A Procedure for Correlating Events in Farm Firm Simulation Models

Alvin M Clements, Jr
Harry P. Mapp, Jr.
Vernon R. Eidman

AGRICULTURAL EXPERIMENT STATION
OKLAHOMA STATE UNIVERSITY

Technical Bulletin T-131

August 1971

Contents

Acknowledgement

The assistance of Dr. Michael Boehlje, Department of Agricultural Economics, in reviewing an early draft of this bulletin and the advice of Dr. P. L. Claypool, Department of Mathematics and Statistics, in a phase of the statistical tests is gratefully acknowledged. The authors accept final responsibility for the content of his bullein.

A Procedure for Correlating Events in Farm Firm Simulation Models

Alvin M. Clements, Jr.*, Harry P. Mapp, Jr.*,
and Vernon R. Eidman**

Department of Agricultural Economics

Simulation has been defined as "essentially a technique that involves setting up a model of a real situation and then performing experiments on that model" [7, p. 2]. Many research and extension workers agree that simulation models can more realistically represent certain aspects of a farm firm (such as uncertain prices and yields, economies of size, discontinuities, indivisibilities and use of decision rules which are not simple minimization or maximization rules) than linear programming, budgeting and other models used in farm management work. However, the degree of realism incorporated in any model, including simulation models, is relative. One can always find additional features that might have been added to make the model more realistically represent the real world.

Simulation models frequently incorporate Monte Carlo procedures to represent uncertainty facing the farm firm. The Monte Carlo applications typically assume that the correlation between any two events (such as the yield of two crops in a given year) is either nonexistent (zero) or perfect (one). This assumption does not realistically represent the covariance between related events and may even introduce artificial and unrealistic variability into the analysis. In some studies the resulting bias may not be great, while in others it appears to be of major importance. Unfortunately, the researcher must rely on his judgement to determine the extent of the bias.

This bulletin discusses a procedure that can be incorporated in Monte Carlo simulation models to correlate two events at any desired level from minus one to plus one. The procedure has been applied to the correlation of two product prices [3, p. 153-154], but can be applied

*Research Assistant
**Associate Professor

Research reported herein was conducted under Oklahoma Station Project 1358 and Grant 14-01-0001-1539 from the Office of Water Resources Research.

to events such as production levels, weather effects, disease and insect infestations, irrigation water requirements, machinery requirements, input prices or other events of interest in a particular application. An example of generating a series of correlated production levels is used to illustrate the procedure.

The following section illustrates the correlation procedure for the two-event case. Then generalized equations are presented for use in correlating outcomes for more than two events and an example of generating production levels for four enterprises is discussed. The final section discusses means of applying this procedure when trend, seasonal and cyclical effects are important. Much of the detail for the examples as well as the computer programs used in applying the procedure are given in the appendices.

Establishing Correlated Outcomes for the Two-Event Case

The Monte Carlo technique can be used to generate outcomes having the desired variance and covariance values for any number of events, assuming the outcomes for each event are normally distributed.[1] The simplest case is generation of a series of n outcomes for two events. For purposes of illustration, assume the problem is to generate a series of yields for n production periods for two crops. Using matrix notation, the yield relationship may be written as

(1) $Y = \bar{Y} + AW$

where Y is a 2 x n matrix of generated yields, $\bar{Y}$ is a 2 x 1 matrix of expected yields for the two enterprises, A is a 2 x 2 matrix of coefficients, and W is a 2 x n matrix of random normal deviates. A Fortran IV computer program has been written which generates correlated yield series based upon the yield relationship in Equation (1). The 2 x 1 matrix of values for $\bar{Y}$ is calculated externally based on available data for the two events. A set of 2 x n values for W is generated by calling Subroutine GAUSS, or a similar random normal deviate generator, available at most computer facilities. Coefficients for the A matrix are calculated externally and are read as data. Having read in the mean values ($\bar{Y}$), the coefficients of A, and generated n sets of two random normal deviates (W), (1) is solved to provide n sets of two yields (Y). The procedure is very straightforward, with perhaps the exception of calculating the coefficients of the A matrix.

[1]The terms "outcomes" and "events" are used throughout this bulletin. An "event" refers to an activity or enterprise. Two events might be the crops wheat and grain sorghum. An "outcome" refers to the yield or price for one event in one time period.

Development of the A Matrix

The variance-covariance matrix of Y (Σ_y) may be expressed as follows [3, p. 153]:

$$(2) \quad \Sigma_y = E(AWW'A')$$

where A is an upper triangular matrix.[2]

Since WW′ is composed of random normal deviates with expected value of zero and variance of one, the expected value of Equation (2) gives the following expression for the variance-covariance matrix of Y.

$$(3) \quad \Sigma_y = AA'$$

Assuming the two events are wheat and grain sorghum yield per acre in a given year, the variance-covariance matrix in Equation (3) may be written as

$$(4) \quad \begin{bmatrix} \sigma_w^2 & \sigma_{wg} \\ \sigma_{wg} & \sigma_g^2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} \begin{bmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{bmatrix}$$

where σ_w^2 and σ_g^2 represent the variance of wheat and grain sorghum yield per acre, respectively, and σ_{wg} represents their covariance. Applying matrix multiplication to (4) results in (5).

$$(5) \quad \begin{bmatrix} \sigma_w^2 & \sigma_{wg} \\ \sigma_{wg} & \sigma_g^2 \end{bmatrix} = \begin{bmatrix} a_{11}^2 + a_{12}^2 & a_{12}a_{22} \\ a_{22}a_{12} & a_{22}^2 \end{bmatrix}$$

Calculation of the A Matrix for the Two-Event Case

Assume that wheat and grain sorghum yields are normally distributed random variables with mean yields of 12.51250 bushels and 10.94370 bushels per acre, respectively; variances of 17.97311 and 31.70388 bushels per acre, respectively; and the covarience is —5.79250. Substituting the appropriate values into (5) results in (6).

$$(6) \quad \begin{bmatrix} 17.97311 & -5.79250 \\ -5.79250 & 31.70388 \end{bmatrix} = \begin{bmatrix} a_{11}^2 + a_{12}^2 & a_{12}a_{22} \\ a_{22}a_{12} & a_{22}^2 \end{bmatrix}$$

The problem is to calculate the elements of a unique triangular matrix A of the form

$$(7) \quad A = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}$$

[2]It will be proven here that A and A′ are upper triangular and lower triangular matrives, respectively. The interested reader is referred to Wagner [8, p. 132] and Naylor [7, p. 98] for a general discussion or to Anderson [1, p. 19] for the appropriate theorem which indicates the existence of a unique triangular matrix A such that $Y = \bar{Y} + AW$.

The procedure is simple for a 2 x 2 matrix. Using values in (6), first calculate a_{22} by taking the square root of a_{22}^2.

(8) $a_{22}^2 = 31.70388$

(9) $a_{22} = 5.63062$

Use the calculated value for a_{22} to compute a_{12}.

(10) $a_{12}a_{22} = -5.79250$

Since $a_{22} = 5.63062$, then

(11) $a_{12} = -5.79250/5.63062$

or

(12) $a_{12} = -1.02875.$

Calculation of the final coefficient of the A matrix, a_{11}, also follows from Equation (6). Since

(13) $a_{11}^2 + a_{12}^2 = 17.97311$

then

(14) $a_{11}^2 = 17.97311 - (-1.02875)^2$
$= 17.97311 - 1.05833 = 16.91478$

or

(15) $a_{11} = 4.11276.$

Substituting the coefficients of the A matrix into Equation (1) results in (16).

(16) $$\begin{bmatrix} Y_{wi} \\ Y_{gi} \end{bmatrix} = \begin{bmatrix} 12.51250 \\ 10.94370 \end{bmatrix} + \begin{bmatrix} 4.11276 & -1.02875 \\ 0 & 5.63062 \end{bmatrix} \begin{bmatrix} W_{1i} \\ W_{2i} \end{bmatrix}$$

Using this procedure with n pairs of random normal deviate values (W_{1i} and W_{2i}), a series of n paired wheat and grain sorghum yields can be generated with the appropriate means, variances and covariances. The 2 x 2 A matrix represents the simplest case of correlating event series. Most applications require correlating values for more than two events. Thus it is useful to generalize the procedure for the m-event case.

