distribution Demo.XLS). For this method a vector of CUSD's is simulated using =CUSD (Correlation Matrix). The CUSD's are used individually to simulate MVN random variables using the Simetar function:

$$=NORM(\hat{X}_j, StdDev_j, CUSD_j)$$

	A	B	C	D	E	F	G	H	I	J	K	L	M
118	Use Correlation Matrix to Generate CUSDs and =NORM() Function to Simulate MVN Distribution												
119		Random				Correlation Matrix							
120		Yields	Means	Std Dev	CUSDs	Corn	Wheat	Sorghum	Formulas in Column B		Formula for Column E		
121	Corn	110.0931	107.559	11.44636	0.5876		1	0.23	0.69	=NORM(C121,D121,E121)	=CUSD(F121:H123)		
122	Wheat	34.0981	35.312	2.3554	0.303102			1	0.24	=NORM(C122,D122,E122)			
123	Sorghum	65.2169	60.006	6.072636	0.804565				1	=NORM(C123,D123,E123)			

Figure 7.2. Simulating a MVN Distribution Using the Correlation Matrix.

The second method for simulating a MVN distribution uses the covariance matrix to simulate stochastic correlated deviations (or CDEVs). A CDEV is the number of deviations from the mean that the random value lies. For this method a vector of CDEV's is simulated using =CSND (Covariance Matrix). The CDEV's are used individually in the formula:

$$\tilde{X}_i = \hat{X}_i + CDEV_i$$

This method is demonstrated for simulating a three variable MVN distribution in Figure 7.3. (See the Multivariate Normal Distribution Demo.XLS for this example.)

	A	B	C	D	E	F	G	H	I	J	K	L	
125	Use the Covariance Matrix to Generate Correlated Deviations (CDEVs)												
126		Random	Histroical										
127	Year 1	Yields	Means	CDEVs	Covariance Matrix			Formulas in Column B		Formula in Column D			
128	Corn	124.5414	107.559	16.98221	302.6983	42.96493	114.6301	=C128+D128		=CSND(E128:G130)			
129	Wheat	38.1642	35.312	2.851893	0	13.38022	17.57007	=C129+D129					
130	Sorghum	65.31816	60.006	5.311842	0	0	62.536	=C130+D130					

Figure 7.3. Example of Using Covariance Matrix to Simulate MVN Distribution in Two Steps.

The third method for simulating a MVN distribution does it all in one step using Simetar's MVNORM function. The one step formula is the easiest method for simulating a MVN distribution over a multiple year planning horizon. The MVNORM function simulates the MVN values for the i^{th} forecast period as:

$$\tilde{X}_{ij} = \text{MVNORM}(\text{Vector of } \hat{X}_{ij}, \text{Covariance Matrix})$$

The one step method for simulating a MVN distribution is demonstrated in Figure 7.4.

	A	B	C	D	E	F	G	H	I	
110	A One Step Approach to Simulating a MVN Distribution.									
111	Assemble the parameters and use the =MVNORM() function									
112		Random		Covariance Matrix						
113		Yields	Means	Corn	Wheat	Sorghum	Formula in Column B			
114	Corn	84.24	107.559	302.6983	42.96493	114.6301	=MVNORM(C114:C116,D114:F116)			
115	Wheat	31.52	35.312		13.38022	17.57007				
116	Sorghum	51.46	60.006			62.536				

Figure 7.4. Example of the One Step Procedure for Simulating a MVN Distribution for One Year.

The one step method for simulating a MVN distribution is demonstrated in Figure 7.5 for a three variable MVN distribution that is simulated for four years.

	A	B	C	D	E	F	G	H
141	Simulate 5 Years for a MVN distribution in One Step							
142	Forecast Means for Each Year to Simulate							
143	Years	Corn	Wheat	Sorghum				
144	27	131.1439	40.34983	69.12418				
145	28	132.8909	40.72298	69.79958				
146	29	134.6379	41.09613	70.47498				
147	30	136.3849	41.46928	71.15038				
148	Covariance Matrix							
149		Corn	Wheat	Sorghum				
150	Corn	302.6983	42.96493	114.6301				
151	Wheat		13.38022	17.57007				
152	Sorghum			62.536				
153	Simulate Stochastic Forecast of the MVN for 4 Years							
154	Years	Corn	Wheat	Sorghum				
155	Y27	106.5542	39.55058	66.31111	=MVNORM(B144:D144,\$B\$150:\$D\$152)			
156	Y28	141.2978	45.432	76.61209	=MVNORM(B145:D145,\$B\$150:\$D\$152)			
157	Y29	103.6693	35.8383	48.31042	=MVNORM(B146:D146,\$B\$150:\$D\$152)			
158	Y30	139.0412	42.41304	65.17593	=MVNORM(B147:D147,\$B\$150:\$D\$152)			

Figure 7.5. Example of the One Step Procedure for Simulating a MVN Distribution for Four Years.

– **Results of Simulating MVN Distribution**

After simulating a multivariate probability distribution one must first test the simulated values to be sure they are appropriately correlated. This involves calculating the correlation matrix for the simulated random variables and comparing this correlation matrix to the original. Simetar includes a statistical test to assist in validating the correlation of multivariate distributions. The validation procedure for a MVN distribution is demonstrated using the example in Multivariate Normal Distribution Demo.XLS.

The correlation test used in validation compares the correlation matrix used for the simulation to the implicit correlation matrix in the simulated variables. For a model with three random variables the Check Correlation in Simetar’s Hypothesis Testing for Data menu, the user provides the location for the three simulated random variables, say B:8; B108 of the SimData worksheet and the location for the correlation matrix to simulate the three random variables. The null hypothesis for the test is that each correlation coefficient for the simulated variables equals the original or assumed correlation coefficient or:

$$H_0: \hat{\rho}_{ij} = \rho_{ij}$$

The test uses a Student –t test and the critical value at the alpha equal 5% or less based on sample size. To reject the null hypothesis the calculated t-test statistic must exceed the critical value of, say, 2.43 at the 98.3% confidence level. The calculated t-test statistics in Figure 7.6 are all less than 2.43 so we fail to reject the null hypothesis and say the simulated correlation coefficients are statistically equal to the assumed (historical) correlation coefficient at the 98% level. This test must be done for each year simulated by the MVN distribution.

