Proceedings of the 1990 Winter Simulation Conference
Osman Balci, Randall P. Sadowski, Richard E. Nance (eds.)

POWER COMPARISONS FOR THE MULTIVARIATE BATCH-MEANS METHOD

John M. Charnes

Department of Management Science
417 Jenkins Building
University of Miami
Coral Gables, Forida 33124-8237

ABSTRACT

The multivariate batch-means (MBM) method of analyzing simulation output has the potential to be a widely-used tool by practitioners. This paper presents a power comparison of selected statistics that can be used for choosing the batches, which is one of the key elements in the successful use of the MBM method. Explicit procedures are given for five batch-selection criteria, and some guidelines for the proper use of the MBM method are stated.

1. INTRODUCTION

There has been some interest recently among researchers in the problem of making statistical inferences simultaneously on more than one output measure of interest in simulation modeling (Chen and Seila 1987; Yang and Nelson 1988; Charnes and Kelton 1988). One technique that has received attention is the multivariate batch-means (MBM) method, perhaps because it is a generalization of the widely-used univariate batch-means method and thus has the potential to be the most widely-used multivariate technique. In light of this potential, the intent of this paper is to examine more closely the procedure used to select the batches, which is a fundamental element of the MBM method.

In making inferences on multivariate processes one must recognize the autocorrelation that also exists with univariate processes, and, in addition, the cross-correlation that may exist among the individual univariate processes comprising the multivariate process. Data are said to be *cross-correlated* if they are observations on a vector-valued stochastic process $\{\mathbf{X}_t\}$ whose specification includes not only the serial dependence of each component series $\{X_{ti}\}$ but also the interdependence between different component series $\{X_{ti}\}$ and $\{X_{tj}\}$. When such cross-correlation exists, it may be useful to exploit the information somehow.

The MBM method attempts to circumvent the autocorrelation problem without losing the information on cross-correlation by grouping the data into (nearly) uncorrelated batches. The purpose of this paper is to compare selected statistics available from multivariate-statistical theory with respect to their performance in choosing batches. §2 describes the MBM method. §3 presents the statistics that were selected for investigation in this study. §4 gives the results of the Monte Carlo experiment that was used to compare the selected statistics, and in §5 some conclusions are drawn and recommendations made.

2. THE MBM METHOD

The MBM method of constructing confidence regions is a generalization of the univariate batch-means method of constructing confidence intervals. Chen and Seila (1987) discuss the application of the multivariate-batch-means method to steady-state, synchronous simulation output. Consider a stationary process that produces a sequence of d-dimensional vector-valued observations $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$. The analyst wishes to estimate the mean vector $\mu = E(\mathbf{X}_t)$ with a joint confidence region. The MBM method calls for dividing the sequence of output vectors into m batches of k (vector) observations each (where $n = mk$) and computing the batch-mean vectors as

$$\mathbf{Y}_i = \frac{1}{k}\sum_{j=1}^{k} \mathbf{X}_{(i-1)k+j} \qquad \text{for} \quad i = 1, \ldots, m.$$

The m vectors of batch means are then treated as if they were uncorrelated vectors of observations and standard multivariate-statistical techniques are used to form a confidence region on the mean vector, μ, as follows. Let $\mathbf{S}$ denote the sample variance-covariance matrix for the $\mathbf{Y}_i$'s:

$$\mathbf{S} = \frac{1}{m-1}\sum_{i=1}^{m}(\mathbf{Y}_i - \hat{\mu})(\mathbf{Y}_i - \hat{\mu})'$$

where $'$ denotes matrix transposition and the point estimator of μ is the d-dimensional column vector

$$\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{Y}_i.$$

An approximate $100(1-\alpha)\%$ confidence region for μ is then given by

$$\left\{\theta \in \mathbb{R}^d : m(\hat{\mu} - \theta)'\mathbf{S}^{-1}(\hat{\mu} - \theta) \le \frac{d(m-1)}{m-d}F_{\alpha;d,m-d}\right\}$$

where $F_{\alpha;d,m-d}$ is the $100(1-\alpha)\%$ quantile of the F distribution with d numerator and $m-d$ denominator degrees of freedom.