Generalization of the Procedure for the m-Event Case

Let m refer to the number of events to be correlated in the general case. For purposes of illustration, assume the problem is to generate a series of yields (outcomes) for m enterprises (events) for each of n pro-

duction periods. Letting equation (1) represent the yield relationship, Y is a m x n matrix of generated yields, $\bar{Y}$ is a m x 1 matrix of expected yields, A is a m x m matrix of coefficients, and W is a m x n matrix of random normal deviates. Again, the only difficulty in applying the procedure is in calculating the coefficients of the m x m A matrix.

Perhaps the simplest procedure to use in calculating the A matrix coefficients is to calculate the a_{mm} coefficient first and proceed up the m^{th} column calculating $a_{m-1,m}$; $a_{m-2,m}$; . . . , until the last coefficient in the m^{th} column, a_{1m} is calculated. Next calculate coefficients in the m-1 column beginning with $a_{m-1,m-1}$, and proceeding until the last coefficient in the column, $a_{1,m-1}$, is calculated. Continue this procedure column by column, until the final coefficient in the A matrix, a_{11}, has been calculated.

The following three equations can be used to calculate the a_{ij}'s.[3]

$$(17)\quad a_{ii} = (\sigma_{ii}^2 - \sum_{k=i+1}^{m} a_{ik}^2)^{1/2},\ 1 \leqslant i \leqslant k < m$$

$$(18)\quad a_{im} = \sigma_{im}/\sigma_{mm},\ 1 \leqslant i \leqslant m\text{-}1$$

$$(19)\quad a_{ij} = \sigma_{ij} - \sum_{k=j+1}^{m} a_{ik}a_{jk}/a_{jj},\ 1 \leqslant i < j \leqslant m\text{-}1$$

Since A is an upper triangular matrix, $a_{ij} = 0$ for all $i > j$.

Illustration of the Procedure with a Four-Event Example

This example of the correlation procedure uses the sixteen years of production data for wheat and grain sorghum yields (in bushels per planted acre) and buy-sell steer and cow-calf yields (in pounds per acre of native range) presented by Greve, Plaxico and Lagrone [4, p. 10]. The yield relationship used to generate correlated yields, (1), may be rewritten as (20) for this four-event example.

$$(20)\quad \begin{bmatrix} Y_{wi} \\ Y_{gi} \\ Y_{si} \\ Y_{ci} \end{bmatrix} = \begin{bmatrix} 12.51250 \\ 10.94370 \\ 39.14990 \\ 23.95000 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix} \begin{bmatrix} W_{1i} \\ W_{2i} \\ W_{3i} \\ W_{4i} \end{bmatrix}$$

Calculation of the coefficients of the A matrix in Equation (20) follows the procedure described above and is explained in detail in Appendix A. The resulting A matrix is shown in equation (21). This

[3]These generalized equations were developed to calculate the a_{ij} elements of an upper triangular A matrix. A similar set of equations is presented by Naylor [7, p. 98] for calculation of the a_{ij} elements of a lower triangular A matrix.

equation can be used with n sets of four random normal deviates to generate a series of n years of wheat, grain sorghum, steer and cow-calf yields having the appropriate means, standard deviations, covariances and correlation coefficients.

$$(21)\quad \begin{bmatrix} Y_{wi} \\ Y_{gi} \\ Y_{si} \\ Y_{ci} \end{bmatrix} = \begin{bmatrix} 12.51250 \\ 10.94370 \\ 39.14990 \\ 23.95000 \end{bmatrix} + \begin{bmatrix} 3.88109 & -1.13808 & 1.25958 & -0.16869 \\ 0.00000 & 5.40159 & 0.48455 & 1.51391 \\ 0.00000 & 0.00000 & 3.85071 & 8.10785 \\ 0.00000 & 0.00000 & 0.00000 & 0.99599 \end{bmatrix} \begin{bmatrix} W_{1i} \\ W_{2i} \\ W_{3i} \\ W_{4i} \end{bmatrix}$$

Testing the Generated Values for the Four-Event Case

The reader interested in a rigorous mathematical proof of the procedure described in this bulletin is referred to an article by Wagner [8]. However, a comparison of the mean, standard deviation and correlation coefficients estimated from the original data with those computed from values generated from (21) provides a measure of empirical validity.

The 16 years of data presented by Greve, Plaxico and Lagrone [4, p. 10] were used to estimate the mean and standard deviation for each of the four enterprises as well as the appropriate covariance and correlation coefficients. These variance and covariance estimates were used in calculating the A matrix in (21). Then (21) was used to generate 1000 wheat, grain sorghum, cow-calf and steer production levels. The mean and standard deviation of these generated outcomes was calculated for each enterprise. In addition the appropriate covariances and correlation coefficients were also calculated.

Table 1 presents the means and standard deviations by events for the original data and for the generated yield series. A comparison of the original and generated values reveals very close agreement for all events. Statistical tests verify this conclusion. The difference between each set of means (original versus generated) was tested for statistical significance by computing a t statistic and comparing the computed values with

Table 1. Means and Standard Deviations by Events for Original and Generated Yield Series

Event	Data Source	Mean	Standard Deviation
Wheat	Original	12.51250	4.23947
(Bushels)	Generated	12.50207	4.22499
Grain Sorghum	Original	10.94370	5.63062
(Bushels)	Generated	10.92747	5.65981
Steers	Original	39.14490	8.97581
(Pounds)	Generated	38.76559	9.09685
Cow-calf	Original	23.95000	0.99599
(Pounds)	Generated	23.89412	0.99568

tabular values at the 95% significance level. The difference between each set of standard deviations was tested for statistical significance by computing a Chi Square X^2 statistic and comparing it with the appropriate tabular value for the 95% level. The differences observed between each set of means and each set of standard deviations are not significant. The tests and computations performed are discussed in Appendix D of this bulletin.

Table 2 presents correlation coefficients by events for the original and generated yield series. Again, close agreement is found between correlation coefficients calculated on the basis of the original data and those calculated on the basis of generated yield series. More variation exists between correlation coefficients than between means and standard deviations. Two factors should be emphasized. First, the correlation coefficients presented in Table 2 are calculated based on one series of 1000 outcomes for each of four events. For one of the events sufficient variation existed that a number of yield outcomes were originally negative. The computer program replaces all negative outcomes with zero. This process slightly alters the generated mean, standard deviation and correlation coefficients. Second, the values generated by the correlation program are sensitive to the base selected for the random normal deviate generator. The user might desire to experiment with alternative bases and select the one which provides the most satisfactory results.

Confidence intervals were calculated for each population correlation coefficient, p. The 95% confidence interval consists of an upper and lower limit. The upper limit is computed by adding to the sample correlation coefficient, r, 1.96 times the standard error of the sample coefficient. The lower limit is computed by subtracting from r 1.96 times the standard error of r. For each confidence interval computed, the population correlation coefficient falls within the 95% confidence interval. Thus, with confidence represented by a probability of .95, it can

Table 2. Coefficients of Correlation by Events for Original and Generated Yield Series.

Event	Data Source	Wheat	Grain Sorghum	Steers	Cow-Calf
Wheat	Original	1.00000	—0.24266	0.09152	—0.03979
	Generated	1.00000	—0.23702	0.09651	—0.02506
Grain Sorghum	Original		1.00000	0.27979	0.30094
	Generated		1.00000	0.26040	0.26094
Steers	Original			1.00000	0.90330
	Generated			1.00000	0.90023
Cow-Calf	Original				1.00000
	Generated				1.00000

be stated that each population value falls between the upper and lower limits set for the population coefficient. The computations performed are discussed in Appendix D.

Applying the Procedure

The mathematical procedure described can be used to correlate a wide variety of events as suggested in the introduction. Computer programs utilizing the procedure described are included in the appendices to facilitate application. The Fortran IV program included in Appendix C is divided into two parts. The first part generates a series of correlated values for the four-event case. The second part computes the mean, standard deviation, correlation coefficients and covariances of the generated values. The second part is used only to verify the accuracy of the procedure, is not necessary to generate correlated series of events and could be eliminated before integrating the procedure into a simulation model.