	O	P	Q	R	S	T
2	Test Correlation Coefficients for Simulated Values in Year 27					
3	Confidence Level		98.3048%			
4	Critical Value		2.43			
5						
6		Wheat	Sorghum			
7	Corn	0.84	0.25			
8	Wheat		0.47			

Figure 7.6. Example of a Student –t Test for a Correlation Matrix.

Simetar provides three non-parametric tests for validating MVN distributions. The tests are:

- Two Sample Hotelling T² Test – tests the historical mean vector vs. the mean vector for the simulated variables.
- Box’s M Test – tests the historical covariance matrix vs. the covariance matrix for the simulated variables.
- Complete Homogeneity Test – simultaneously tests the historical means vector and covariance matrix vs. the means vector and covariance matrix for the simulated variables.

The tests are described in Chapter 3 and 16 in more detail. If the tests “Fail to Reject the Null Hypothesis” that the simulated parameters equal the historical parameters, then the simulation process simulated the MVN distribution appropriately. An example of using the non-parametric tests to validate a MVN distribution is provided in Figure 7.7. When the Test Values are less than the Critical Values, the tests will Fail to Reject the null hypothesis and this is what we want.

	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	A
28	Distribution Comparison of Corn & Corn													
29	Confidence Level		95.0000%											
30		Test Value	Critical Val	P-Value										
31	2 Sample Hotelling T ² Test	0.00	8.17	1.000	<i>Fail to Reject the Ho that the Mean Vectors are Equal</i>									
32	Box's M Test	0.61	12.59	0.996	<i>Fail to Reject the Ho that the Covariance Matrices are Equivalent*</i>									
33	Complete Homogeneity Test	0.58	16.92	1.000	<i>Fail to Reject the Ho that the Mean Vectors and Covariance Matrices are Equivalent, Respectively*</i>									
34	<i>*Based on asymptotic distribution</i>													

Figure 7.7. Non-Parametric Tests for Testing Simulated Values from MVN Distributions.

Multivariate Empirical (MVE) Distribution

Two or more correlated random variables can be simulated as a multivariate distribution, even if the variables are not normally distributed. For example, non-normal MV distributions can be simulated as a MVE distribution. The generalized MV procedure presented here for the MVE distribution allows one to correlate non-normal distributions in a simulation model. Richardson and Condra first introduced the procedure in 1978 and an extension of their original procedure is used here and expanded upon in Chapter 8. King later reported the procedure to simulate a multivariate beta distribution. The workbook Multivariate Empirical Demo.XLS demonstrates the steps for MVE parameter estimation and simulation described in this section.

A MVE distribution has three parameters or components to be estimated. The MVE is described here for a model with four random variables, even though the procedure can easily be expanded for an m variable MVE. The three components/parameters are:

- Deterministic component for each of the four variables, \hat{X}_j for $j = 1, 2, 3, 4$.
- Stochastic component for each of the four variables, $S_{\hat{e}_j}$ for $j = 1, 2, 3, 4$
- Multivariate component for the four variables $\rho_{4 \times 4}$.

– Parameters for a MVE Distribution

The deterministic component is the projected value based on the mean, trend regression, multiple regression, or time series model for each of the random variables, such as:

$$\hat{X}_{ij} = \hat{a} + \hat{b}_1 T_i + \hat{b}_2 X_{i-1} + \hat{b}_3 Z_i$$

or simply the mean:

$$\hat{X}_{ij} = \bar{X}_{ij}$$

where \hat{X}_{ij} are the predicted values for all random variables X_j , $j = 1, 2, 3, \dots, m$, and i denotes the periods (years, months, etc.) over which the variable is to be simulated.

The stochastic component for the MVE distribution is the measure of the dispersion about the deterministic component or $S_{\hat{\epsilon}}$. The dispersion measure for an empirical distribution is the vector of sorted deviations from the deterministic component, expressed as a fraction of the forecasted values at each historical period i . The $S_{\hat{\epsilon}}$ values are calculated for each random variable as:

$$\begin{aligned} \hat{\epsilon}_{ij} &= X_{ij} - \hat{X}_{ij} \\ F_{\hat{\epsilon}ij} &= \hat{\epsilon}_{ij} / \hat{X}_{ij} \\ S_{\hat{\epsilon}ij} &= \text{Sorted}(F_{\hat{\epsilon}ij}) \end{aligned}$$

where $S_{\hat{\epsilon}ij}$ are the sorted fractional residuals for each of the random variables X_j , $j = 1, 2, 3, \dots, m$ over the historical period $i = 1, 2, 3, \dots, T$

Multivariate component for the MVE distribution is the correlation matrix of rank m for the m random variables. The correlation matrix must be calculated using the unsorted residuals ($\hat{\epsilon}_{ij}$), i.e., the stochastic component. For a 4 variable model the ρ matrix is:

$$\rho = \begin{pmatrix} 1.0 & \rho_{12} & \rho_{13} & \rho_{14} \\ & 1.0 & \rho_{23} & \rho_{24} \\ & & 1.0 & \rho_{34} \\ & & & 1.0 \end{pmatrix}$$

– **Parameter Estimation for the MVE Distribution**

The steps for estimating the parameters for a MVE distribution are:

1. Calculate the best model possible to predict each of the random variables, whether this is simply the mean or a complex econometric model, such as, a trend regression, a multiple regression, or a time series model.
 \hat{X}_{ij} = econometric model
2. Calculate the residuals, $\hat{\epsilon}_{ij}$, from the econometric forecast for each random variable.
3. Calculate the $m \times m$ correlation matrix for all of the random variables using the unsorted residuals. (Note: Use the residuals to calculate the correlation matrix because the residuals are the stochastic component of the variables to be correlated. Calculating the correlation matrix from the actual data is equivalent to calculating the multivariate measures about the mean which is not the same as the correlation for the residuals.