3. SELECTING A STATISTIC FOR GROUPING

An important step in using the MBM method is the determination of the number of vector observations per batch, k (or, equivalently, the number of batches, m). The usual method of making this determination is to assume that the batch-means process can be sufficiently approximated by the first-order, vector-autoregressive (VAR(1)) model

$$\mathbf{Y}_i = \Phi\mathbf{Y}_{i-1} + \epsilon_i \qquad \text{for} \quad i = 1, \ldots, m,$$

where Φ is a $(d \times d)$ matrix of autoregression coefficients and the ϵ_i are $(d \times 1)$ independent and identically distributed vectors of random error drawn from the multivariate normal distribution. Then k is chosen such that H_0: $\Phi = \mathbf{0}$ is not rejected. Implicit in the use of this model is the assumption that if the first-order serial correlation is zero, then the higher-order serial correlations will be zero also.

Anderson (1978) suggests that H_0 can be tested with one of the criteria given in Anderson (1984) for testing the general linear hypothesis. The tests considered in this paper are the Lawley-Hotelling trace criterion, the Bartlett-Nanda-Pillai criterion, and three slightly different forms of the Wilks likelihood-ratio criterion.

3.1 Test Procedures

All of the test procedures studied here use the following $d \times d$ matrices in the calculation of the test statistics

$$\begin{aligned}
\mathbf{S}^*(0) &= \sum_{i=1}^{m-1}(\mathbf{Y}_i - \hat{\mu})(\mathbf{Y}_i - \hat{\mu})' \\
\mathbf{S}(0) &= \sum_{i=2}^{m}(\mathbf{Y}_i - \hat{\mu})(\mathbf{Y}_i - \hat{\mu})' \\
\mathbf{S}(1) &= \sum_{i=2}^{m}(\mathbf{Y}_i - \hat{\mu})(\mathbf{Y}_{i-1} - \hat{\mu})' \\
\mathbf{H} &= \mathbf{S}(1)\mathbf{S}^*(0)^{-1}\mathbf{S}(1)' \\
\mathbf{G} &= \mathbf{S}(0) - \mathbf{S}(1)\mathbf{S}^*(0)^{-1}\mathbf{S}(1)',
\end{aligned}$$

where the number of degrees of freedom of $\mathbf{G}$ is $df = m - d - 1$. The procedures are

1. The Lawley-Hotelling trace procedure, where significance points for the statistic
$$L = \frac{m-d-1}{d} \cdot \mathrm{tr}\left\{\mathbf{H}\mathbf{G}^{-1}\right\}$$
are tabulated in Anderson (1984, pp. 616–629). ($\mathrm{tr}\{\mathbf{Z}\} = \sum_{i=1}^{d} Z_{ii}$ denotes the trace of the $(d \times d)$ matrix $\mathbf{Z}$.)

2. The Wilks likelihood-ratio procedure, where values of the statistic
$$W = -(m - d - 1 - \frac{1}{2})\log\left\{\frac{|\mathbf{G}|}{|\mathbf{G}+\mathbf{H}|}\right\}$$
are compared to critical values of the $\chi^2(d^2)$ distribution adjusted with the values tabulated in Anderson (1984, pp. 609–615) for finite samples.

3. The Bartlett-Nanda-Pillai trace procedure, where significance points for the statistic
$$P = \frac{m-1}{d} \cdot \mathrm{tr}\left\{\mathbf{H}(\mathbf{G}+\mathbf{H})^{-1}\right\}$$
are tabulated in Anderson (1984, pp. 630–633).