The first part of the program in Appendix C has been expanded to generate 1000 outcomes for 15 events. A copy of the expanded program, including the definition of arrays, array dimensions and an explanation of card input data required to execute the program is contained in Appendix B. The program in Appendix B can be used for applications involving two to fifteen events without changing array dimensions. However, applications with less than five events can be run more efficiently using the program in Appendix C.

Researchers may want to use this correlation procedure in generating observations for events involving trend, seasonal, cyclical or combinations of these movements. The trend variable may be particularly important in yields, while researchers may want to include all three types of movements in generating price series. In general, these types of movements can be considered by modifying equation (1) to include another term for trend, one for seasonal effect and one for cyclical effect. In an application in which only trend is important, the relationship could be written as (22) if trend is considered a linear function of time.

$$(22) \quad Y = \bar{Y} + BT + AW$$

In this case, B is m x 1 and T is a scalar taking the values 1, 2, ... , n, for the n time periods. Thus the equation for the first event in (22) is of the form

$$(23) \quad y_{1i} = \bar{y}_1 + b_1 t_i + a_{11} w_{1i} + \ldots . + a_{1m} w_{mi}$$

If trend is considered quadratic, the relationship may also be written as shown in (22), where B is now m x 2 and T is 2 x 1. The equation for the first event in (22), assuming trend is a quadratic function of time, is of the form

(24) $y_{1i} = \bar{y}_1 + b_{11}t_i + b_{12}t_i^2 + a_{11}w_{1i} + \ldots + a_{1m}w_{mi}$

Obviously, many other types of trend relationships could also be included.

If a cyclical or seasonal movement is important, equation (1) could be modified by adding the appropriate terms of a mathematical relationship, such as a sine or cosine curve, to include the desired effect. Applications involving both a trend and cyclical effect can be handled by including terms for each.

The computer programs included in Appendices B and C do not include provisions for incorporating trend, seasonal or cyclical effects. Appropriate modifications must be made to apply the procedure for correlating yields to problems that include these effects. In making such modifications, deviations about the trend and cyclical components included will have the correlations specified in developing the coefficients of the A matrix. However, the generated series of outcomes will not have the original correlations because trend and cyclical rather than just mean values have been incorporated into the generating equation. This suggests that researchers desiring to base the coefficients for the correlation procedure on a historical series including trend and/or cyclical values may want to remove trend and cyclical movements from the historical data before obtaining the residuals used to develop the variance-covariance matrix. This variance-covariance matrix can be used to develop the coefficients for the A matrix. Then the appropriate form of equation (1) can be specified, to include trend and/or cyclical components. The equation will generate a series of outcomes having trend, cyclical and random components similar to the historical series.

Summary and Conclusions

The procedure described in this bulletin can be used to correlate the outcome of two or more events at any desired level from minus one to plus one in Monte Carlo simulation models. Application of the correlation procedure will add another dimension of realism to many farm firm simulation models.

In general, the procedure involves defining an upper triangular A matrix of coefficients, calculating the numerical values of these coefficients using the variance-covariance matrix, and combining the estimated coefficients with a series of random normal deviates to generate the correlated outcomes. The procedure is illustrated briefly for both a two-event and a four-event case. Parameters of the distribution for the four-event case do not differ significantly from those of the original data. Applications of the procedure to events involving trend, cyclical and seasonal effects are also discussed.

Appendix A illustrates the calculation of coefficients for the A matrix in detail. Appendix B presents a general program capable of generating 1000 correlated observations for each of two to fifteen events. Appendix C discusses the program used to generate yield series and to verify the accuracy of the correlation procedure for the four-event case. Program modifications are also discussed for those readers who wish to verify results for correlation programs containing more than four activities. The statistical tests used to compare the mean, variance and correlation coefficients of the original and generated series are discussed in Appendix D.

The analysis in this bulletin has only considered one form of the probability distribution—the normal distribution. While the normal is very frequently used to describe the random nature of events in farm firm simulation models, researchers may also wish to use other forms of the probability distribution. To the authors' knowledge, the procedure described has not been tested with other probability distributions.

References

1. Anderson, T. W., *An Introduction to Multivariate Statistical Analysis,* John Wiley and Sons, New York, 1958.
2. *CRC Handbook of Tables for Probability and Statistics, Second Edition,* The Chemical Rubber Co., Cleveland, Ohio, 1968.
3. Eidman, Vernon R., "Optimum Production Plans for California Turkey Growers with Chance-Constrained Programming", unpublished Ph.D. Thesis, University of California, Berkeley, 1965.
4. Greve, Robert W., James S. Plaxico, and William F. Lagrone, *Production and Income Variability of Alternative Farm Enterprises in Northwest Oklahoma,* Bulletin B-563, Agricultural Experiment Station, Oklahoma State University, Stillwater, August, 1960.
5. Hoel, Paul G., *Introduction to Mathematical Statistics,* John Wiley and Sons, Inc., New York, 1964.
6. Mills, Frederick C., *Statistical Methods,* Henry Holt and Company, New York, 1955.
7. Naylor, Thomas H., Joseph L. Balintfy, Donald S. Burdick, and Kong Chu, *Computer Simulation Techniques,* John Wiley and Sons, New York, April, 1968.
8. Wagner, Harvey M., "A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations", *Econometrica,* Vol. 26, No. 1, January 1958, pp. 117-133.

APPENDIX A

Calculation of the A Matrix for a Four-Event Case

This appendix discusses in detail the calculation of the A matrix for a situation involving more than two events. The example chosen for illustration involves generating correlated production levels of wheat, grain sorghum, steers, and a cow-calf operation. The variance-covariance matrix of yield per acre was derived from the original production data presented by Greve, Plaxico and Lagrone [4, p. 10] and is presented in (25).

$$(25)\quad \Sigma_y = \begin{bmatrix} \sigma_w^2 & \sigma_{wg} & \sigma_{ws} & \sigma_{wc} \\ \sigma_{gw} & \sigma_g^2 & \sigma_{gs} & \sigma_{gc} \\ \sigma_{sw} & \sigma_{sg} & \sigma_s^2 & \sigma_{sc} \\ \sigma_{cw} & \sigma_{cg} & \sigma_{cs} & \sigma_c^2 \end{bmatrix} = \begin{bmatrix} 17.97311 & -5.79250 & 3.48258 & -0.16801 \\ -5.79250 & 31.70388 & 14.14041 & 1.50784 \\ 3.48258 & 14.14041 & 80.56517 & 8.07534 \\ -0.16801 & 1.50784 & 8.07534 & 0.99200 \end{bmatrix}$$

Since it has been shown that

$$(26)\quad \Sigma_y = AA'$$

where A and A′ are upper and lower triangular matrices, Equation (26) may be rewritten as (27).

$$(27)\quad \Sigma_y = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix} \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ a_{12} & a_{22} & 0 & 0 \\ a_{13} & a_{23} & a_{33} & 0 \\ a_{14} & a_{24} & a_{34} & a_{44} \end{bmatrix}$$

Performing the appropriate matrix multiplication results in Equation (28).

$$(28)\quad \Sigma_y = \begin{bmatrix} a_{11}^2+a_{12}^2+a_{13}^2+a_{14}^2 & a_{12}a_{22}+a_{13}a_{23}+a_{14}a_{24} & a_{13}a_{33}+a_{14}a_{34} & a_{14}a_{44} \\ a_{12}a_{22}+a_{13}a_{23}+a_{14}a_{24} & a_{22}^2+a_{23}^2+a_{24}^2 & a_{23}a_{33}+a_{24}a_{34} & a_{24}a_{44} \\ a_{13}a_{33}+a_{14}a_{34} & a_{23}a_{33}+a_{24}a_{34} & a_{33}^2+a_{34}^2 & a_{34}a_{44} \\ a_{14}a_{44} & a_{24}a_{44} & a_{34}a_{44} & a_{44}^2 \end{bmatrix}$$