4. Calculate the fractional residuals for each variable and then sort these values for each of the random variables. Calculate the pseudo minimums and maximums for each variable using the sorted fractional residuals.
5. Assign probabilities to each of the sorted fractional residuals including a zero to the pseudo minimum and a one to the pseudo maximum.

– Parameter Estimation Using Simetar

The EMP icon on the Simetar toolbar can be used to calculate all of the parameters for an MVE distribution. The EMP icon calculates the S_{eij} values for multiple variables assuming the MVE uses either actual data, deviations from mean or deviations from trend. If more complex forecasting models are needed to project \hat{X}_{ij} , use the residuals from the econometric model and select the Actual Data option in the EMP icon to estimate the parameters.

The steps for estimating the parameters for a MVE distribution are presented in Multivariate Empirical Demo.XLS. The worksheet starts with six random variables and goes through the steps described above for estimating the parameters. The data do not have statistically significant trends so fractional deviations from the mean are used to estimate the S_i parameters for the empirical distributions. The resulting MVE distribution is simulated three different ways in Steps 7-9.

– Simulating a MVE Distribution

A MVE distribution can be simulated several ways in Excel using Simetar functions. Three methods are presented here starting with the easiest and proceeding to the most complex. But first the general procedure is described in matrix notation for completeness.

The first step is to simulate an $M \times 1$ vector of correlated uniform standard deviates or CUSD's. The Simetar array function =CUSD() performs the necessary calculations and simulation. In its most simple form the CUSD function is =CUSD (Correlation Matrix) so the result is m cells with CUSDs as demonstrated in Figure 7.8.

	E	F	G	H	I	J	K	L	M
160	CUSDs			Correlation Matrix					Formula in Column E
161	0.3354	1.000	0.869	0.981	-0.299	-0.175	-0.529	=CUSD(F161:K166)	
162	0.5353	0.000	1.000	0.816	-0.159	-0.367	-0.342		
163	0.2788	0.000	0.000	1.000	-0.288	-0.169	-0.565		
164	0.8504	0.000	0.000	0.000	1.000	0.531	0.587		
165	0.3809	0.000	0.000	0.000	0.000	1.000	0.473		
166	0.8180	0.000	0.000	0.000	0.000	0.000	1.000		

Figure 7.8. Example of Simulating a Vector of CUSD's.

The final step in simulating a MVE distribution is to use the CUSD's in an empirical distribution. The formula varies depending upon the format of the S_j values:

- S_j are actual data

$$\tilde{X}_j = \text{EMP}(S_j, P(S_j), \text{CUSD}_j)$$

- S_j are absolute deviations from means

$$\tilde{X}_j = \bar{X}_j + \text{EMP}(S_j, P(S_j), \text{CUSD}_j)$$

- S_j are fractional deviations from means

$$\tilde{X}_j = \bar{X}_j + \bar{X}_j * \text{EMP}(S_j, P(S_j), \text{CUSD}_j)$$

- S_j are fractional deviations from a forecast

$$\tilde{X}_j = \hat{X}_j + \hat{X}_j * \text{EMP}(S_j, P(S_j), \text{CUSD}_j)$$

The EMP() Simetar function simulates an empirical distribution defined by S_j and $P(S_j)$ using the uniform standard deviate indicated by the CUSD_j .

There are two ways to simulate the MVE distribution with Simetar. The first method uses the =MVEMP() function and it takes only one step. The one step method calculates all of the parameters in the background and also generates the CUSD's. Program the one step MVE distribution function as:

$$=\text{MVEMP}(\text{Actual Data},,,, \text{Forecasted Xs}, \text{Code})$$

where: Actual Data is the location for the original historical data,
 The ,,,, must be provided and indicate optional parameters that are not provided,
 Forecasted Xs represent the \hat{X}_{ij} for the i^{th} period to simulate,
 Code is a switch to specify the format for the Sjs, as 0 for actual data,
 1 for fractional deviates from mean,
 2 for fractional deviates from trend, and
 3 for differences from mean.

The =MVEMP() function is an array function so highlight m cells representing the m random variables in the MVE distribution. Additionally, the function is completed by pressing Control Shift Enter. An example of simulating a six variable MVE distribution using the one step method is presented in Figure 7.9.

	A	B	C	D	E	F	G
148		Random					
149	Names	Values	Formulas Used to Simulate the MVE in One Step				
150	Corn Pr	2.98	=MVEMP(B11:G23,,,B144:G144,1)				
151	Wheat Pr	3.63					
152	Sorghum Pr	2.64	Note B11:G23 is the actual data and B144:G144 is the				
153	Corn Yld	127.04	forecasted deterministic component of the MVE distribution				
154	Wheat Yld	43.05					
155	Sorghum Yld	67.53					

Figure 7.9. One Step Method to Simulate an MVE Distribution.

The second method for simulating an MVE distribution with Simetar is a two step process. First generate an $m \times 1$ vector of CUSD's. Next use the CUSD's in the EMP function applying the appropriate formula based on the form of the S_j to each of the m random variables:

$$=\hat{X}_j + \hat{X}_j * \text{EMP}(S_j, P(S_j), \text{CUSD}_j)$$

Repeat this formula for each of the m variables in the distribution.

This method is demonstrated in Figure 7.10 for a six variable MVE distribution where the S_j 's are specified as fractional deviations from the mean.