4. The F-approximation to the Wilks likelihood-ratio procedure suggested by Rao (1951)
$$R = \frac{ks - r}{d^2} \cdot \frac{1 - U^{1/s}}{U^{1/s}}$$
where
$$U = \frac{|\mathbf{G}|}{|\mathbf{G}+\mathbf{H}|}.$$
R has approximately the F-distribution with d^2 and $ks - r$ degrees of freedom, where
$$s = \sqrt{\frac{d^4 - 4}{2d^2 - 5}}, \qquad r = \frac{d^2}{2} - 1,$$
and k is $m - \frac{1}{2}$

5. The statistic used in the Wilks likelihood-ratio procedure can be shown to have an asymptotic χ^2 distribution, and was used for the fifth procedure in the comparison. This statistic,
$$A = -(m - d - 1 - \frac{1}{2})\log\left\{\frac{|\mathbf{G}|}{|\mathbf{G}+\mathbf{H}|}\right\},$$
is computed identically to W, but is compared to critical values obtained from the $\chi^2(d^2)$ distribution.

4. EXPERIMENTAL DESIGN

The statistics were compared in a designed experiment with factors: α, the Type I error rate; d, the dimension of the output vectors; df, the number of degrees of freedom of $\mathbf{G}$; and the cross correlation among elements of the output vectors. The factors and their levels are listed in Table 1.

Table 1. Experimental Design

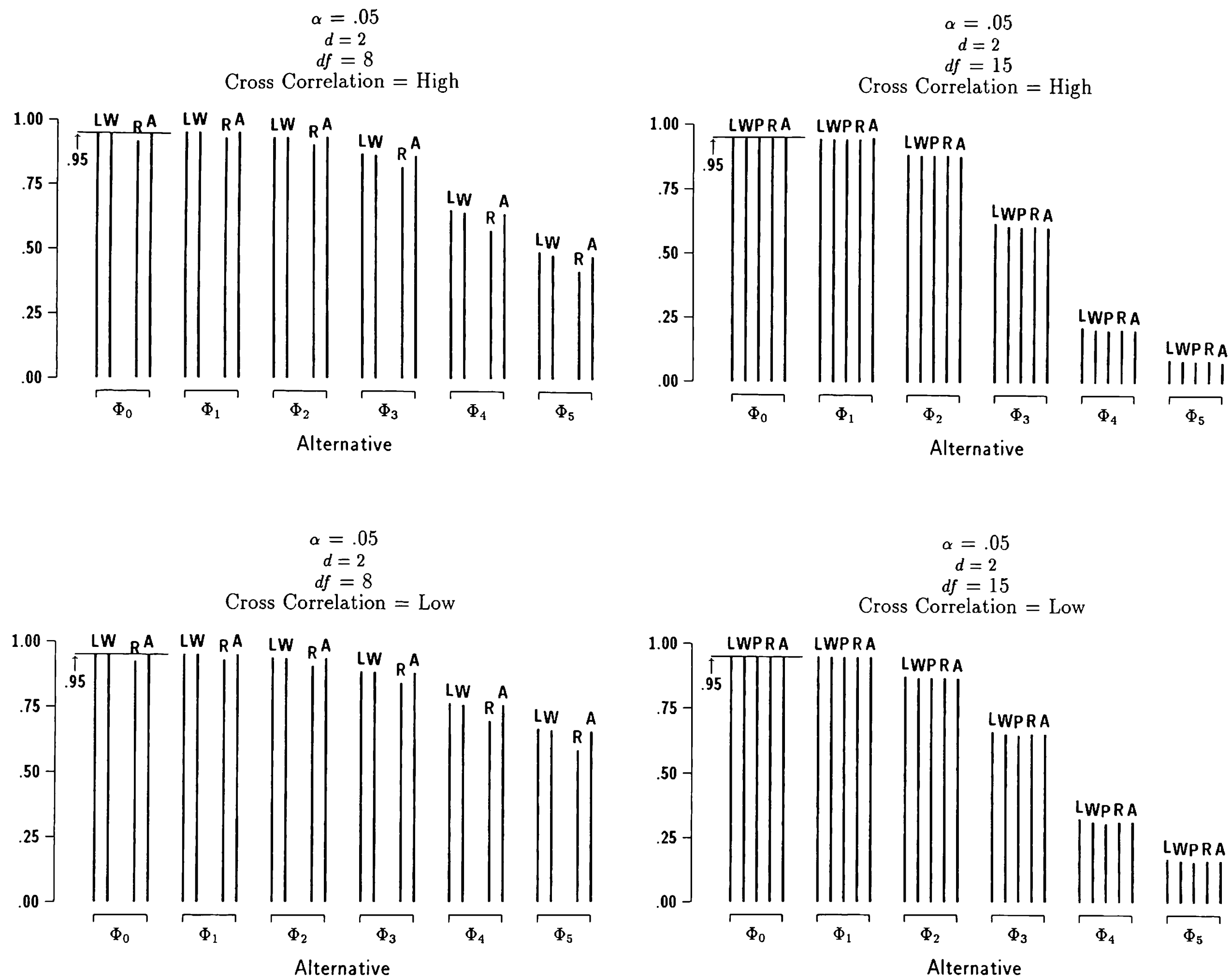
Factor	Levels
α	.01, .05
d	2, 4
df	8, 15, 30, 60
Cross Correlation	High, Low