To derive the elements of the upper triangular A matrix in Equation (27) used to generate correlated yields, proceed by calculating the elements of the fourth column, beginning with a_{44} and ending with a_{14}. Equations (25) and (28) provide the necessary numerical data.

$$(29)\quad a_{44}^2 = .99200$$
$$a_{44} = .99599$$

$$(30)\quad a_{34}a_{44} = 8.07534$$
$$a_{34} = 8.07534/.99599$$
$$a_{34} = 8.10785$$

$$(31)\quad a_{24}a_{44} = 1.50784$$
$$a_{24} = 1.50784/0.99599$$
$$a_{24} = 1.51391$$
$$(32)\quad a_{14}a_{44} = -0.16801$$
$$a_{14} = -0.16801/.99599$$
$$a_{14} = -0.16869$$

Elements of the third column are calculated next, beginning with a_{33} and ending with a_{13}.

$$(33)\quad a_{33}^2+a_{34}^2 = 80.56517$$
$$a_{33}^2 = 80.56517 - (8.10785)^2$$
$$a_{33}^2 = 80.56517 - 65.73723$$
$$a_{33}^2 = 14.82794$$
$$a_{33} = 3.85071$$
$$(34)\quad a_{23}a_{33}+a_{24}a_{34} = 14.14041$$
$$a_{23}a_{33} = 14.14041 - 12.27456$$
$$a_{23} = 1.86585/3.85071$$
$$a_{23} = 0.48455$$
$$(35)\quad a_{13}a_{33}+a_{14}a_{34} = 3.48258$$
$$a_{13}a_{33} = 3.48258 - (-1.36771)$$
$$a_{13} = 4.85027/3.85071$$
$$a_{13} = 1.25958$$

Elements of the second column of the A matrix are calculated next followed by the only element in the first column, a_{11}.

$$(36)\quad a_{22}^2+a_{23}^2+a_{24}^2 = 31.70388$$
$$a_{22}^2 = 31.70388 - .23479 - 2.29192$$
$$a_{22}^2 = 29.17717$$
$$a_{22} = 5.40159$$
$$(37)\quad a_{12}a_{22}+a_{13}a_{23}+a_{14}a_{24} = -5.79250$$
$$a_{12}a_{22} = -5.79250 - .61033 - (-0.25538)$$
$$a_{12} = -6.14745/5.40159$$
$$a_{12} = -1.13808$$
$$(38)\quad a_{11}^2+a_{12}^2+a_{13}^2+a_{14}^2 = 17.97311$$
$$a_{11}^2 = 17.97311-1.29523-1.58654-0.02846$$
$$a_{11}^2 = 15.06288$$
$$a_{11} = 3.88109$$

Coefficients calculated in (29) through (38) are used to form the upper triangular A matrix (39).

$$(39)\quad A = \begin{bmatrix} 3.88109 & -1.13808 & 1.25958 & -0.16869 \\ 0.00000 & 5.40159 & 0.48455 & 1.51391 \\ 0.00000 & 0.00000 & 3.85071 & 8.10785 \\ 0.00000 & 0.00000 & 0.00000 & 0.99599 \end{bmatrix}$$

This is the same A matrix, shown in Equation (22) of the body of this bulletin, which was utilized in generating appropriately correlated yield series for the four events considered.

Coefficients for the A matrix may also be calculated using generalized equations, (17), (18) and (19), presented in the body of this bulletin. The first generalized equation, rewritten below, allows calculation of A matrix coefficients a_{44}, a_{33}, a_{22}, and a_{11}.

$$(40)\quad a_{ii} = (\sigma_{ii}^2 - \sum_{k=i+1}^{n} a_{ik}^2)^{1/2}\ ,\ 1 \leqslant i \leqslant k < m$$

The above four coefficients are calculated as follows from Equation (40).

$$(41)\quad a_{44} = (\sigma_{44}^2)^{1/2}$$
$$(42)\quad a_{33} = (\sigma_{33}^2 - a_{34}^2)^{1/2}$$
$$(43)\quad a_{22} = (\sigma_{22}^2 - a_{23}^2 - a_{24}^2)^{1/2}$$
$$(44)\quad a_{11} = (\sigma_{11}^2 - a_{12}^2 - a_{13}^2 - a_{14}^2)^{1/2}$$

The four preceding equations correspond to Equations (29), (33), (36) and (38), respectively.

The second generalized equation, restated below, may be used to calculate coefficients a_{14}, a_{24} and a_{34} of the A matrix.

$$(45)\quad a_{im} = \sigma_{im}/\sigma_{mm}\ ,\ 1 \leqslant i \leqslant m\text{-}1$$

The above three coefficients are calculated as follows from Equation (45).

$$(46)\quad a_{14} = \sigma_{14}/\sigma_{44}$$
$$(47)\quad a_{24} = \sigma_{24}/\sigma_{44}$$
$$(48)\quad a_{34} = \sigma_{34}/\sigma_{44}$$

The three preceding equations correspond to Equations (32), (31) and (30), respectively.

The third generalized equation, restated below, is used to calculate A matrix coefficients a_{12}, a_{13}, and a_{23}.

$$(49)\quad a_{ij} = \sigma_{ij} - \sum_{k=j+1}^{m} a_{ik}a_{jk}/a_{jj}\ ,\ 1 \leqslant i < j \leqslant m\text{-}1$$

The three coefficients are calculated as follows from Equation (49).

$$(50)\quad a_{12} = (\sigma_{12} - a_{13}a_{23} - a_{14}a_{24})/a_{22}$$
$$(51)\quad a_{13} = (\sigma_{13} - a_{14}a_{34})/a_{33}$$
$$(52)\quad a_{23} = (\sigma_{23} - a_{24}a_{34})/a_{33}$$

The three preceding equations correspond to Equations (37), (35), and (34), respectively.

The validity of the generalized equations has been demonstrated. Regardless which method of calculation is utilized, certain coefficients

must be known before other coefficients can be computed. Consequently, even when the generalized equations are used, it is necessary to begin with the elements in the mth column of the A matrix, then move to the m-1 column followed by the m-2 column, etc., as suggested previously.

APPENDIX B

Definition of Arrays, Discussion of Array Dimensions and Explanation of Correlation Program

The correlation program discussed in this appendix can be incorporated in a simulation model with a minimum of modification. This appendix identifies the relevant arrays, discusses the array dimensions and explains the attached correlation program which is designed to accommodate a maximum of 15 events and generate 1000 outcomes for each event.

The relevant arrays and their dimensions are as follows:

A(ACT,ACT): This matrix contains coefficients for the A matrix. These coefficients are calculated externally and are read from cards. ACT represents the number of events to be included within the program. The attached program accommodates a maximum of 15 events and A is dimensioned as follows: A(15,15).

AMEAN(ACT): The AMEAN array contains the mean of each event to be included in the correlation program. These means are calculated externally and are read into the correlation program from cards. ACT represents the number of events to be correlated. Since 15 events are included in the attached program, AMEAN is dimensioned as follows: AMEAN(15).

X(ACT,OBS): The X matrix receives and retains the random normal deviates generated by calling SUBROUTINE GAUSS, or a comparable random normal deviate generator. ACT represents the number of events and OBS the number of outcomes for each event. If 1000 outcomes are desired for each of 15 events, then X is dimensioned X(15,1000), as in the attached correlation program.

D1(ACT,ACT): The D1 matrix is used in matrix multiplication of the A and X matrices. The A matrix is multiplied by each row of the X matrix, the results are temporarily stored in D1 and are subsequently passed to matrix AK. The D1 matrix is dimensioned according to the number of events to be considered, or D(15,15) in the attached program.

AK(ACT,OBS): The matrix AK contains the product of the A and X matrices. Since A is a 15 x 15 matrix and X is a 15 x 1000 matrix in the attached program, their product, AK, must be dimensioned as follows: AK(15,1000).

Y(ACT,OBS): The Y matrix contains the final outcomes, appropriately correlated. It is obtained by adding the mean of each event from AMEAN to the product of the A and X matrices (AK). Y is dimensioned as Y(15,1000) in the attached program to represent 1000 outcomes for each of 15 events.