	A	B	C	D	E	F	G	H	I	J	K
159		Random		Correlated							
160	Names	Values	Means	Deviates as %	CUSDs			Correlation Matrix			
161	Corn Pr	2.23	2.325	-0.0395	0.3354	1.000	0.869	0.981	-0.299	-0.175	-0.529
162	Wheat Pr	3.32	3.294	0.0063	0.5353	0.000	1.000	0.816	-0.159	-0.367	-0.342
163	Sorghum Pr	1.90	2.128	-0.1066	0.2788	0.000	0.000	1.000	-0.288	-0.169	-0.565
164	Corn Yld	132.49	118.363	0.1194	0.8504	0.000	0.000	0.000	1.000	0.531	0.587
165	Wheat Yld	36.04	37.152	-0.0300	0.3809	0.000	0.000	0.000	0.000	1.000	0.473
166	Sorghum Yld	69.96	64.855	0.0787	0.8180	0.000	0.000	0.000	0.000	0.000	1.000
167		=C166+C166*D166			=CUSD(F161:K166)						
168				=EMP(G\$122:G\$136,\$A\$122:\$A\$136,E166)							

Figure 7.10. Example of Using a Two Step Method to Simulate a MVE Distribution.

If there is a one step method, what is the need for the two step method? The two step method is used when not all of the random variables have the same form for the S_j 's. For example, if the first three X s have no trend so they are simulated as actual data and the next three X 's have a trend and must be simulated as fractional deviations from trend.

– Results of Simulating MVE Distribution

A validation test to insure that the random variables are appropriately correlated should be done before using the random values in a decision model. Simulate the MVE distribution collecting values for the stochastic variables and then test the correlation implicit in the simulated values against the correlation matrix for the historical data. Validation of the simulated random variables is particularly important for non-normal distributions because the procedure is not widely used and understood. Simetar provides a correlation test and three non-parametric tests for validating MVE distributions. All four tests should be used to validate all multivariate distributions.

The six variable MVE distribution simulated in the Multivariate Empirical Demo.XLS is used to demonstrate the four validation tests in the SimData worksheet. The results of the four tests are summarized in Figure 7.11. The results of the correlation test reveal that all of the correlation coefficients implicit in the simulated variables are statistically equal to their counterparts in the actual data's correlation matrix, at the 99% level. Additionally the mean vectors and covariance matrices for the simulated data and the actual data are statistically equal at the 95 percent level.

	I	J	K	L	M	N	O	P	Q	R	S	T	U	
7	Test Correlation Coefficients													
8	Confidence Level	99.6586%												
9	Critical Value	3.00												
10														
11		Wheat Pr	Sorghum F	Corn Yld	Wheat Yld	Sorghum Yld								
12	Corn Pr	0.38	1.19	0.58	0.83	1.27								
13	Wheat Pr		1.13	0.68	0.68	0.55								
14	Sorghum Pr			0.72	0.70	0.93								
15	Corn Yld				1.55	0.37								
16	Wheat Yld					1.29								
17														
18	Distribution Comparison of Corn Pr & Corn Pr													
19	Confidence Level	95.0000%												
20	Test Value	Critical Val	P-Value											
21	2 Sample t	1.77	13.73	0.945	<i>Fail to Reject the Ho that the Mean Vectors are Equal</i>									
22	Box's M Test	12.83	32.67	0.914	<i>Fail to Reject the Ho that the Covariance Matrices are Equivalent*</i>									
23	Complete L	14.69	40.11	0.974	<i>Fail to Reject the Ho that the Mean Vectors and Covariance Matrices are Equivalent, Respectively*</i>									
24					<i>*Based on asymptotic distribution</i>									

Figure 7.11. Validation Tests for MVE Distributions.

Mixed Multivariate Probability Distributions

When parameter estimation (or the problem being analyzed) requires simulation of a distribution where the random variables are not all normal or all empirical, what do you do? Ignoring the correlation of these variables would bias the key output variables in the model by over-or under-stating their means and risk. The procedure presented in this section allows for the appropriate correlation of random variables with different distributions. For example, one variable can be normal, another can be empirical, and another can be uniform or beta and yet the probability distribution can be appropriately correlated in simulation, so historical variability and correlation will be observed in the simulation. Richardson and Condra reported this procedure in 1987.

– Parameter Estimation

The deterministic component for each random variable must be quantified using the best model possible, such as, mean, trend regression, multiple regression, or time series model. The stochastic component for each variable is the residual from the deterministic component and must be calculated for each variable in the MV distribution, i.e., estimate the $\hat{\epsilon}_{ij}$'s. The multivariate component of the MV distribution (ρ matrix) must be estimated using the unsorted residuals or $\hat{\epsilon}_{ij}$'s. The parameters to quantify the stochastic component for each of the random variables are estimated based on the appropriate parameters for each variable's assumed distribution, i.e., $\hat{\sigma}$ for the normal, S_i and $P(S_i)$ for the empirical, minimum and maximum for the uniform, and so on.

– Simulation Steps

1. Generate m correlated uniform standard deviates using the correlation matrix and the Simetar command =CUSD for an $m \times 1$ array of $CUSD_i$, where $i = 1, 2, 3, \dots, m$.
2. Use the inverse transform formulas to simulate random values for each variable by applying the appropriate parameters for the variable's deterministic and stochastic components and the variable's respective $CUSD_i$.

– Simulating a Mixed Multivariate Distribution

Assume the mixed MV distribution to simulate has 4 random variables, defined and distributed as follows:

$$X \sim \text{Normal}(\bar{X}, \hat{\sigma})$$

$$Y \sim \text{Empirical}(S_i, P(S_i))$$

$$Z \sim \text{Empirical}(S_i, P(S_i))$$

$$W \sim \text{Uniform}(\min, \max)$$

The multivariate component for the mixed MV distribution is the correlation matrix for the four random variables. The correlation coefficients for the ρ matrix must be estimated using the residuals from the deterministic component for the X, Y, and Z variables and the historical data for variable W. The actual historical data has to be used for W because of the nature of the uniform distribution. The resulting correlation matrix is:

$$\rho = \begin{matrix} \begin{matrix} \rho_{\hat{x}_t, \hat{x}_t} & \rho_{\hat{x}_t, \hat{y}_t} & \rho_{\hat{x}_t, \hat{z}_t} & \rho_{\hat{x}_t, w_t} \\ \rho_{\hat{y}_t, \hat{y}_t} & \rho_{\hat{y}_t, \hat{z}_t} & \rho_{\hat{y}_t, w_t} \\ \rho_{\hat{z}_t, \hat{z}_t} & \rho_{\hat{z}_t, w_t} \\ \rho_{w_t, w_t} \end{matrix} \end{matrix}$$

Generate a vector of correlated uniform standard deviates (CUSDs) of size $m \times 1$ using the array function =CUSD as follows:

Block a 4×1 array and type the command =CUSD (Correlation Matrix Range), then press the Control Shift Enter keys.