For each combination of factor levels, observations were generated on the models

$$\mathbf{Y}_i = \Phi_j \mathbf{Y}_{i-1} + \epsilon_i \tag{1}$$

for $i = 1, \ldots, m(= df + d + 1)$ and $j = 0, 1, \ldots, 5$. The Φ_js are shown in Figure 1. Φ_0 is the zero matrix and thus generated observations for which H_0 was true. $\Phi_1, \ldots, \Phi_5$ generated observations for which H_0 was false. The ϵ_i were generated from a multivariate normal distribution with mean vector $\mathbf{0}$ and a variance-covariance matrix, Σ, that was varied to induce either high or low cross correlation. The different values for Σ that were used in the experiment are also shown in Figure 1. Random numbers were generated with Marse and Roberts' (1983) portable generator, and normal deviates were obtained with the "polar" method.

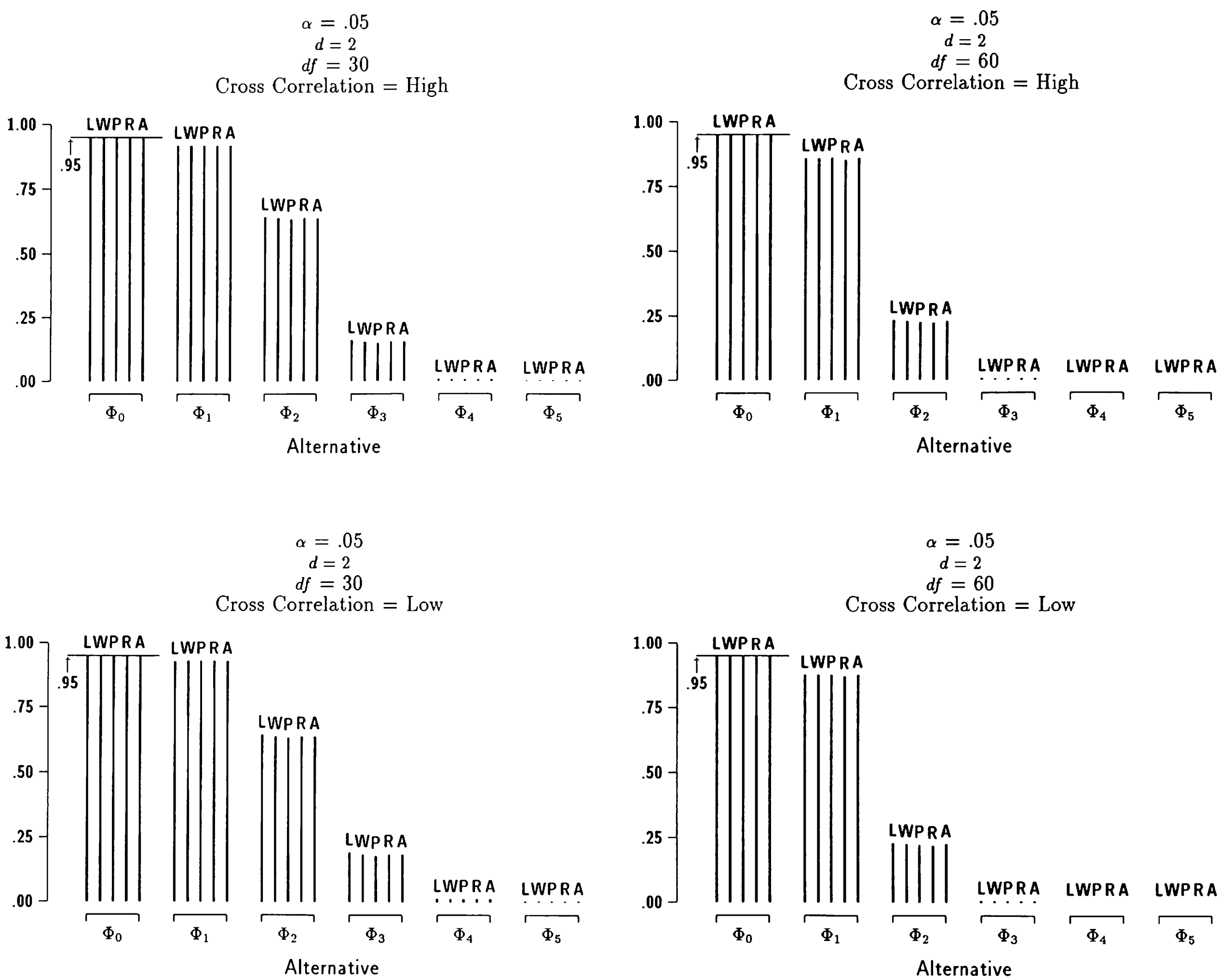
The response measured for each test procedure was the proportion of times (out of 10,000 runs) that the test statistic fell below the critical value at the α level. Thus the response for Φ_0 was the proportion of times that a test procedure indicated correctly that H_0 was true (which should be very close



Matrices used for different alternatives in (1).
(For $d = 2$, the upper left (2×2) sub-matrices were used.)

$$\Phi_0 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \Phi_1 = \begin{pmatrix} .10 & .01 & .01 & .01 \\ 0 & .15 & .01 & .01 \\ 0 & 0 & .20 & .01 \\ 0 & 0 & 0 & .25 \end{pmatrix}$$