IX: Any odd integer value with seven or less digits may be chosen as the base for the random normal deviate generator. In this program, IX was arbitrarily set equal to 99999.

The following program is designed to be quite general in nature. Correlated outcomes may be generated for a minimum of two and a maximum of 15 events without changing array dimensions or Fortran statements. Naturally, larger numbers of events or outcomes may be accommodated within the same general framework by simply enlarging the array dimension and changing READ statements, DO LOOP parameters and the WRITE statement. Comment cards have been added to enable the reader to locate relevant portions of the program if changes are required to handle his problem. Since the event means and A matrix coefficients are read from cards, comment cards have been added to the program to explain the data input requirements.

```
                                        80/80 LIST

      00000000011111111112222222222333333333344444444455555555556666666666777
      12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012
CARD
0001  C*************************************************************************
0002  C*************************************************************************
0003  C                          CORRELATION PROGRAM
0004  C*************************************************************************
0005  C*************************************************************************
0006  C
0007  C*************************************************************************
0008  C*************************************************************************
0009  C     THIS PROGRAM IS DESIGNED TO GENERATE SERIES OF CORRELATED
0010  C     EVENTS. IT WILL ACCOMODATE A MAXIMUM OF 15 EVENTS AND GENERATE
0011  C     1000 OUTCOMES FOR EACH EVENT WITHOUT ALTERING ARRAY DIMENSIONS OR
0012  C     FORTRAN STATEMENTS. THE PROGRAM IS GENERAL AND MAY BE ALTERED TO
0013  C     HANDLE AS MANY EVENTS AS DESIRED.
0014  C*************************************************************************
0015  C*************************************************************************
0016        DIMENSION A(15,15),AMEAN(15),X(15,1000),D1(15,15),
0017       1AK(15,1000),Y(15,1000)
0018  C*************************************************************************
0019  C     INITIALIZE ARRAYS TO ZERO
0020  C*************************************************************************
0021        DO 4 I=1,15
0022        AMEAN(I)=0.0
0023        DO 4 J=1,15
0024        A(I,J)=0.0
0025        D1(I,J)=0.0
0026      4 CONTINUE
0027        DO 5 I=1,15
0028        DO 5 J=1,1000
0029        X(I,J)=0.0
0030        AK(I,J)=0.0
0031        Y(I,J)=0.0
0032      5 CONTINUE
0033  C*************************************************************************
0034  C     READ MEAN FOR EACH EVENT
0035  C*************************************************************************
0036        READ(5,2)(AMEAN(I),I=1,15)
0037      2 FORMAT(5F10.5)
0038  C*************************************************************************
0039  C     READ "A" MATRIX
0040  C*************************************************************************
0041        READ(5,3)((A(I,J),J=1,15),I=1,15)
0042      3 FORMAT(5F10.5)
0043  C*************************************************************************
0044  C     GENERATE RANDOM NORMAL DEVIATES
0045  C*************************************************************************
0046        IX=99999
0047        DO 6 J=1,1000
0048        DO 6 I=1,15
0049        CALL GAUSS (IX,1.0,0.0,X(I,J))
0050      6 CONTINUE
0051  C*************************************************************************
0052  C     CALCULATE CORRELATED OUTCOMES
0053  C*************************************************************************
0054        DO 7 K=1,1000
```

```
                                    80/80 LIST

     0000000001111111111222222222233333333334444444445555555556666666666777
     1234567890123456789012345678901234567890123456789012345678901234567890123
CARD
0055          DO 7 I=1,15
0056          DO 7 J=1,15
0057          D1(I,J)=A(I,J)*X(J,K)
0058          AK(I,K)=AK(I,K)+D1(I,J)
0059        7 CONTINUE
0060          DO 9 K=1,1000
0061          DO 9 I=1,15
0062          Y(I,K)=AMEAN(I)+AK(I,K)
0063          IF(Y(I,K)-0.0) 8,9,9
0064        8 Y(I,K)=0.0
0065        9 CONTINUE
0066    C*************************************************************************
0067    C       WRITE CORRELATED OUTCOMES
0068    C*************************************************************************
0069          WRITE (6,10)
0070       10 FORMAT(60X,'MATRIX OF YIELDS')
0071          DO 11 K=1,1000
0072          WRITE(6,12)(Y(J,K),J=1,15)
0073       11 CONTINUE
0074       12 FORMAT(15F8.4)
0075          STOP
0076          END
0077    C
0078    C*************************************************************************
0079    C       THE INPUT DATA FOR THE PROGRAM IS READ IN AS FOLLOWS:
0080    C*************************************************************************
0081    C
0082      12.51250   10.94370   39.14990   23.95000    0.0
0083       0.0         0.0         0.0         0.0      0.0
0084       0.0         0.0         0.0         0.0      0.0
0085       3.88109   -1.13808    1.25958   -0.16869    0.0
0086       0.0         0.0         0.0         0.0      0.0
0087       0.0         0.0         0.0         0.0      0.0
0088       0.0         5.40159     0.48455     1.51391  0.0
0089       0.0         0.0         0.0         0.0      0.0
0090       0.0         0.0         0.0         0.0      0.0
0091       0.0         0.0         3.85071     8.10785  0.0
0092       0.0         0.0         0.0         0.0      0.0
0093       0.0         0.0         0.0         0.0      0.0
0094       0.0         0.0         0.0         0.99599  0.0
0095       0.0         0.0         0.0         0.0      0.0
0096       0.0         0.0         0.0         0.0      0.0
0097       0.0         0.0         0.0         0.0      0.0
0098       0.0         0.0         0.0         0.0      0.0
0099       0.0         0.0         0.0         0.0      0.0
0100       0.0         0.0         0.0         0.0      0.0
0101       0.0         0.0         0.0         0.0      0.0
0102       0.0         0.0         0.0         0.0      0.0
0103       0.0         0.0         0.0         0.0      0.0
0104       0.0         0.0         0.0         0.0      0.0
0105       0.0         0.0         0.0         0.0      0.0
0106       0.0         0.0         0.0         0.0      0.0
0107       0.0         0.0         0.0         0.0      0.0
0108       0.0         0.0         0.0         0.0      0.0
```

```
                                                80/80 LIST

      0000000001111111111222222222233333333334444444444555555555566666666677
      1234567890123456789012345678901234567890123456789012345678901234567890123
CARD
0109      0.0         0.0         0.0         0.0         0.0
0110      0.0         0.0         0.0         0.0         0.0
0111      0.0         0.0         0.0         0.0         0.0
0112      0.0         0.0         0.0         0.0         0.0
0113      0.0         0.0         0.0         0.0         0.0
0114      0.0         0.0         0.0         0.0         0.0
0115      0.0         0.0         0.0         0.0         0.0
0116      0.0         0.0         0.0         0.0         0.0
0117      0.0         0.0         0.0         0.0         0.0
0118      0.0         0.0         0.0         0.0         0.0
0119      0.0         0.0         0.0         0.0         0.0
0120      0.0         0.0         0.0         0.0         0.0
0121      0.0         0.0         0.0         0.0         0.0
0122      0.0         0.0         0.0         0.0         0.0
0123      0.0         0.0         0.0         0.0         0.0
0124      0.0         0.0         0.0         0.0         0.0
0125      0.0         0.0         0.0         0.0         0.0
0126      0.0         0.0         0.0         0.0         0.0
0127      0.0         0.0         0.0         0.0         0.0
0128      0.0         0.0         0.0         0.0         0.0
0129      0.0         0.0         0.0         0.0         0.0
0130  C
0131  C*************************************************************************
0132  C      THE FIRST THREE CARDS IN THIS DATA SET CONTAIN MEANS FOR EACH OF
0133  C      THE 15 EVENTS. THE REMAINING CARDS MAY BE SUBDIVIDED INTO THREE-
0134  C      CARD SETS. EACH SET REPRESENTS ONE ROW OF COEFFICIENTS FOR THE "A"
0135  C      MATRIX. THESE COEFFICIENTS ARE CALCULATED EXTERNALLY UTILIZING THE
0136  C      PRECEDURES EXPLAINED IN THE BODY OF THIS PUBLICATION
0137  C*************************************************************************
```

APPENDIX C

Correlation Program for the Four-Event Case Including Statements Necessary to Verify Appropriate Correlations Among Events

This appendix discusses the correlation program used for the four-event case explored in the body of this bulletin. The program is divided into two parts. The first part is essentially the same as the correlation program discussed in Appendix B. The array and matrix names and definitions are the same for the first part, but the array and matrix dimensions reflect the fact that four instead of 15 events are being correlated.