The resulting 4×1 array has four values that change as the F9 key is pressed and during simulation, Simetar will generate new values for each iteration. The four values in the array are correlated based on the correlation matrix.

Apply the appropriate Simetar function to simulate the random numbers for each random variable being sure to use each variables $CUSD_i$,

$$\tilde{X} = \text{NORM}(\hat{X} + \hat{\sigma}, \text{CUSD}_1)$$

$$\tilde{Y} = \hat{Y} + \hat{Y} * \text{EMP} (S_{y_i}, F(S_i), \text{CUSD}_2)$$

$$\tilde{Z} = \hat{Z} + \hat{Z} * \text{EMP} (S_{z_i}, F(S_i), \text{CUSD}_3)$$

$$\tilde{W} = \text{UNIFORM}(\text{Min}, \text{Max}, \text{CUSD}_4)$$

The Simetar functions for generating random variables all allow for specification of a random uniform standard deviate (see Chapter 16). In the case of a MV distribution you must specify a CUSD rather than an independent USD.

The Multivariate Mixed Probability Distribution Demo.XLS spreadsheet provides an example of how to simulate random variables that have different probability distributions, as a multivariate distribution. The first variable is assumed to be normally distributed, the second and third are distributed empirical and the last is distributed uniform. The MV distributions simulated as indicated in (cells B79-B82) use their respective CUSD's in the inverse transform formulas Figure 7.12.

	A	B	C	D	E	F	G	H	I	J
77		Random								
78		Values				Mean	Std Dev	Min	Max	CUSDs
79	Var X Normal	2.1823438	=NORM(F79,G79,J79)			2.127	0.369			0.5595
80	Var Y EMP	2.5497697	=F80*(1+EMP(F32:F43,E32:E43,J80))			2.361				0.7477
81	Var Z EMP	172.11025	=F81*(1+EMP(G32:G43,E32:E43,J81))			160.7				0.6124
82	Var W Uniform	89.959372	=UNIFORM(H82,I82,J82)					61.310	92.740	0.9115

Figure 7.12. Example of Simulating a MV Mixed Distribution Using CUSD's and the Inverse Transform Functions.

Four validation tests were used to statistically determine if the MV mixed distribution reproduced the historical correlation matrix, means vector, and covariance matrix. The null hypothesis that the simulated test statistic (ρ , \bar{X} and Σ) equal their historical counterparts were not rejected, indicating that the procedure worked. The test statistics are reported in Figure 7.13.

	H	I	J	K	L	M	N	O	P	Q	R	S	T
8	Test Correlation Coefficients												
9	Confidence Level	99.1488%											
10	Critical Value	2.64											
11													
12		Var Y EM	Var Z EMP	Var W Uniform									
13	Var X Nor	1.58	0.15	0.34									
14	Var Y EMP		0.49	1.06									
15	Var Z EMP			0.71									
16													
17	Distribution Comparison of Var X Normal & Var X												
18	Confidence Level	95.0000%											
19		Test Value	Critical Value	P-Value									
20	2 Sample	4.52	9.62	0.344	Fail to Reject the Ho that the Mean Vectors are Equal								
21	Box's M T	1.42	18.31	0.999	Fail to Reject the Ho that the Covariance Matrices are Equivalent*								
22	Complete	2.00	23.68	1.000	Fail to Reject the Ho that the Mean Vectors and Covariance Matrices are Equivalent, Respectively								
23					*Based on asymptotic distribution								

Figure 7.13. Validation Tests for a MV Mixed Distribution.

Simulating Large Multivariate Distributions

For multivariate (MV) distributions containing a large number of variables, it is often impossible to use the procedures described for the MVN and MVE distributions. The symptom that the MVE and MVN methods will not work, is that Excel returns “# VALUE” in the CUSD or CSND arrays. The function fails because the correlation matrix is not positive definite, and therefore the correlation matrix (ρ) can not be factored by the square root method (Choleski decomposition). A factored matrix (ρ) is required to simulate the MVE and MVN distributions. The equations for simulating the MVE and MVN distributions show the dependence of the method on the R matrix:

$$\begin{aligned} \text{Let } \rho &= N \times N \text{ correlation matrix} \\ \text{and } R_{(N \times N)} &= \sqrt{\rho}_{(N \times N)} \\ \text{CSND}_{(N \times 1)} &= R_{(N \times N)} * \text{ISND}_{(N \times 1)} \\ \text{also } \text{CUSD}_{(N \times 1)} &= \text{ERF}(R_{(N \times N)} * \text{ISND}_{(N \times 1)}) \end{aligned}$$

where: ERF is the error function for intergrating the area under a standard normal distribution from $-\infty$ to z and is calculated using Excel’s function =NORMSDIST.

The Choleski decomposition of the correlation matrix is calculated in Simetar using the =MSQRT (Correlation Matrix) function. If the correlation matrix is positive definite, the MSQRT function will return a non-zero value in every cell in the upper right triangle and the main diagonal of the result matrix. On the other hand, if MSQRT returns a matrix with zeros in the upper right triangle or main diagonal (or #VALUE) the distribution cannot be simulated using the MVE or MVN procedure. See the correlation matrix and its factored matrix in Figure 7.14 for an example of a matrix which is not positive definite. The example comes from the Bad Correlation Matrix Demo.XLS.