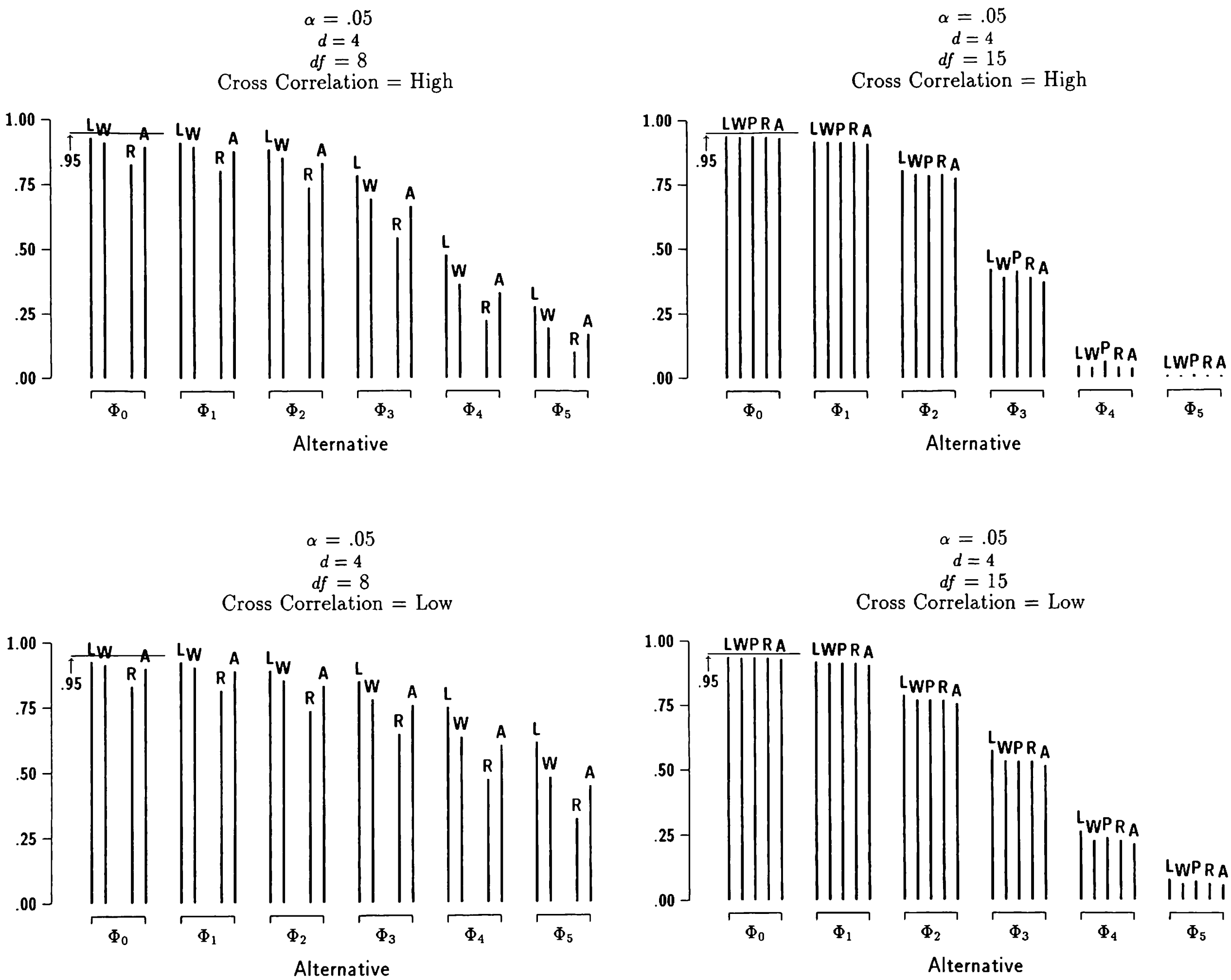
Figure 1. Power Comparison Results



Matrices used for different alternatives in (1).
(For $d = 2$, the upper left (2×2) sub-matrices were used.)

$$\Phi_2 = \begin{pmatrix} .30 & .03 & .03 & .03 \\ 0 & .35 & .03 & .03 \\ 0 & 0 & .40 & .03 \\ 0 & 0 & 0 & .45 \end{pmatrix} \qquad \Phi_3 = \begin{pmatrix} .50 & .05 & .05 & .05 \\ 0 & .50 & .05 & .05 \\ 0 & 0 & .50 & .05 \\ 0 & 0 & 0 & .50 \end{pmatrix}$$

Figure 1. (continued) Power Comparison Results

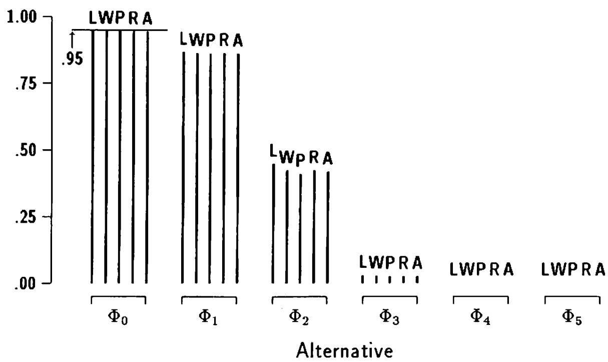


Matrices used for different alternatives in (1).
(For $d = 2$, the upper left (2×2) sub-matrices were used.)