The second part of the attached program simply verifies that the outcomes generated by the first part of the program are appropriately correlated. If the program user is confident that the first portion of the program is generating satisfactory values, the second part is not needed. However, to test the procedure used in this bulletin, statements were added to find the means, standard deviations, correlation coefficients and covariances for each of the events included.

The arrays used in the second part, or verification phase, of the program, their dimensions and uses are as follow:

B1(ACT): The B1 array contains the sum of each column of outcomes generated by the correlation program. The dimension of B1 is equal to the number of events being correlated (ACT), or B1(4) in the program attached to this appendix.

YM(ACT): The YM array contains the calculated mean of each column of event outcomes generated by the correlation program. Its dimension equals the number of events, or YM(4), in the attached program.

TN: This constant is set equal to the total number of outcomes to be generated by the correlation program. If 1000 outcomes are to be generated, TN=1000. TN is not an array and thus is undimensioned.

CORCOL(ACT,OBS): The matrix CORCOL contains each outcome after it has been corrected for the mean. Its dimensions must be the same as the outcome matrix (Y). For the attached program it is dimensioned as follows: CORCOL (4, 1000).

SQ(OBS): This array temporarily stores the square of each outcome after correction for the mean. Its dimension must equal the number of outcomes, or SQ(1000) in the attached program.

SSQ(ACT): This array contains the sum of squares for each event after correction for the mean. Its dimension must equal the number of events, or SSQ(4) in the attached program.

VAR(ACT): This array contains an unbiased estimate of the variance for each event. It must be dimensioned according to the number of events to be correlated, or VAR(4) in the attached program.

STD1(ACT): This array contains the standard deviation of the outcomes for each event. It must be dimensioned equal to the number of events, or STD1(4), in the attached program.

SUM1(OBS), SUM2(OBS), SUM3(OBS): These arrays contain the values resulting from cross-product array multiplication of corrected outcome columns to be used subsequently in calculating covariances and correlation coefficients. Each array must have a dimension equal to the number of outcomes generated or SUM1(1000), SUM2(1000) and SUM3(1000) in the attached program.

SUM4(NCC): This array contains the sums of each of the cross-product multiplications. There is one sum for each of the correlation coefficients calculated. NCC stands for number of correlation coefficients and is equal to six for a four-event program. Thus, SUM4 is dimensioned SUM4(6) in the attached program.

CALCOV(NCC): This array contains the calculated covariances among events of the correlation program. A covariance is calculated for each combination of events and thus six are calculated for the four-event program. It is dimensioned as CALCOV(6) in the attached program.

CRCOEF(NCC): This array contains the correlation coefficients calculated on the basis of outcomes generated within the correlation program. It is also dimensioned to accommodate the number of correlation coefficients to be calculated or as CRCOEF(6) in the attached program.

It should be emphasized that the latter portion of the attached program merely verifies that the correlation program is generating series that are appropriately correlated. If this portion of the program is to be incorporated into the correlation program attached to Appendix B, modifications must be made. To accomodate 15 activities, the array dimensions must be changed accordingly. In addition, the DO LOOP parameters must reflect 15 rather than four events. More importantly, 15 events necessitate calculation of 105 correlation coefficients and covariances. Arrays whose dimensions are based upon the number of correlation coefficients to be calculated must be enlarged. In addition, a number of Fortran statements must be added following Statement 28 in the attached program. One Fortran statement is required for each correlation coefficient to be calculated. If 105 correlation coefficients are required for 15 events, 99 additional statements would have to be added to the program to accommodate all correlation coefficients.

As with the 15-event program, the card input requirements for the four-event program are explained by comment cards following the Fort-