As a further test of a problem matrix, the determinate of the full symmetric correlation matrix or the covariance matrix should be calculated to make sure it is positive or that it is not so close to zero it causes exponent overflows in =MSQRT. The correlation icon on the Simetar toolbar can be used to calculate the full symmetric correlation or covariance matrix. The determinate of a square matrix can be calculated using the determinate function in the Simetar Matrix Operations dialog box. If the determinate for either the correlation or covariance matrix is negative or almost zero the matrix cannot be factored. This result is generally due to the fact that there are too many “large” correlation coefficients or the number of correlation coefficients outside the ± 0.50 causes the MV distribution to be over specified.

When the MVE and MVN procedures described in this Chapter cannot be used, there are two options for simulating a MV distribution: (a) use a bootstrap simulation technique or (b) re-arrange the correlation matrix.

	A	B	C	D	E	F	G	H	I	J	K
21	Correlation Matrix of 10x10 Variables										
22		Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9	Var 10
23	Var 1	1	0.21	-0.05	0.30	-0.12	0.35	0.09	0.19	-0.47	0.00
24	Var 2		1	0.78	0.76	0.89	0.64	0.84	0.60	0.17	0.52
25	Var 3			1	0.80	0.93	0.79	0.88	0.26	0.68	0.62
26	Var 4				1	0.79	0.80	0.92	0.33	0.26	0.70
27	Var 5					1	0.66	0.91	0.35	0.50	0.60
28	Var 6						1	0.79	0.26	0.43	0.40
29	Var 7							1	0.32	0.34	0.69
30	Var 8								1	-0.06	0.46
31	Var 9									1	0.31
32	Var 10										1
33											
34		Factor a 10x10 Square, Symmetric Matrix. 12/14/2003 3:46:01 PM									
35		#####	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
36		#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
37		#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
38		#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
39		#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
40		#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
41		#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
42		#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
43		#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
44		#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!

Figure 7.14. Example of a Correlation Matrix that Can Not be Factored by the Square Root Method.

– **Bootstrapping a Multivariate Distribution**

Bootstrap methodology can be used to simulate a multivariate empirical distribution if there are a large number of observations for the random variables. Bootstrap simulation is based on Elfon’s (1979) work on bootstrapping univariate distributions. For a complete description of univariate bootstrapping see Chapter 11. If applied properly bootstrap simulation of MV distributions maintains not only the correlation among variables, but also higher order moments and any multi-modal characteristics of the variables. To simulate a MV distribution using bootstrap simulation methodology, do the following:

1. Prepare the historical data in a table with all random variables contiguous and in the proper temporal order, $X_{1i}, X_{2i}, \dots X_{mi}$ for m variables, with years i for the rows. See Figure 7.15 for a data table in the proper format.
2. Use Simetar’s =BOOTSTRAPPER function to randomly draw rows from the data matrix. The array function is programmed as:

=BOOTSTRAPPER(Data Matrix, TRUE)

where: TRUE instructs Simetar to draw m values from the m columns of the Data Matrix and that all of M values must come from one row.

Simetar draws rows of values from the data matrix at random if the function is used properly as an array function and is concluded by pressing Control Shift Enter. See the example in Figure 7.15 taken from the Bootstrap Multivariate Distribution Demo.XLS.

	A	B	C	D	E	F	G	H	I	J
77		Random								
78		Values				Mean	Std Dev	Min	Max	CUSDs
79	Var X Normal	2.1823438	=NORM(F79,G79,J79)			2.127	0.369			0.5595
80	Var Y EMP	2.5497697	=F80*(1+EMP(F32:F43,E32:E43,J80))			2.361				0.7477
81	Var Z EMP	172.11025	=F81*(1+EMP(G32:G43,E32:E43,J81))			160.7				0.6124
82	Var W Uniform	89.959372	=UNIFORM(H82,I82,J82)					61.310	92.740	0.9115

Figure 7.15. Using the Bootstrapping Function to Simulate a MV Distribution.

The benefits to the bootstrap simulation methodology for a MV empirical distribution are: (a) it draws random \tilde{X}_j observations in historically observed paired groups so the historical intratemporal correlation is maintained, (b) the historical data are used to simulate each variable's distribution so no assumptions about distributional shapes or types are required, and (c) the method is efficient in that no parameters must be calculated. The disadvantage to the methodology is that the values simulated are the discrete historical values because interpolation between observed points is not employed. To minimize this being a significant disadvantage, the original sample in the data matrix must be sufficiently large to define the population being simulated, as no new values are generated by interpolating among observed values.