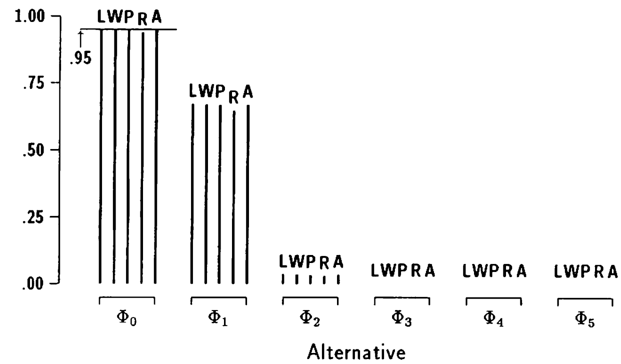
$$\Phi_4 = \begin{pmatrix} .70 & .07 & .07 & .07 \\ 0 & .65 & .07 & .07 \\ 0 & 0 & .60 & .07 \\ 0 & 0 & 0 & .55 \end{pmatrix} \qquad \Phi_5 = \begin{pmatrix} .80 & .09 & .09 & .09 \\ 0 & .75 & .09 & .09 \\ 0 & 0 & .70 & .09 \\ 0 & 0 & 0 & .65 \end{pmatrix}$$

Figure 1. (continued) Power Comparison Results

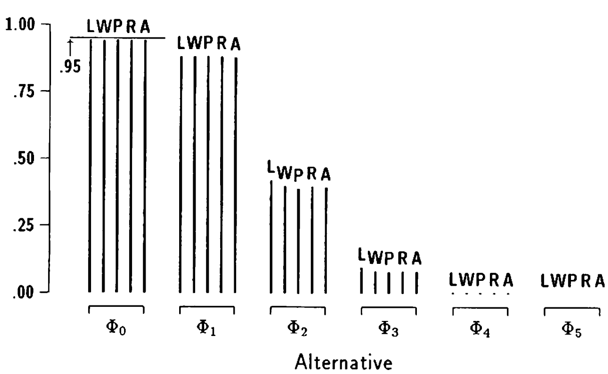
$\alpha = .05$
$d = 4$
$df = 30$
Cross Correlation = High



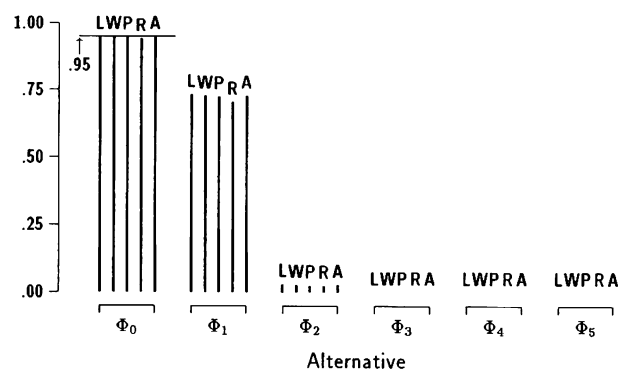
$\alpha = .05$
$d = 4$
$df = 60$
Cross Correlation = High



$\alpha = .05$
$d = 4$
$df = 30$
Cross Correlation = Low



$\alpha = .05$
$d = 4$
$df = 60$
Cross Correlation = Low



Matrices used to induce High and Low Cross Correlation.
(For $d = 2$, the upper left (2×2) sub-matrices were used.)