ran program. For the four-event case, only five input cards are required. The first card contains the mean of each event to be correlated and the next four cards contain coefficients of the A matrix.

```
                                    80/80 LIST

     00000000011111111112222222222333333333344444444445555555556666666666777
     12345678901234567890123456789012345678901234567890123456789012345678901234
CARD
0001  C*************************************************************************
0002  C*************************************************************************
0003  C                          CORRELATION PROGRAM
0004  C*************************************************************************
0005  C*************************************************************************
0006  C
0007  C*************************************************************************
0008  C*************************************************************************
0009  C      THIS PROGRAM IS DESIGNED TO GENERATE SERIES OF CORRELATED OUTCOMES
0010  C      FOR FOUR EVENTS.  IT IS DIVIDED INTO TWO PARTS.  THE FIRST PORTION
0011  C      GENERATES A SERIES OF 1000 YIELD OUTCOMES FOR WHEAT, GRAIN SORGHUM
0012  C      STEER AND COW-CALF EVENTS, CORRELATED AT THE APPROPRIATE LEVELS.
0013  C*************************************************************************
0014  C*************************************************************************
0015        DIMENSION B1(4),STD1(4),Y(4,1000),D1(4,4),AK(4,1000),AMEAN(4),
0016       1A(4,4),X(4,1000),YM(4),SSQ(4),VAR(4),SQ(1000),
0017       1CORCOL(4,1000),CALCOV(6),CRCOEF(6),SUM1(1000),SUM2(1000),
0018       1SUM3(1000),SUM4(6)
0019  C*************************************************************************
0020  C      INITIALIZE ARRAYS TO ZERO
0021  C*************************************************************************
0022        DO 4 I=1,4
0023        YM(I)=0.
0024        B1(I)=0.
0025        SSQ(I)=0.
0026        VAR(I)=0.
0027        STD1(I)=0.
0028        AMEAN(I)=0.
0029      4 CONTINUE
0030        DO 97 I=1,1000
0031        SQ(I)=0.
0032        SUM1(I)=0.
0033        SUM2(I)=0.
0034        SUM3(I)=0.
0035     97 CONTINUE
0036        DO 98 I=1,4
0037        DO 98 J=1,4
0038        D1(I,J)=0.
0039        A(I,J)=0.
0040     98 CONTINUE
0041        DO 99 I=1,4
0042        DO 99 K=1,1000
0043        AK(I,K)=0.
0044        Y(I,K)=0.
0045        CORCOL(I,K)=0.
0046        X(I,K)=0.
0047     99 CONTINUE
0048        DO 100 I=1,6
0049        CALCOV(I)=0.
0050        CRCOEF(I)=0.
0051        SUM4(I)=0.
0052    100 CONTINUE
0053  C*************************************************************************
0054  C      READ MEAN FOR EACH EVENT
```

```
      00000000011111111112222222222333333333344444444445555555556666666666777
      12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012
CARD
0055  C*************************************************************************
0056        READ(5,2)(AMEAN(I),I=1,4)
0057      2 FORMAT(4F10.5)
0058  C*************************************************************************
0059  C     READ "A" MATRIX
0060  C*************************************************************************
0061        READ(5,3)((A(I,J),J=1,4),I=1,4)
0062      3 FORMAT(4F10.5)
0063  C*************************************************************************
0064  C     GENERATE RANDOM NORMAL DEVIATES
0065  C*************************************************************************
0066        IX=99999
0067        DO 7 J=1,1000
0068        DO 7 I=1,4
0069        CALL GAUSS(IX,1.0,0.0,X(I,J))
0070      7 CONTINUE
0071  C*************************************************************************
0072  C     CALCULATE CORRELATED OUTCOMES
0073  C*************************************************************************
0074        DO 9 K=1,1000
0075        DO 9 I=1,4
0076        DO 9 J=1,4
0077        D1(I,J)=A(I,J)*X(J,K)
0078        AK(I,K)=AK(I,K)+D1(I,J)
0079      9 CONTINUE
0080        DO 10 K=1,1000
0081        DO 10 I=1,4
0082        Y(I,K)=AMEAN(I)+AK(I,K)
0083        IF(Y(I,K)-0.0)30,10,10
0084     30 Y(I,K)=0.0
0085     10 CONTINUE
0086  C
0087  C*************************************************************************
0088  C*************************************************************************
0089  C     THE SECOND PART OF THE PROGRAM VERIFIES CALCULATION OF CORRELATED
0090  C     OUTCOMES BY CALCULATING THE MEAN AND VARIANCE FOR EACH EVENT AND
0091  C     COVARIANCES AND CORRELATIONS BETWEEN ALTERNATIVE EVENTS.
0092  C*************************************************************************
0093  C*************************************************************************
0094  C
0095  C*************************************************************************
0096  C     CALCULATE MEANS, VARIANCES AND STANDARD DEVIATIONS
0097  C*************************************************************************
0098        TN=1000.0
0099     13 DO 14 J=1,1000
0100        DO 14 I=1,4
0101        B1(I)=B1(I)+Y(I,J)
0102     14 CONTINUE
0103        DO 15 I=1,4
0104        YM(I)=B1(I)/TN
0105     15 CONTINUE
0106        DO 16 I=1,4
0107        DO 16 J=1,1000
0108        CORCOL(I,J)=Y(I,J)-YM(I)
```

```
      00000000011111111112222222223333333334444444445555555556666666667777
      12345678901234567890123456789012345678901234567890123456789012345678901234
CARD
0109        SQ(J)=CORCOL(I,J)*CORCOL(I,J)
0110        SSQ(I)=SSQ(I)+SQ(J)
0111     16 CONTINUE
0112        DO 17 I=1,4
0113        VAR(I)=SSQ(I)/(TN-1.0)
0114        STD1(I)=VAR(I)**.5
0115     17 CONTINUE
0116  C*************************************************************************
0117  C      CALCULATE COVARIANCES AND CORRELATION COEFFICIENTS
0118  C*************************************************************************
0119        DO 23 I=1,3
0120        DO 23 J=1,1000
0121        SUM1(J)=CORCOL(1,J)*CORCOL(I+1,J)
0122        SUM4(I)=SUM4(I)+SUM1(J)
0123     23 CONTINUE
0124        DO 24 I=2,3
0125        DO 24 J=1,1000
0126        SUM2(J)=CORCOL(2,J)*CORCOL(I+1,J)
0127        SUM4(I+2)=SUM4(I+2)+SUM2(J)
0128     24 CONTINUE
0129        DO 25 I=1,1000
0130        SUM3(I)=CORCOL(3,I)*CORCOL(4,I)
0131        SUM4(6)=SUM4(6)+SUM3(I)
0132     25 CONTINUE
0133        DO 28 I=1,6
0134        CALCOV(I)=SUM4(I)/(TN-1.0)
0135     28 CONTINUE
0136        CRCOEF(1)=CALCOV(1)/(STD1(1)*STD1(2))
0137        CRCOEF(2)=CALCOV(2)/(STD1(1)*STD1(3))
0138        CRCOEF(3)=CALCOV(3)/(STD1(1)*STD1(4))
0139        CRCOEF(4)=CALCOV(4)/(STD1(2)*STD1(3))
0140        CRCOEF(5)=CALCOV(5)/(STD1(2)*STD1(4))
0141        CRCOEF(6)=CALCOV(6)/(STD1(3)*STD1(4))
0142  C*************************************************************************
0143  C      WRITE CORRELATED OUTCOMES
0144  C*************************************************************************
0145        WRITE(6,12)
0146     12 FORMAT(14X,'MATRIX OF YIELDS')
0147        DO 21 K=1,1000
0148        WRITE(6,20)(Y(J,K),J=1,4)
0149     21 CONTINUE
0150     20 FORMAT(4F12.5)
0151  C*************************************************************************
0152  C      WRITE CALCULATED MEANS, STANDARD DEVIATIONS, COVARIANCES AND
0153  C      CORRELATION COEFFICIENTS
0154  C*************************************************************************
0155        WRITE(6,19)(YM(I),I=1,4)
0156     19 FORMAT(12X,'MEANS OF YIELD TOTALS',/4F12.5)
0157        WRITE(6,18)(STD1(I),I=1,4)
0158     18 FORMAT(8X,'STANDARD DEVIATION OF YIELDS',/4F12.5)
0159        WRITE(6,27)
0160     27 FORMAT(21X,'CALCULATED COVARIANCES')
0161        WRITE(6,50)(CALCOV(I),I=1,6)
0162        WRITE(6,26)
```

```
                                        80/80 LIST

      00000000011111111112222222222333333333344444444455555555556666666667777
      12345678901234567890123456789012345678901234567890123456789012345678901234
CARD
0163      26 FORMAT(22X,'CORRELATION COEFFICIENT')
0164         WRITE(6,50)(CRCOEF(I),I=1,6)
0165      50 FORMAT(3F12.5,/,12X,2F12.5,/,24X,F12.5)
0166         STOP
0167         END
0168  C
0169  C*************************************************************************
0170  C      THE INPUT DATA FOR THE PROGRAM IS READ IN AS FOLLOWS:
0171  C*************************************************************************
0172  C
0173     12.51250  10.94370  39.14990  23.95000
0174      3.88109  -1.13808   1.25958  -0.16869
0175      0.00000   5.40159   0.48455   1.51391
0176      0.00000   0.00000   3.85071   8.10785
0177      0.00000   0.00000   0.00000   0.99599
0178  C
0179  C*************************************************************************
0180  C*************************************************************************
0181  C      THE FIRST CARD IN THE DATA SET CONTAINS MEANS FOR EACH EVENT. THE
0182  C      SECOND THROUGH FIFTH CARDS IN THE DATA SET CONTAIN VALUES FOR THE
0183  C      "A" MATRIX. THESE VALUES ARE CALCULATED EXTERNALLY UTILIZING THE
0184  C      PROCEDURES EXPLAINED IN THE BODY OF THIS PUBLICATION.
0185  C*************************************************************************
```

APPENDIX D

Tests of Significance for the Parameters of the Generated Distributions

This appendix discusses tests applied to determine if the mean and variance of the generated outcomes for each event as well as the correlation of outcomes for each pair of events differ significantly from the original values specified. For purposes of these tests, the original values specified in Tables 1 and 2 for the means, variances and correlations may be interpreted as "true population parameters". The means, variances and correlations computed from the generated series are considered "sample estimates" of the population parameters. The purpose of the tests is to determine if the observed differences between the parameters read into the procedure and those computed from the generated series (sample) can be expected to occur due to chance. Tests to compare the means, variances and correlation coefficients are discussed in order.

In testing the difference between original and generated means, the null hypothesis H_0 is that the true difference is zero. The alternative hypothesis H_1 is that the true difference is not equal to zero. The test statistic is given in (53).

$$(53) \quad T = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

where $\bar{x}$ = the mean of the sample of outcomes,
μ = the population mean,
$\sigma_{\bar{x}} = \sigma/\sqrt{N}$ where σ is the population standard deviation and N is the sample size.

The value for the population parameters μ and σ, representing the population mean and standard deviation are presented in Table 1. The sample of 1000 outcomes for each event may be considered a random sample drawn from a normal population for which mean and standard deviations have also been computed. Thus to test H_0 of no difference between means for each event, merely substitute the appropriate values in (53). For wheat

$$(54) \quad T_w = \frac{12.50207 - 12.51250}{4.23947/\sqrt{1000}} = \frac{-0.01043}{4.23947/31.62278}$$

$$= \frac{-0.01043}{0.13406} = -0.07780$$

The quantity T is normalized by expressing the difference between means in units of standard deviation of the distribution of means, and may be interpreted as a normal deviate [6, p. 213-214]. The computed value of T_w (—0.07780 for wheat) is compared to the 95% tabular value in a table of Normal Areas and Ordinates. The computed T value would have to exceed 1.96 for a significant difference to exist between population and sample means at the 95% level.[4]

Computation of T values for grain sorghum, steers and cow-calf events are reflected in (55), (56) and (57), respectively.

$$(55)\ T_g = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{10.92747 - 10.94370}{5.63062/31.62278} = \frac{-0.01623}{0.17806} = -0.09115$$

$$(56)\ T_s = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{38.76559 - 39.14490}{8.97581/31.62278} = \frac{-0.37931}{0.28384} = -1.33635$$

$$(57)\ T_c = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{23.89412 - 23.95000}{0.99599/31.62278} = \frac{-0.05588}{0.03150} = -1.77397$$

All computed T values are less than 1.96. Thus there is no basis for rejecting the null hypothesis at the 95% level of significance.