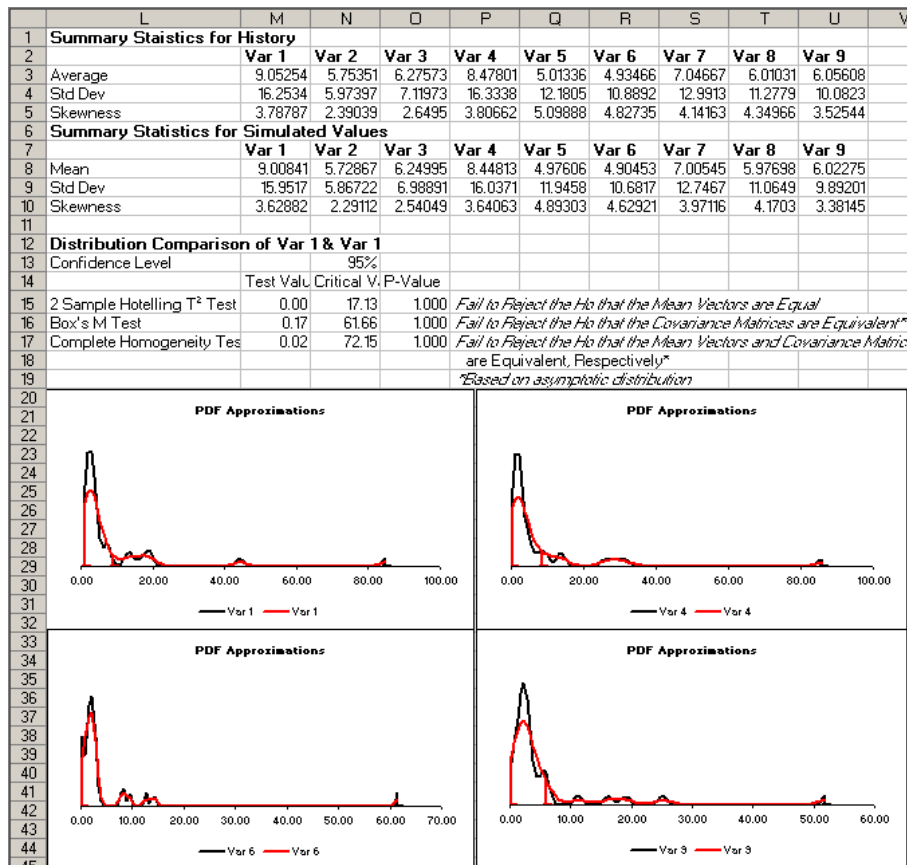


Figure 7.16. Validation Tests and PDFs for a Bootstrapped MV Distribution.

When using bootstrap simulation it is recommended that the number of iterations should be expanded to 1000 or more (Conover). The question remains, how well does this method work? For the example MVE in Figure 7.15, the results from a 1,000 iteration simulation are summarized in Figure 7.16. The means for the simulated and historical series are statistically equal as are the covariance matrices, based on three non-parametric validation tests. The skewness for the distributions are also very similar between the simulated and the historical distributions. The PDFs for four of the random variables are presented to show that the simulated distributions conform closely to the historical distributions. In particular the PDFs show that the multi-modal aspects of the historical distributions are matched by the simulated values.

– **Two Stage Correlation for Multivariate Distributions**

The second method for simulating a MV distribution defined by a correlation matrix that will not factor is to make two correlation matrices and factor the two matrices. The new correlation matrix is re-estimated after eliminating one or two variables and a second correlation matrix between the deleted variables and one remaining variable is estimated.

To demonstrate the procedure, the 10x10 correlation matrix in Figure 7.14 is used because it will not factor. The 3rd variable in the original data set is removed and a new correlation matrix is estimated (Figure 7.17). The 3rd variable was deleted because of its large number of correlation coefficients outside the ± 0.5 range. (The 5th variable could have been the eliminated variable for the same reason.) The resulting 9x9 correlation matrix factors as demonstrated by the third part of Figure 7.17. The second correlation matrix is a 2x2 between variables 3 and 5 (or any other variable remaining in the 9x9) (see Figure 7.18).

Once the two correlation matrices have been estimated and factored, the CSNDs for the 10 variables are generated in three steps.

1. Use the first correlation matrix and the =CSND (9x9 Correlation Matrix) function to generate an array of nine CSNDs (Step 1 in Figure 7.18).
2. Use the second correlation matrix and the =CSND (2x2 Correlation Matrix, 2x1 ISND Array) function to generate two CSNDs for variables 3 and 5 (Step 2 in Figure 7.18). The two ISNDs in the CSND function are actually one ISND for variable 3 (=NORM()) and the CSND₅ generated for variable 5 in Step 1.
3. Assemble the 10 CSNDs into a final 10x1 array using the CSNDs for variables 1-2 and 4-9 from Step 1 and variable 3’s CSND from Step 2. In Figure 7.18 the final array is G96:G105.

The assembled 10x1 CSND array can be used to generate MVN values or converted to CUSDs using =NORMSDIST() and used to generate MVE values.

	A	B	C	D	E	F	G	H	I	J	
52	Drop out the Var 3 Column and re-build the Correlation Matrix										
53		Var 1	Var 2	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9	Var 10	
54		1	25	22	38	25	31	24	31	32	30
55		2	33	34	35	30	25	22	44	29	33
56		3	23	30	32	30	22	23	30	24	26

Before using the CSNDs for generating random numbers, they need to be tested to insure that they are appropriately correlated. To test the correlation select the assembled array of CSNDs as the simulation output variables, simulate the workbook, and test the implicit correlation for the simulated values against the original 10x10 correlation matrix. The results of the correlation test for the problem in Figure 7.18 indicate that all but two of the correlation coefficients in the matrix are statistically correlated as they were over the historical period. If the correlation test indicates that many of the correlation coefficients are statistically different from the original matrix, select a different variable to eliminate and repeat the process. Experience suggests that the variable with the most large correlation coefficients is the best candidate for elimination and for some problems more than one variable must be eliminated and then added back via the second matrix. The variable selected to correlate back to (5 in this example) must be highly correlated to the eliminated variables and to the remaining variables in the correlation matrix. The Bad Correlation Matrix Demo.XLS workbook contains the example described in this section.

	A	B	C	D	E	F	G	H	I
87	Use the correlation coefficient from the original								
88	matrix between Var 3 and 5, to create a new correlation matrix.								
89		Var 3	Var 5						
90	Var 3	1	0.9311	Note that Variable 5 is the independent variable in the matrix					
91	Var 5		1	and Variable 3 depends upon Variable 5.					
92									
93	Step 1. CSNDS for 9 variable								
94	Correlation Matrix				Step 3. Assemble the CSNDS for the				
95		CSND			10 Variables and use the CSNDS.				
96	Var 1	-0.13614	=CSND(B66:J74)		Var 1	-0.13614			
97	Var 2	-0.4416			Var 2	-0.4416			
98	Var 4	-0.01293			Var 3	-0.67602			
99	Var 5	-0.20708			Var 4	-0.01293			
100	Var 6	-0.71311			Var 5	-0.20708			
101	Var 7	-0.18983			Var 6	-0.71311			
102	Var 8	-0.21889			Var 7	-0.18983			
103	Var 9	0.08733			Var 8	-0.21889			
104	Var 10	0.85187			Var 9	0.08733			
105					Var 10	0.85187			
106	Step 2. Calculate CSNDS for Variables 3 and 5								
107	For this step use the CSND generated for Variable 5 in Step 1								
108		CSND for Var 3		CSNDs					
109	Var 3	-1.324787	=NORM()	-0.676024	=CSND(B90:C91,B109:B110)				
110	Var 5	-0.207082	=B99	-0.207082					

Figure 7.18. Steps for Simulating CSNDS from a 9x9 and a 2x2 Correlation Matrix.