Low

$$\Sigma = \begin{pmatrix} 1. & 0 & 0 & 0 \\ 0 & 2. & 0 & 0 \\ 0 & 0 & 3. & 0 \\ 0 & 0 & 0 & 4. \end{pmatrix}$$

High

$$\Sigma = \begin{pmatrix} 1.00 & 1.27 & 1.39 & 1.40 \\ 1.27 & 2.00 & 1.71 & 1.70 \\ 1.39 & 1.71 & 3.00 & 2.08 \\ 1.40 & 1.70 & 2.08 & 4.00 \end{pmatrix}$$

Figure 1. (continued) Power Comparison Results

to $1-\alpha$ for a good procedure). The responses for $\Phi_1, \ldots, \Phi_5$ were the proportion of times that the test procedures indicated incorrectly that H_0 was true, and thus are point estimates of $\beta = P(\text{Type II Error})$, or $1 -$ Power, for each procedure. Note that the point estimates are quite precise, for if the actual power level of a procedure is .5, the standard error of the estimate is .005.

5. RESULTS

The results of the experiment for $\alpha = .05$ are shown in Figure 1. The results for $\alpha = .01$ are similar and are not included here to conserve space, but are available from the author upon request. At the design points for which $df = 8$, responses for the Bartlett-Nanda-Pillai statistic, P, are not plotted because tabulated significance values of that statistic for low values of df were not available.

Figure 1 shows that no test procedure appears to be uniformly more powerful than the others for the factors and levels tested here. All of the test statistics perform quite well at Φ_0, where nominal coverage is nearly obtained by all the statistics. However, there is some degradation of performance of the statistics, especially R, for those design points where $df = 8$ and $d = 4$.

The statistics seem to be slightly more powerful when cross correlation is high rather than low, although the differences may not be significant. None of the statistics tested here are very powerful at low df, a fact that has implications for the minimum number of observations necessary for successful use of the MBM method.

6. CONCLUSION

Because action is taken upon *not* rejecting H_0 in the MBM method, the probability of making a Type II error in testing H_0 is more important than the probability of a Type I error.

This comparison shows that one needs many more than 8 degrees of freedom for $\mathbf{G}$ to have a reasonable chance of not making a Type II error with the alternative hypotheses that were considered. Furthermore, even with $df = 60$, one will not often detect small departures of Φ from $\mathbf{0}$, such as Φ_1.

The drop in power of the statistics in going from $d = 2$ to $d = 4$ indicates that the number of simultaneous inferences one wishes to make with the MBM method will significantly affect the number of batches that will be necessary for good use of the MBM method. Future research will involve the exploration of that factor at other levels.

ACKNOWLEDGEMENT

This research was supported partially by the Minnesota Supercomputer Institute.

REFERENCES

Anderson, T.W. (1978), "Repeated Measurements on Autoregressive Processes," *Journal of the American Statistical Association 73*, 362, 371-378.

Anderson, T.W. (1984), *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, NY.

Charnes, J.M. and W.D. Kelton (1988), "A Comparison of Confidence Region Estimators for Multivariate Simulation Output," In *Proceedings of the 1988 Winter Simulation Conference*, M. Abrams, P. Haigh, and J.C. Comfort, Eds. IEEE, Piscataway, NJ, 458-465.

Chen, D.R. and A.F Seila (1987), "Multivariate Inference in Stationary Simulation Using Batch Means," In *Proceedings of the 1987 Winter Simulation Conference*, A. Thesen, H. Grant, and W.D. Kelton, Eds. IEEE, Piscataway, NJ, 302-304.

Marse, K. and S.D. Roberts (1983), "Implementing a Portable FORTRAN Uniform (0,1) Generator," *Simulation 41*, 135-139.

Yang, W. and B.L. Nelson (1988), "Multivariate Estimation and Variance Reduction in Terminating and Steady-State Simulation," In *Proceedings of the 1988 Winter Simulation Conference*, M. Abrams, P. Haigh, and J.C. Comfort, Eds. IEEE, Piscataway, NJ, 466-472.