For testing the difference between original and generated variances, the null hypothesis is that the variance for an event is equal to a specified value, or $H_o{:}\sigma^2 = \sigma_o^2$. The alternative hypothesis is that the variance is not equal to the specified value, $H_1{:}\sigma^2 \neq \sigma_o^2$. The specified value may be represented by s^2, an unbiased estimate of σ^2.

A Chi-Square (X^2) test statistic can be used to test the null hypothesis. Hoel [5, p. 268] presents a theorem which states that if x is normally distributed with variance σ^2, and s^2 is the sample variance based on a random sample of size n, then ns^2/σ^2 ahs a X^2 distribution with n-1 degrees of freedom. Since the computer program in Appendix C calculates unbiased estimates of sample variance, the appropriate test statistic is $Z = (n\text{-}1)s^2/\sigma^2$ which is distributed as X^2 with n-1 degrees of freedom.

Such a statistic, denoted as Z, was computed for each of the four events and is reported for wheat, grain sorghum, steers and cow-calf events, respectively, in (58), (59), (60) and (61).

[4]The reader may want to question if the 95% level is appropriate for this test. Selecting a lower significance level would result in using a smaller tabular T value. This would increase the probability of accepting H_1 when H_0 is true (Type I error), but decrease the probability of accepting H_0 when H_1 is true (Type II error). A similar question can be raised concerning the statistical tests for significance between "actual" and "sample" values for the variance and correlation coefficients. This indicates researchers making these tests must decide which type of error is more important and select the level of significance accordingly.

$$(58)\quad Z_w = \frac{999 \times 17.85054}{17.97311} = 992.18718$$

$$(59)\quad Z_g = \frac{999 \times 32.03345}{31.70388} = 1{,}009.38486$$

$$(60)\quad Z_s = \frac{999 \times 82.75268}{80.56517} = 1{,}026.12490$$

$$(61)\quad Z_c = \frac{999 \times 0.99138}{0.99199} = 999.39275$$

The appropriate test is conducted by comparing computed values of Z for each event with the tabular values of the X^2 distribution at the 95% level of significance for 999 degrees of freedom. The tabular value is approximately 1070. Since all calculated Z_i are less than the tabular value, there is no statistical basis for rejecting the null hypothesis of equal variances.

A possible difficulty in employing this test is that tabular values for the X^2 distribution with up to 1000 degrees of freedom are difficult to locate. Most standard textbooks have tabular values for no more than 30 degrees of freedom. A simple transformation allows use of an alternative distribution known as "X^2 divided by degrees of freedom" or X^2/d.f., for which tabular values for 1000 or more degrees of freedom are readily available [2, p. 295]. The X^2/d.f. distribution has a test statistic $Z = s^2/\sigma^2$ distributed as X^2 with n-1 degrees of freedom.

Test statistics $Z_i = s^2/\sigma^2$ for wheat, grain sorghum, steer and cow-calf events are computed in (62), (63), (64) and (65).

$$(62)\quad Z_w = \frac{17.85054}{17.97311} = 0.99318$$

$$(63)\quad Z_g = \frac{32.03345}{31.70388} = 1.01040$$

$$(64)\quad Z_s = \frac{82.75268}{80.56517} = 1.02715$$

$$(65)\quad X_c = \frac{0.00183}{0.099199} = 0.99984$$

These computed values must be compared with a tabular X^2/d.f. value at the 95% level of significance for 999 degrees of freedom. The

closest tabular value located was 1.0747 for 1000 degrees of freedom. Since no computed value exceeds 1.0747, the null hypothesis of no difference between the original variance and the sample estimate is not rejected.

Test of hypotheses were not conducted for the original versus generated correlation coefficients. Rather confidence intervals were calculated for each population correlation coefficient, ρ_{ij}. A 95% confidence interval for any ρ is specified by computing the sample estimate of the correlation coefficient, r, plus and minus 1.96 times the standard error of r. The standard error of r may be calculated from the general expression

$$(66) \quad \sigma_r = \frac{1 - \rho^2}{\sqrt{N\text{-}1}}$$

where σ_r = the standard error of r,
$_r$ = the sample estimate of ρ,
N = the numbers of observations.

There are two restrictions on the use of (66). First, it calls for knowledge of ρ, which is not generally known. In this example, ρ was calculated from the original data and is known. The second restriction is that the sampling distribution of r may depart widely from the normal type, particularly for small and moderate size sampling distributions of r [6, pp. 298-299]. However, with a sample size of 1000 the sampling distribution of r should closely approximate the normal distribution.

Equation (66) was used to compute the standard error of each r and 95% confidence limits were then computed for each ρ. For wheat vs. grain sorghum, the standard error of r is given by

$$(67) \quad \sigma_{r_{wg}} = \frac{1 - (-0.24266)^2}{31.60696} = 0.02978$$

The sample estimate of the correlation coefficient, plus and minus 1.96 times the standard error of r specifies a 95% confidence interval for ρ.

$$(68) \quad -0.23702 \pm (1.96 \times 0.02978) = -0.23702 \pm 0.05837$$

The upper limit is —0.17865 and the lower limit is —0.29539. The population value of —0.24266 falls near the center of the interval.

Similar computations were made for wheat vs. steers, wheat vs. cow-calf, grain sorghum vs. steers, grain sorghum vs. cow-calf and steers vs. cow-calf and are presented in (69), (70), (71), (72) and (73), respectively

For wheat vs. steers, the standard error of r_{ws} is

$$(69) \quad \sigma_{r_{ws}} = \frac{1 - (0.09152)^2}{31.60696} = 0.03137$$

and confidence limits are 0.09651 ± (1.96 x 0.03137) = 0.09651 ± 0.06149.

The upper and lower limits for ρ_{ws} are 0.15800 and 0.03502, respectively. The population value of 0.09152 is near the center of this interval.

For wheat vs. cow-calf, the standard error of r_{wc} is

$$(70) \quad \sigma_{r_{wc}} = \frac{1 - (-0.03979)^2}{31.60696} = 0.03159$$

and confidence limits are —0.02506 ± (1.96 x 0.03159) = 0.02506 ± 0.06192.

The upper and lower limits for ρ_{wc} are 0.03686 and —0.08698, respectively. The population value of —0.03979 is near the center of this interval.

For grain sorghum vs. steers, the standard error of r_{gs} is

$$(71) \quad \sigma_{r_{gs}} = \frac{1 - (-0.27979)^2}{31.60696} = 0.02916$$

and confidence limits are —0.26040 ± (1.96 x 0.02916) = 0.26040 ± 0.05715.

The upper and lower limits for ρ_{gs} are 0.31755 and 0.20325, respectively. The population correlation coefficient, 0.27979 falls within this interval.

For grain sorghum vs. cow-calf, the standard error of r_{gc} is

$$(72) \quad \sigma_{r_{gc}} = \frac{1 - (0.30094)^2}{31.60696} = 0.02877$$

and confidence limits are 0.26094 ± (1.96 x 0.02877) = 0.26094 ± 0.05639.

The upper and lower limits for ρ_{gc} are 0.31733 and 0.20455, respectively. The population value of 0.30094 falls within the 95% confidence interval.

For steers vs. cow-calf, the standard error of r_{sc} is

$$(73) \quad \sigma_{r_{sc}} = \frac{1 - (0.90330)^2}{31.60696} = 0.00582$$

and confidence limits are 0.90023 ± (1.96 x 0.00582) = 0.90023 ± 0.01141.

The upper and lower limits for ρ_{sc} are 0.91164 and 0.88882, respectively. The population value of 0.90330 is within this interval.

0871/1.6M