Appendix: Simulation of a MVN Distribution

In matrix notation a four variable MVN distribution is simulated using the R matrix, which is the factored correlation matrix. The Choleski decomposition can be used to factor the correlation matrix, i.e., to calculate the square root of the correlation matrix. Appendix Figure 7.1 shows the matrix notation for the steps to simulate a four variable MVN distribution. The ISND vector is an Mx1 vector of independent standard normal deviates which is multiplied by the R matrix to calculate a vector of correlated standard normal deviates (CSND's). Multiplying the MxM standard deviation diagonal matrix by a vector of CSND's and adding the product to the means gives a vector of correlated stochastic values that are distributed MVN. The CSND vector is not presented in Appendix Figure 7.1 as this is an intermediate step, and is shown in Step 8 of the Multivariate Normal Distribution Demo.XLS.

$$\underbrace{\begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \tilde{X}_3 \\ \tilde{X}_4 \end{bmatrix}}_{\text{Stochastic Results}} = \underbrace{\begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \\ \hat{X}_3 \\ \hat{X}_4 \end{bmatrix}}_{\text{Deterministic Component}} + \underbrace{\begin{bmatrix} \hat{\sigma}_1 & & & \\ & \hat{\sigma}_2 & & \\ & & \hat{\sigma}_3 & \\ & & & \hat{\sigma}_4 \end{bmatrix}}_{\text{Stochastic Component}} * \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ & r_{22} & r_{23} & r_{24} \\ & & r_{33} & r_{34} \\ & & & r_{44} \end{bmatrix}}_{\text{Correlation Component Results in CSNDs}} * \underbrace{\begin{bmatrix} \text{ISND}_1 \\ \text{ISND}_2 \\ \text{ISND}_3 \\ \text{ISND}_4 \end{bmatrix}}_{\text{ISND Vector}}$$

Appendix Figure 7.1. Simulation of a MVN Distribution.

The MVN distribution simulation depicted in Appendix Figure 7.1 can be written as individual equations for each random variable, much like a univariate normal variable. Recall that a univariate normal variable is simulated as:

$$\tilde{X} = \hat{X} + \sigma_{\tilde{e}} * \text{SND}$$

where SND is an independent standard normal deviate distributed $N(0,1)$.

For a MVN distribution each random variable is simulated as:

$$\tilde{X}_i = \hat{X}_i + \hat{\sigma}_{\tilde{e}_i} * \text{CSND}_i$$

where CSND_i is the i^{th} correlated standard normal deviate in the CSND vector.

Simetar calculates the CSND vector two ways. Both methods are used in Multivariate Normal Distribution Demo.XLS and are described here:

1. The user can specify the vector of ISND's. The 3 variable MVN distribution example in

Appendix Figure 7.2 shows that the vector of ISND's is provided as input in column I. The CSND function is an array function so highlight 3 cells and enter the function =CSND (correlation matrix, ISND vector) and press Control Shift Enter.

	E	F	G	H	I	J	K	L
121	CSNDs	Correlation Matrix			ISNDs	Formula in Column E		
122	1.644655	1.0000	0.2335	0.6943	1.977	=CSND(F122:H124,I122:I124)		
123	0.47652	0.0000	1.0000	0.2373	0.421			
124	0.286447	0.0000	0.0000	1.0000	0.286			

Appendix Figure 7.2. Example of CSND Calculated With the ISND Vector Provided as Input.

- The user does not specify the vector of ISND's. The 3 variable MVN distribution example in Appendix Figure 7.3 shows that the CSND vector is calculated using only the correlation matrix as input. In this case Simetar generates the ISND's without the user explicitly including this step.

	E	F	G	H	I	J	K
113	CSNDs	Correlation Matrix				Formula in Column E	
114	-1.0016	1.0000	0.2335	0.6943		=CSND(F114:H116)	
115	-1.3480	0.0000	1.0000	0.2373			
116	-0.8221	0.0000	0.0000	1.0000			

Appendix Figure 7.3. Example of CSND Vector Calculated Without Explicitly Providing the ISND Vector.

References

- Clements, A.M., Jr., H.P. Mapp, Jr., and V.R. Eidman. "A Procedure for Correlating Events in Farm Firm Simulation Models." Technical Bulletin T-131, Oklahoma Agricultural Experiment Station, August 1971.
- Conover, W. J. Practical Nonparametric Statistics. New York: John Wiley & Sons, Inc. 1999.
- Eidman, V.R. (editor). "Agricultural Production Systems Simulation." Proceedings of a Workshop by the Southern Farm Management Research Committee. Stillwater: Oklahoma State University, May 1971.
- Elfon, B. "Bootstrap Methods: Another Look at the Jackknife." Annals of Statistics 7(1979): 1-26.
- Fackler, P.L. "Modeling Interdependence: An Approach to Simulation and Elicitation." American Journal of Agricultural Economics, 73(1991): 1091-1097.
- King, R.P. "Operational Techniques for Applied Decision Analysis Under Uncertainty." Ph.D. dissertation, Department of Agricultural Economics, Michigan State University, 1979.
- McCarl, B. "Forming Probability Distributions." Department of Agricultural Economics, Texas A&M University, 1988.
- Mjelde, J.W., D.P. Anderson, K. Coble, B. Maufluk, J.L. Outlaw, J.W. Richardson, J.R. Stokes, and V. Sundarapotes. "Tutorial on Density Function Estimation and Use." Department of Agricultural Economics, Texas A&M University, FP-94-2, 1993.
- Richardson, J.W., and G.D. Condra. "A General Procedure for Correlating Events in Simulation Models." Department of Agricultural Economics, Texas A&M University, 1978.
- Taylor, C. Robert. "Two Practical Procedures for Estimating Multivariate Nonnormal Probability Density Functions." American Journal of Agricultural Economics, 72(1990): 210-217.